Visual search in naturalistic scenes from foveal to peripheral vision: A comparison between dynamic and static displays

Antje Nuthmann
Institute of Psychology, Kiel University, Kiel, Germany
Psychology Department, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK

Teresa Canas-Bajo
Vision Science Graduate Group, University of California, Berkeley, Berkeley, CA, USA
Psychology Department, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK

How important foveal, parafoveal, and peripheral vision are depends on the task. For object search and letter search in static images of real-world scenes, peripheral vision is crucial for efficient search guidance, whereas foveal vision is relatively unimportant. Extending this research, we used gaze-contingent *Blindspots* and *Spotlights* to investigate visual search in complex dynamic and static naturalistic scenes. In Experiment 1, we used dynamic scenes only, whereas in Experiments 2 and 3, we directly compared dynamic and static scenes. Each scene contained a static, contextually irrelevant target (i.e., a gray annulus). Scene motion was not predictive of target location. For dynamic scenes, the search-time results from all three experiments converge on the novel finding that neither foveal nor central vision was necessary to attain normal search proficiency. Since motion is known to attract attention and gaze, we explored whether guidance to the target was equally efficient in dynamic as compared to static scenes. We found that the very first saccade was guided by motion in the scene. This was not the case for subsequent saccades made during the scanning epoch, representing the actual search process. Thus, effects of task-irrelevant motion were fast-acting and short-lived. Furthermore, when motion was potentially present (*Spotlights*) or absent (*Blindspots*) in foveal or central vision only, we observed differences in verification times for dynamic and static scenes (Experiment 2). When using scenes with greater visual complexity and more motion (Experiment 3), however, the differences between dynamic and static scenes were much reduced.

Introduction

Visual search, a task relevant to everyday life, has been studied for decades (J. M. Wolfe, 2020; J. M. Wolfe & Horowitz, 2017). In the vast majority of studies, observers searched arbitrary static displays without moving their eyes. More recently, researchers have begun to study visual search with eye movements using photographs of real-world scenes as stimuli (e.g., Castelhano & Heaven, 2010; Foulsham & Underwood, 2011; Malcolm & Henderson, 2009). Given that the visual world across our field of view is full of information, some of these studies have investigated how important the different regions of the visual field are to the search process (Clayden et al., 2020; Nuthmann, 2013, 2014; Nuthmann et al., 2021). Here, we report three experiments in which we extend this research by investigating visual search for a nonmoving target embedded in complex dynamic and static real-world scenes.

The resolution of the visual system drops off from the fovea into the periphery gradually rather than with sudden transitions (Loschky et al., 2005), but for descriptive convenience, researchers commonly divide the visual field into three major regions: foveal, parafoveal, and peripheral. In visual-cognition research, the foveal region is considered to extend from 0° to 1° eccentricity and the parafoveal region from 1° to 4–5°, whereas the peripheral region encompasses the remainder of the visual field (Larson & Loschky, 2009; Loschky et al., 2019). The fovea and parafovea together are oftentimes referred to as central vision, whereas extrafoveal vision comprises the parafovea and the periphery. Figure 1 provides a visualization.

Citation: Nuthmann, A., & Canas-Bajo, T. (2022). Visual search in naturalistic scenes from foveal to peripheral vision: A comparison between dynamic and static displays. *Journal of Vision*, *22*(1):10, 1–31, https://doi.org/10.1167/jov.22.1.10.

https://doi.org/10.1167/jov.22.1.10 Received September 12, 2021; published January 19, 2022 ISSN 1534-7362 Copyright 2022 The Authors

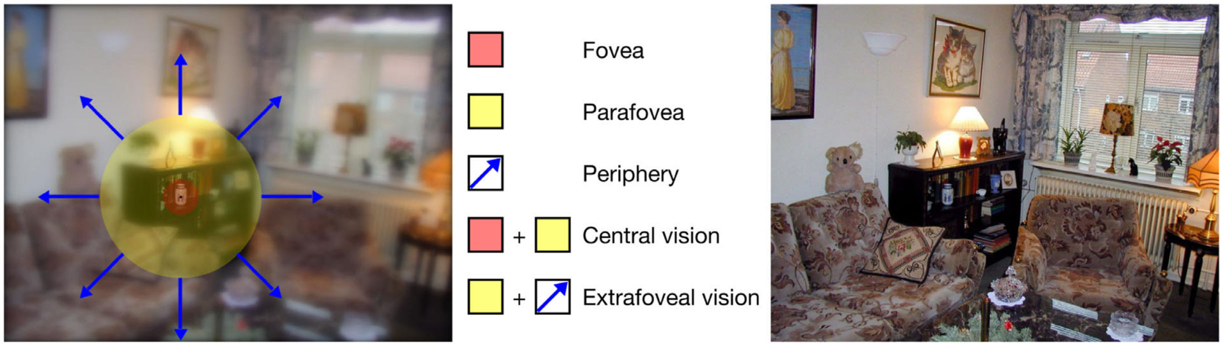


Figure 1. The drop-off in visual resolution with eccentricity and the different regions of the visual field. For the image of a British living room scene on the right, the foveated image on the left shows a reconstruction of the image on the retina, assuming an eye fixation on the vase. The image becomes more blurred away from the fixation point (black dot), which mimics the way our vision gradually loses the ability to see fine detail away from the center of vision. Superimposed are the different regions of the visual field, as defined in the text; the calculation of the extensions of the fovea and parafovea was based on the settings in which the reported experiments were conducted (size of the scene stimuli on the monitor, viewing distance). The foveated image was created with the Space Variant Imaging Toolbox by Geisler and Perry, http://svi.cps.utexas.edu/software.shtml.

As does visual acuity, motion sensitivity declines at greater eccentricities (Basler, 1906). Thus, central vision is more sensitive to motion than the periphery (Finlay, 1982). However, the decrease in sensitivity from center to periphery is proportionately less than the similar decrease in visual acuity (Post & Johnson, 1986), which is consistent with the commonly held assumption that peripheral vision is relatively specialized for motion perception.

In the real world, we move our eyes all the time to direct the high-resolution fovea to points of interest in the environment. This situation is different from basic laboratory search tasks, where observers often search simple displays covertly while holding fixation. In these experiments, the spatial limits of search are defined by the useful field of view (UFOV; Mackworth, 1965) or functional visual field (FVF; Sanders, 1970); see Hulleman and Olivers (2017) and B. Wolfe et al. (2017) for recent reviews. Ball et al. (1988) defined the UFOV as “the total visual field area in which useful information can be acquired without eye and head movements” (p. 2210). In commonly used laboratory search tasks, observers are able to decide about the presence or absence of the target without foveation of the display items (J. M. Wolfe, 2015). This is an interesting finding in itself, as it implies that targets can be discriminated outside foveal vision.

In contrast, visual-cognition researchers study visual search with eye movements and images of naturalistic scenes as stimuli. In the experiments, observers are oftentimes asked to acquire a target with their eyes, whereby the target is always present in the scene (Zelinsky, 2008). In this literature, the counterpart to the UFOV or FVF is the perceptual span, defined as the area of the visual field from which useful information is extracted during each eye fixation (see Rayner, 2014, for a review). Using the same equipment that was used for the present experiments (including stimulus size and viewing distance), Nuthmann (2013) used circular moving windows to measure the size of the perceptual span during the acquisition of a verbally cued target object in each of the scenes (e.g., a briefcase in an office scene). The span size was about 8° (radius), with the critical window capturing 44% of the entire scene (Nuthmann, 2013). Thus, the span was found to be large, highlighting the importance of parafoveal and peripheral vision to the search process.

However, active search also involves foveal analysis of the stimulus. On the way to the target, observers foveate potential target candidates and, ultimately, the target itself. This raises the question of how important foveal, parafoveal, and peripheral vision are for different subprocesses of visual search. The importance of the different regions of the visual field for the localization and recognition of objects in scenes can be directly tested by selectively denying high-resolution (or any) information in a selected region using the gaze-contingent *Moving Mask* and *Moving Window* techniques (see van Diepen et al., 1998, for a review of early scene-viewing studies).

Nuthmann (2014) applied this logic to the same object-in-scene search task that was used by Nuthmann (2013). One of her *Blindspot* or *Moving Mask* conditions simulated the absence of central vision, whereas the corresponding *Spotlight* or *Moving Window* condition simulated the absence of peripheral vision. In both conditions, search times were elevated compared to a natural vision baseline condition. At the same time, search times suggested an equal performance level for search without central vision and search without peripheral vision. Interestingly, decomposing search times into temporal epochs that are associated with particular subprocesses of search (Malcolm & Henderson, 2009) revealed a clear dissociation in behavior. When searching the scene with a central scotoma, participants were selectively impaired at verifying the identity of the target (prolonged verification time) but not in locating it. In contrast, when searching with a peripheral scotoma, participants were selectively impaired in locating the target (prolonged scanning time) but not at identifying it. This pattern of results (see also Nuthmann et al., 2021) is consistent with the assumption of a central-peripheral dichotomy, according to which peripheral vision is mainly for looking or selecting and central vision is mainly for seeing or recognizing (Zhaoping, 2019).

In the study by Nuthmann (2014), foveal vision was not necessary at all to attain normal search performance (see also McIlreavy et al., 2012). This latter finding was surprising, given the importance of foveal vision to visual search within alphanumeric displays (Bertera & Rayner, 2000) and for sentence reading (Rayner & Bertera, 1979).

Scene-viewing studies using variable-size *Moving Windows* and/or *Moving Masks* have reported systematic effects of *Moving Window*/*Mask* size on saccade amplitudes and fixation durations. It is a well-known effect that shrinking the moving window leads to shorter saccade amplitudes (e.g., Loschky & McConkie, 2002; Nuthmann, 2013, 2014; Reingold et al., 2003; Saida & Ikeda, 1979; Shioiri & Ikeda, 1989). The opposite effect is observed when variable-size moving masks are used: As the moving mask radius increases, saccade amplitudes increase (Miellet et al., 2010; Nuthmann, 2014). More generally put, observers have a tendency to fixate more locations in the undegraded scene area and fewer in the degraded area (Laubrock et al., 2013; Nuthmann, 2014). Furthermore, reducing the radius of the high-resolution moving window increases fixation durations (Loschky & McConkie, 2002; Nuthmann, 2013, 2014; Parkhurst et al., 2000).

Research using static images is neglecting the fact that naturalistic scenes change dynamically. In the real world, objects, such as people, animals, vehicles, and environmental elements (e.g., wind-generated waves), move relative to a static background. Therefore, researchers have begun to use moving images (i.e., videos) as stimuli in laboratory experiments, ranging from unedited videos of real-world scenes (e.g., Cristino & Baddeley, 2009; Smith & Mital, 2013) to segments from feature films (e.g., Hinde et al., 2017; Loschky et al., 2015; Valuch & Ansorge, 2015).

The results from several studies suggest that the presence of motion in dynamic scenes leads to different viewing patterns. During free viewing, motion and flicker are the strongest independent predictors of gaze location (Carmi & Itti, 2006; Itti, 2005; Mital et al., 2011). Moreover, when viewing dynamic scenes, the gaze of multiple viewers exhibits a much higher degree of clustering in space and time (e.g., Dorr et al., 2010; Goldstein et al., 2007; Smith & Mital, 2013), although this *attentional synchrony* (Smith & Henderson, 2008) is modulated by the specific task demands (Smith, 2013, for review). In comparison with the free viewing of dynamic scenes, during a spot-the-location task, observers' gaze exhibited less attentional synchrony (Smith & Mital, 2013).

To directly compare dynamic and static scene viewing, one frame is extracted from the original video and serves as the static scene (Açik et al., 2014; Smith & Mital, 2013). While the static scene does not include motion, its semantic content is almost identical to the original video, as long as the video does not include editorial cuts. Mean saccade amplitudes and mean fixation durations were both found to be longer for dynamic scenes than for static scenes, both for free-viewing and spot-the-location tasks (Smith & Mital, 2013).

To date, research on visual search and/or target acquisition in naturalistic dynamic scenes is still rare. One exception is the study by Reingold and Loschky (2002). In one of their experiments, observers were asked to acquire salient ring targets that moved in a straight line over video clips shot from a helicopter flying over landscapes. In an additional experiment, still images and nonmoving targets were used. In both experiments, scenes were degraded outside a gaze-contingent moving window, which was found to delay target acquisition. This "windowing effect" was larger for dynamic as compared to static scenes. However, it is not clear whether this was due to the scene type factor, because of other differences between the experiments (e.g., helicopter videos vs. residential interiors, moving target vs. nonmoving target, 3°-radius circular window vs. 12° square window).

In the present study, we used moving windows (*Spotlights*) and masks (*Blindspots*) of different sizes to investigate the importance of the different regions of the visual field to target acquisition in dynamic and static scenes. To allow for a direct comparison between dynamic and static scenes, we used videos of real-world scenes with no cuts. Moreover, we used nonmoving targets that, in most cases, did not interact with moving objects in the scene.

When using more ecologically valid stimuli like images of real-world scenes, using contextually relevant scene objects as targets seems like a natural choice (e.g., Miellet et al., 2010; Nuthmann, 2014). However, such targets are likely to show natural variation in, for example, their size and visual salience (Nuthmann et al., 2021). Moreover, scene context has a strong influence on object search (Torralba et al., 2006). To alleviate these issues, the use of (predefined) contextually irrelevant targets has proven to be useful, ranging from single letters (Clayden et al., 2020) to spatial distortions (McIlreavy et al., 2012). Importantly, scene processing and object identification also take place when observers are asked to search for a contextually irrelevant target (T. H. W. Cornelissen & Võ, 2017). For the present dynamic and static scenes, identifying suitable contextually relevant target objects proved to be very difficult. Therefore, we used ring targets (cf. Reingold & Loschky, 2002) that were inserted at the same location for the dynamic and static versions of a given scene.

Search for dynamic targets in stationary displays and search for stationary targets in dynamic displays have previously been investigated using simple displays in which a prespecified target is arrayed among distractors. This research has shown that a moving target can be found efficiently among static distractors (e.g., Hillstrom & Yantis, 1994). Moreover, the onset of uninformative motion captures attention (e.g., Abrams & Christ, 2003), and motion per se (i.e., in the absence of a motion onset) may also attract attention under certain circumstances (Abrams & Christ, 2006; Franconeri & Simons, 2003, 2005). Thus, dynamic items can attract attention in an automatic manner. Interestingly, Pinto, Olivers, and Theeuwes (2006) found that the inverse can also be true: When all items were blinking or moving except one, the dynamic items could be largely ignored and attention could be efficiently directed to the static target (see also Pinto et al., 2008). However, this was not the case when a more complex multielement asynchronous dynamic search task was used (Kunar & Watson, 2014, Experiment 8).

Of course, our experiments are different again as the search target was not the only stationary object in the scene. Moreover, using naturalistic scenes implied that there were different types of dynamic distractors in a given scene. Motion in scenes includes objects that move systematically (e.g., cars) but also elements that move in an irregular manner (e.g., tree leaves in the wind). A defining feature of moving objects is that they change position. Oftentimes, dynamic scenes not only depict object motion per se but also the onset and offset of motion. Some scenes also include abrupt onsets and offsets, that is, abrupt appearances and disappearances of nonmoving objects (e.g., traffic lights).

Research using static scenes has shown that the purpose of inspection can provide a cognitive override that renders stimulus-driven salience secondary (Underwood et al., 2006). During visual search, cognitive override of large-scale dynamic distractors in static scenes takes a couple of saccades after stimulus onset, while similar static distractors can be overridden immediately (Einhäuser et al., 2008). During free viewing of MTV-style videos, motion contrast predicted gaze location best for the first saccade after a jump cut, followed by a slow decrease over the next couple of saccades (Carmi & Itti, 2006). In a spot-the-location task, however, participants were able to direct their attention away from dynamic-scene features, that is, areas of flicker and people (Smith & Mital, 2013).

In our experiments, the target itself did not move and scene motion was not predictive of target location. Therefore, if motion attracts attention and gaze, guidance to the target may be less efficient in dynamic as compared to static scenes. Given that scene features in extrafoveal and peripheral vision are relevant for the selection of the next saccade target, search guidance may only be affected when searching dynamic scenes with a moving *Blindspot* or without a scotoma but not with a moving *Spotlight*. If, however, task-irrelevant motion can be overridden, search guidance should be equally efficient in dynamic and static scenes. During fixations, observers are thought to make accept/reject decisions about whether the fixated region contains the target (Malcolm & Henderson, 2009). When searching with a *Blindspot*, scene motion that is associated with the currently fixated location is covered by the *Blindspot* but may have been visible on the previous fixation. Conversely, *Spotlights* prevent motion from being processed in extrafoveal or peripheral vision. Therefore, the obstruction of motion by a *Blindspot* and the sudden appearance of motion within a *Spotlight* could differentially affect the fixation-based accept/reject decisions, in particular during the process of target verification.

To investigate these issues, three experiments were conducted. Experiment 1 was a conceptual replication in that we transferred the experimental design of Nuthmann (2014) from static to dynamic scenes. In the experiment, we included all six scotoma conditions (small, medium, and large *Blindspots* and *Spotlights*) that were used in the reference study. The goal of Experiment 2 was to compare dynamic and static scenes in a more direct manner. Including a manipulation of scene type required us to reduce the number of scotoma conditions. Based on the results observed in Experiment 1, the large *Blindspots* and the large *Spotlights* were dropped in Experiment 2. Experiment 3 was designed to follow up on the results from Experiment 2. Specifically, we tested the *Blindspot* but not the *Spotlight* conditions from Experiment 2 and included scenes that had a higher visual complexity and contained more motion.

General method

Participants

All participants had normal or corrected-to-normal vision by self-report. The Psychology Department at the University of Edinburgh granted ethics approval for the study, which conformed to the Declaration of Helsinki.

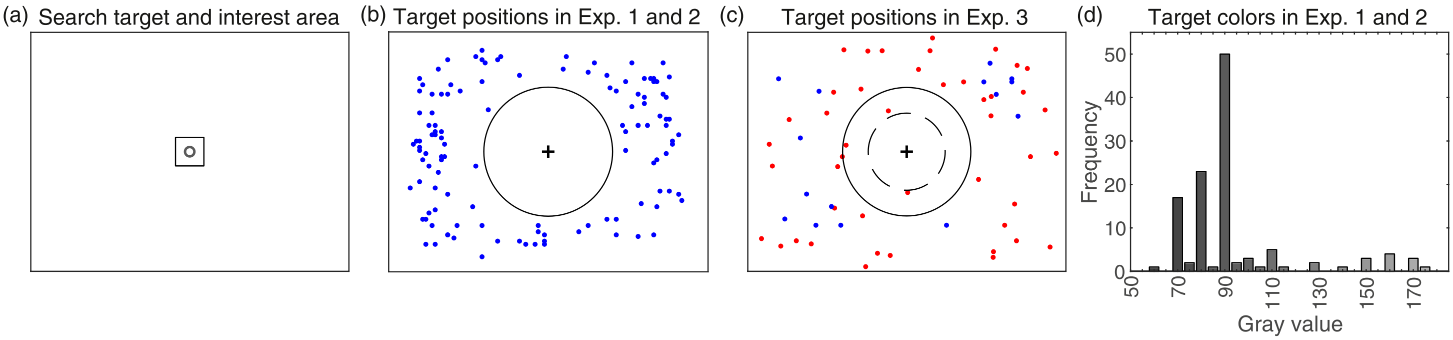


Figure 2. Search targets, one of which was present in each scene. (a) The circular search target and the square area of interest used for gaze scoring. (b) Positions of search targets in the 120 scenes used in Experiments 1 and 2. (c) Search target positions for the 60 scenes used in Experiment 3; the blue dots represent scenes that were also used in Experiments 1 and 2, whereas the red dots represent new scenes. In panels (b) and (c), the cross is the central fixation cross, and the circle with solid perimeter represents the central viewing area (radius 5°). (d) Frequency of occurrence of the different gray values in which the targets were displayed in Experiments 1 and 2, with target colors on the *x*-axis ranging from dark gray to light gray.

Apparatus

Stimuli were presented with a 140-Hz refresh rate on a 21-in. ViewSonic cathode-ray tube (CRT) monitor positioned 90 cm from participants, taking up a 24.8° × 18.6° (width × height) field of view. A chin and forehead rest was used to keep the participants' head position stable. During stimulus presentation, the eye movements of the participants were recorded binocularly with an SR Research EyeLink 1000 Desktop mount system with high accuracy (0.15° best, 0.25–0.5° typical) and high precision (0.01° RMS). The Eyelink 1000 was equipped with the 2,000-Hz camera upgrade, allowing for binocular recordings at a sampling rate of 1,000 Hz per eye. The experiments were programmed in MATLAB (The MathWorks, Natick, MA, USA) using the OpenGL-based Psychophysics Toolbox 3 (PTB-3; Brainard, 1997; Kleiner et al., 2007), which incorporates the EyeLink Toolbox extensions (F. W. Cornelissen et al., 2002). A game controller was used to record participants' behavioral responses.

Stimulus materials: scenes and search target

The stimulus material consisted of colored silent videos. Most of the 120 videos used in Experiments 1 and 2 were shot at various locations around Edinburgh and Fife (Scotland, United Kingdom); a few scenes were recorded in Berlin (Germany). In Experiment 3, 45 real-world videos shot in London (England, United Kingdom) were used along with 15 scenes from the previous experiments. The majority of videos showed outdoor scenes (e.g., roads, bridges, city streets, or parks). A few scenes were recorded in indoor environments (e.g., shopping malls). All scenes had to contain moving objects (typically people, means of transportation, or animals) or weather elements (e.g., waves); see https://youtu.be/Lkwvg8Jf_1s for three example scenes. During most recordings, the camera was fixed to a tripod for stability. The shooting speed (frame rate) was 25 frames per second (fps).

The original video clips were edited to create experimental scenes that were 20 s long and had a 4:3 aspect ratio and a 1,024 × 768 resolution. The videos did not include any editorial cuts, that is, abrupt image changes from one frame to the next. Experiments 2 and 3 included a manipulation of scene type such that dynamic scenes were compared to static counterparts. Static scenes were created by randomly choosing a frame from the original video and exporting the frame in PNG file format (cf. Smith & Mital, 2013).

In the experiments, a gray annulus was superimposed on each scene as the search target (Figure 2a). The annulus had a radius of 0.36° (15 pixels) and was drawn with the PTB function argument *penWidth* set to 2. When using images of real-world scenes as stimuli in visual-search experiments, unintended effects can arise due to the individual characteristics of these scenes. To control for item effects, we manipulated the experimental factors within scenes, without repeating scenes within participants. In a given experiment, each participant viewed each scene exactly once. By counterbalancing scene items across experimental conditions, we ensured that, across participants, each scene appeared in each condition an equal number of times. In this case, the target can be placed randomly in the respective scene (cf. McIlreavy et al., 2012), as long as the chosen position is the same for all experimental conditions. However, random placement would inevitably lead to considerable differences in target salience between scenes, that is, the degree to which the target stands out from its environment (Clayden et al., 2020). To reduce this variability, we positioned the target at an individually determined location in each scene. The location and color (shades of gray, Figure 2d) of the target were chosen such that the target was not highly salient, as this would make the search task very easy. Moreover, the target was placed away from the center of the scene (Figure 2b,c), because this is where observers began their search. In Experiments 1 and 2, locations within 5° of visual angle from the center were excluded (Figure 2b), whereas this was reduced to 3° in Experiment 3 (Figure 2c, circle with dashed perimeter), following Clayden et al. (2020). In addition, the target was placed at a location where there was no or very little motion throughout the video. In one of the videos used in Experiments 1 and 2, the target collided with a moving car for a short period of time, as revealed by post hoc motion analysis. In a few other videos, there was some motion in the background of the target (e.g., due to trees moving in the wind). For the London videos used in Experiment 3, the "no-motion criterion" was less strictly applied. In any case, since the target was superimposed on the video, it was not occluded by moving objects at any time. In summary, it is fair to say that scene motion was not predictive of target location. For the scenes used in Experiments 1 and 2, pilot studies (three people) allowed locations and color to be revised and finalized. For the 45 scenes shot in London that were used in Experiment 3, the gray value of the target was always set to 90.

Creation of gaze-contingent *Blindspots* and *Spotlights*

To implement the visual-field manipulations, gaze-contingent display change techniques were used (Reder, 1973). On the one hand, we utilized the *moving mask* technique, which was first used in the context of sentence reading (Rayner & Bertera, 1979). When applied to scene viewing, the moving mask paradigm is analogous to viewing the scene with a "blindspot": Information in the center of vision is blocked from view, while information outside the window is unaltered (Miellet et al., 2010; Nuthmann, 2014). On the other hand, we used the gaze-contingent *moving window* technique (McConkie & Rayner, 1975, for reading). Applied to scene viewing, the moving window paradigm is analogous to viewing the scene through a "spotlight": A defined region in the center of vision contains unaltered scene content, while the scene content outside the window is blocked from view (Caldara et al., 2010; Nuthmann, 2014). In this article, we use the terms *Blindspots* and *Spotlights*; when referring to *Blindspots* and *Spotlights* simultaneously, we call them scotomas.

By manipulating the radius of the gaze-contingent *Blindspots* and *Spotlights*, we created different types of scotomas, ranging from foveal scotomas (or small *Blindspots*) and central scotomas (or large *Blindspots*) to extrafoveal scotomas (or small *Spotlights*) and peripheral scotomas (or large *Spotlights*). To record baseline behavior, we implemented natural vision control conditions in which the radius of the scotoma was either 0 (*Blindspots*) or infinitely large (*Spotlights*). In that sense, the control conditions are part of our experimental manipulations of scotoma size.

The general idea underlying our scotoma implementation is to mix a foreground image and a background image via a mask image (van Diepen et al., 1994). The foreground image is formed by the experimental stimulus, that is, by the current video frame. The background image defines the content of the masked area. In the present experiments, the background image was a monochrome image (gray, RGB-value: 128, 128, 128), which implies that the moving scotomas were drawn in that color (Clayden et al., 2020). Using a uniform background image is different from Nuthmann (2014), in which low-resolution images were used instead. In this previous study, a low-resolution image was generated for each scene by applying strong spatial blurring to the original high-resolution image. However, this method is not readily applicable to dynamic scenes because residual motion remains when applying spatial blurring to individual frames.

The mask image defines the shape and size of the moving scotoma. In Experiment 1, symmetric Gaussian mask images were used, and the size of the scotoma was defined as the standard deviation of the two-dimensional Gaussian distribution (Nuthmann, 2014). The rationale for using Gaussian masks was to avoid perceptual problems that may result from sharp-boundary circular windows (see Reingold & Loschky, 2002). In Experiments 2 and 3, we improved our method by using smooth-boundary circular mask images instead of Gaussian mask images (Clayden et al., 2020; Nuthmann, 2013). When investigating the importance of the different regions of the visual field—in particular small central regions—it appears to be more appropriate to define the size of the scotoma as the radius of a circle rather than the standard deviation of a Gaussian. To avoid sharp-boundary scotomas, the perimeter of the circular mask or window was slightly faded through low-pass filtering (Clayden et al., 2020).

Working with gaze-contingent displays requires minimizing the latency of the system (Loschky & Wolverton, 2007; Saunders & Woods, 2014). For the present experiments, we used an eye tracker with a binocular sampling rate of 1,000 Hz and fast online access of new gaze samples. More precisely, the eye tracker computed a new gaze position every millisecond and made it available in less than 3 ms. The experiments were programmed using PTB-3 for MATLAB, which offers fast creation of gaze-contingent scotomas using texture mapping and OpenGL (Open Graphics Library). Gaze contingency was realized by moving the mask image across the stimulus, thereby avoiding the need for computationally expensive real-time image synthesis. Since scene images typically occupy the entire monitor space, a full refresh cycle is required to update the screen. In the experiments, the stimuli were displayed on a 140-Hz CRT monitor, which means that it took 7.14 ms for one refresh cycle to complete. Throughout the experimental trial, gaze position was continuously evaluated online. The algorithm first checked whether new valid binocular gaze samples were available. If that was the case, the center of the mask was realigned with the average horizontal and vertical position of the two eyes (Nuthmann, 2013, for discussion). If a new video frame was available, it was used as the current foreground image. The screen was updated to show the current video frame along with the scotoma at the last available gaze position. More details about the implementation, with reference to static images, are provided in Nuthmann (2013, 2014).

Procedure

In all three experiments, participants performed a visual search task in which they had to look for a gray annulus in a series of complex real-world scenes. At the beginning of the experiment, participants were informed about the gaze-contingent manipulations. In Experiments 2 and 3, participants were also told that the experiment had two parts, one in which videos were presented and another in which still images were displayed, and that their task was the same for both parts.

After instructions, a 9-point calibration, followed by a 9-point calibration accuracy test, was performed. A trial sequence began with the presentation of a black cross in the center of a gray screen on which participants were instructed to fixate. The fixation check was judged successful if gaze position, averaged across both eyes, continuously remained within an area of 40 × 40 pixels (0.97° × 0.97°) for 200 ms. If this condition was not met, the fixation check timed out after 500 ms. In this case, the fixation check procedure was either repeated or replaced by another calibration procedure. If the fixation check was successful, the scene image or video was displayed on the screen. Once subjects had found the target, they were instructed to fixate their gaze on it and press a button on the controller to end the trial (cf. Clayden et al., 2020; Glaholt et al., 2012; Nuthmann, 2014). Trials timed out 20 s after stimulus presentation if no response was made. There was an intertrial interval of 1 s before the next fixation cross was presented.

Data analysis

The raw data obtained by the eye tracker were converted into a fixation sequence matrix using the SR Research Data Viewer software. The default settings were used; in particular, the right eye was used as the reference eye. In dynamic scenes, due to the movement of (task-irrelevant) objects, it is possible that observers' gaze traces include smooth pursuit movements (SPMs). The SR Research EyeLink eye-movement event parser does not include a specific SPM detector, as it is difficult to reliably distinguish smooth pursuit movements from saccades and fixations in dynamic scenes algorithmically (but see Larsson et al., 2016; Startsev et al., 2019).

Typically, smooth pursuit movements exhibit lower velocity and acceleration than saccadic eye movements (Orban de Xivry & Lefèvre, 2007). For detection of eye-movement events during static and dynamic-scene search, we used the EyeLink's standard settings for cognitive research. Thus, saccades were detected using a 50°/s velocity threshold combined with an 8,000°/s^2 acceleration threshold. The acceleration threshold ensured that smooth pursuit movements were not misclassified as saccades (Mital et al., 2011). At the same time, this implies that the fixations in the dynamic-scene conditions may contain proportions of smooth pursuit movements.

Previous research suggests that spontaneous smooth pursuit eye movements are infrequent when viewing dynamic scenes (Smith & Mital, 2013). Nevertheless, we performed a smooth pursuit check following the procedure suggested by Hutson et al. (2017). These authors evaluated the change in gaze position during a given intersaccadic interval and removed individual intervals for which the displacement exceeded 1° from their data. The intersaccadic cleaning was applied to ensure that group differences in fixation durations were not affected by the presence of smooth pursuit.

The analysis conducted by Hutson et al. (2017) requires one to calculate the Euclidean distance between the two-dimensional gaze position at the end of the previous saccade and the one at the start of the next saccade (with the current fixation in between). Due to fixational eye movements, the eyes are never completely still (Martinez-Conde et al., 2009). Therefore, linear displacements during intersaccadic intervals also occur in the absence of smooth pursuit. In the presence of smooth pursuit, however, the displacement during the intersaccadic interval is expected to be larger than without smooth pursuit. Our analyses focused on Experiments 2 and 3, in which the data obtained for static scenes provided a baseline to which the data for dynamic scenes could be compared. The analyses considered all valid fixations and saccades from correct trials (see below). For Experiment 2, the mean displacement values did not differ between dynamic scenes ($M = 0.443°$, $SD = 0.069°$) and static scenes ($M = 0.436°$, $SD = 0.075°$), $t(39) = 0.98$, $p = 0.167$ (one-sided test). For Experiment 3, there was also no difference between dynamic ($M = 0.410°$, $SD = 0.038°$) and static ($M = 0.406°$, $SD = 0.041°$) scenes, $t(23) = 0.42$, $p = 0.340$. Of course, this does not mean that smooth pursuit never occurred during intersaccadic intervals. Instead, the results suggest that the proportion of potential smooth pursuits was low, which is consistent with previous findings (Smith & Mital, 2013). Therefore, we decided not to remove individual intersaccadic intervals from the data.

The behavioral and eye-movement data were further processed and analyzed using the R system for statistical computing (R Development Core Team, Vienna, Austria). Analyses of saccade amplitudes and fixation durations excluded fixations that were interrupted with blinks. Analysis of fixation durations disregarded fixations that were the first or last fixation in a trial. Fixation durations that are very short or very long are typically discarded, based on the assumption that they are not determined by online cognitive processes (Inhoff & Radach, 1998). This precaution was not followed here because the presence of a simulated scotoma may affect eye movements (e.g., fixations were predicted to be longer than normal).

Distributions of continuous response variables were positively skewed. In this case, variables are oftentimes transformed to produce model residuals that are more normally distributed. To find a suitable transformation, we estimated the optimal λ-coefficient for the Box–Cox power transformation (Box & Cox, 1964) using the *boxcox* function of the R package *MASS* (Venables & Ripley, 2002) with $y(\lambda) = (y^{\lambda} - 1)/\lambda$ if $\lambda \neq 0$ and $log(y)$ if $\lambda = 0$. For all continuous dependent variables, the optimal λ was different from 1, making transformations appropriate. Whenever λ was close to 0, a log transformation was chosen. We analyzed both untransformed and transformed data. As a default, we report the results for the raw untransformed data and additionally supply the results for the transformed data when they differ from the analysis of the untransformed data.

Statistical analysis using mixed models

Search accuracy was analyzed using binomial generalized linear mixed-effects models (GLMMs) with a logit link function. Continuous response variables (search times, fixation durations, saccade amplitudes) were analyzed using linear mixed-effects models (LMMs). The analyses were conducted with the R package *lme4* (version 1.1.-23; Bates et al., 2015), using the bobyqa optimizer for LMMs and a combination of Nelder–Mead and bobyqa for GLMMs. Separate (G)LMMs were estimated for each dependent variable.

A mixed-effects model contains both fixed effect and random-effect terms. Fixed effect parameters were estimated via contrast coding (Schad et al., 2020). Subjects and scene items were included as crossed random factors. The overall mean for each subject and scene item was estimated as a random intercept. In principle, the variance–covariance matrix of the random effects not only includes random intercepts but also random slopes as well as correlations between intercepts and slopes. Random slopes estimate the degree to which each main effect and/or interaction varies across subjects and/or scene items.

To select an optimal random-effects structure for (G)LMMs, we pursued a data-driven approach using backward model selection. To minimize the risk of Type I error, we started with the maximal random-effects structure justified by the design (Barr et al., 2013). This maximal structure was backward reduced using the *step* function of the R package *lmerTest* (version 3.1–2; Kuznetsova et al., 2017). If the final fitted model returned by the algorithm had convergence issues, we proceeded to fit zero-correlation parameter (zcp) models in which the random slopes are retained but the correlation parameters are set to zero (Matuschek et al., 2017; Seedorff et al., 2019). The full random-effects structure of the zcpLMM was backward reduced to arrive at a model that was justified by the data. For GLMMs, we report random intercept models because random slope models did not converge.

LMMs were estimated using the restricted maximum likelihood criterion. GLMMs were fit by Laplace approximation. For the coded contrasts, coefficient estimates (b) and their standard errors (SE) along with the corresponding t values (LMM: $t = b/SE$) or z values (GLMM: $z = b/SE$) are reported. For GLMMs, p values are additionally provided. For LMMs, a two-tailed criterion ($|t| > 1.96$) was used to determine significance at the alpha level of .05 (Baayen et al., 2008).

In the (G)LMM analyses, data from individual trials (subject–item combinations) were considered. For data figures, means were calculated for each subject, and these were then averaged across subjects. Figures were created using MATLAB. During stimulus preparation, one of the videos that was used in Experiments 1 and 2 was erroneously edited to be 10 s long only. One of the London videos used in Experiment 3 included minor viewpoint changes and zooms; another London video included image manipulations, making it less suitable as a search scene. These scenes were removed from all analyses.

Experiment 1: dynamic scenes only

In Experiment 1, participants searched for a target in dynamic real-world scenes. In the majority of trials, different types of gaze-contingent moving scotomas obstructed different regions of the participant's visual field to assess their importance to the task at hand.

The small *Blindspot* was meant to obscure foveal vision. Conversely, the small *Spotlight* left foveal vision intact but obscured extrafoveal vision. The medium *Blindspot* covered foveal and part of parafoveal vision, with the medium *Spotlight* being the inverse manipulation. Finally, the large *Blindspot* obstructed both foveal and parafoveal vision (i.e., central vision). Conversely, the large *Spotlight* left central vision intact but obstructed peripheral vision. We wanted to explore whether and how the presence of a gaze-contingent *Blindspot* or *Spotlight* and its size affected behavioral and oculomotor measures when visual search took place in dynamic scenes. If participants are able to preset their attentional biases to ignore motion, then the qualitative pattern of results should be similar to the one observed for static scenes (Nuthmann, 2014). Alternatively, if such presetting is difficult—given the complexity of natural scenes and the unpredictability of knowing where motion will occur—search efficiency may be adversely affected by the motion in the scene.

Methods: Participants, stimulus materials, and design

Thirty-two participants (19 women, 13 men) between 18 and 39 years of age ($M = 22.8$ years) took part in the experiment. Figure 3 displays still images from eight out of 120 dynamic scenes that were used in the experiment.

In Experiment 1, the visual field manipulation (*Blindspot* vs. *Spotlight*) was crossed with three different scotoma sizes. The specific radii used to create the small (radius: 1.6°), medium (2.9°), and large (4.1°) scotomas were adopted from previous studies (Loschky & McConkie, 2002; Nuthmann, 2014). The visual field manipulation was blocked such that participants completed two blocks of trials: In one block, observers' vision was impaired by a moving *Blindspot*, and in the other block, it was obstructed by a moving *Spotlight*. Each block included a separate normal-vision control condition. In sum, we implemented a 2 × 4 within-participants design, which is visualized in Figure 4. There were 15 trials in each of the eight experimental conditions. Each block started with four practice trials, one for each scotoma size condition. The order of blocks was counterbalanced across subjects. Within a block, scenes were presented randomly. A demo visualizing the gaze data for an exemplary trial from search with a small *Spotlight* is available on https://youtu.be/e8S8C2OWNWE.

Results

We analyzed behavioral indices reflecting search efficiency, in particular search accuracy and search time. Moreover, we analyzed saccade amplitudes and fixation durations across the viewing period as global visuo-oculomotor indices. For a given dependent variable, the four *Blindspot* conditions and the four *Spotlight* conditions were examined in separate analyses (Nuthmann, 2014). To test the effect of scotoma size, *backward difference coding* (also known as sliding differences or repeated contrasts) was used to compare the mean of the dependent variable for one level of the ordered factor with the mean of the dependent variable for the prior adjacent level (Venables & Ripley, 2002). The factor scotoma size was ordered according to expected task difficulty; for *Blindspots*, the ordering was No-Small-Medium-Large (see *lower* legend in Figure 5,

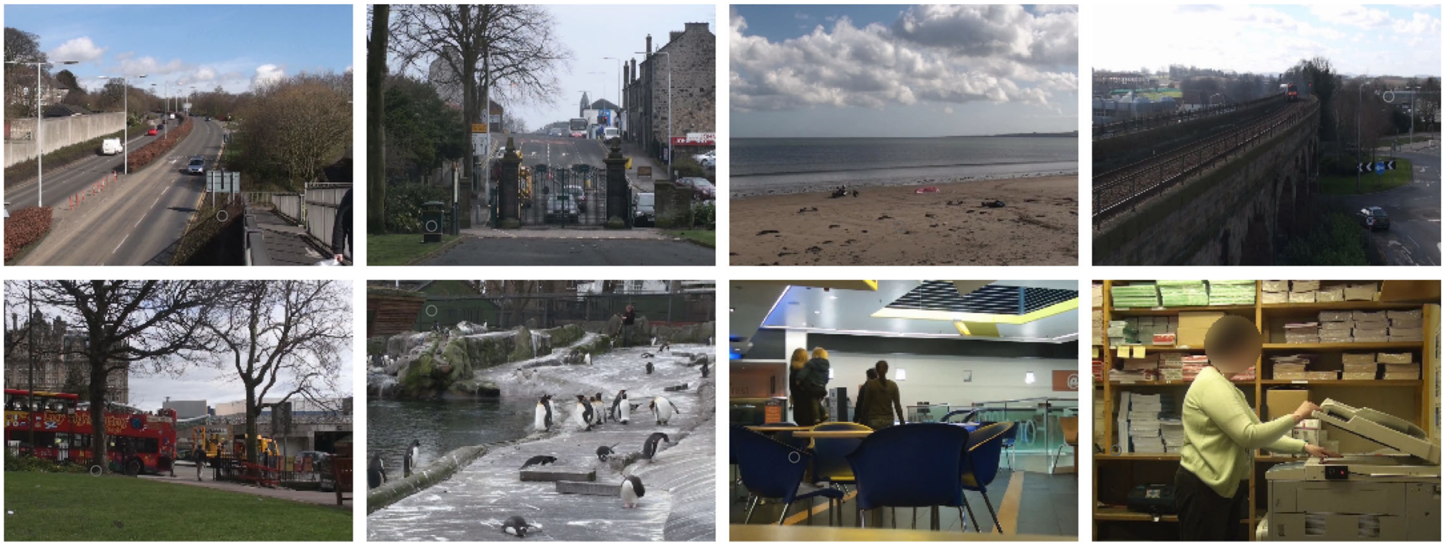
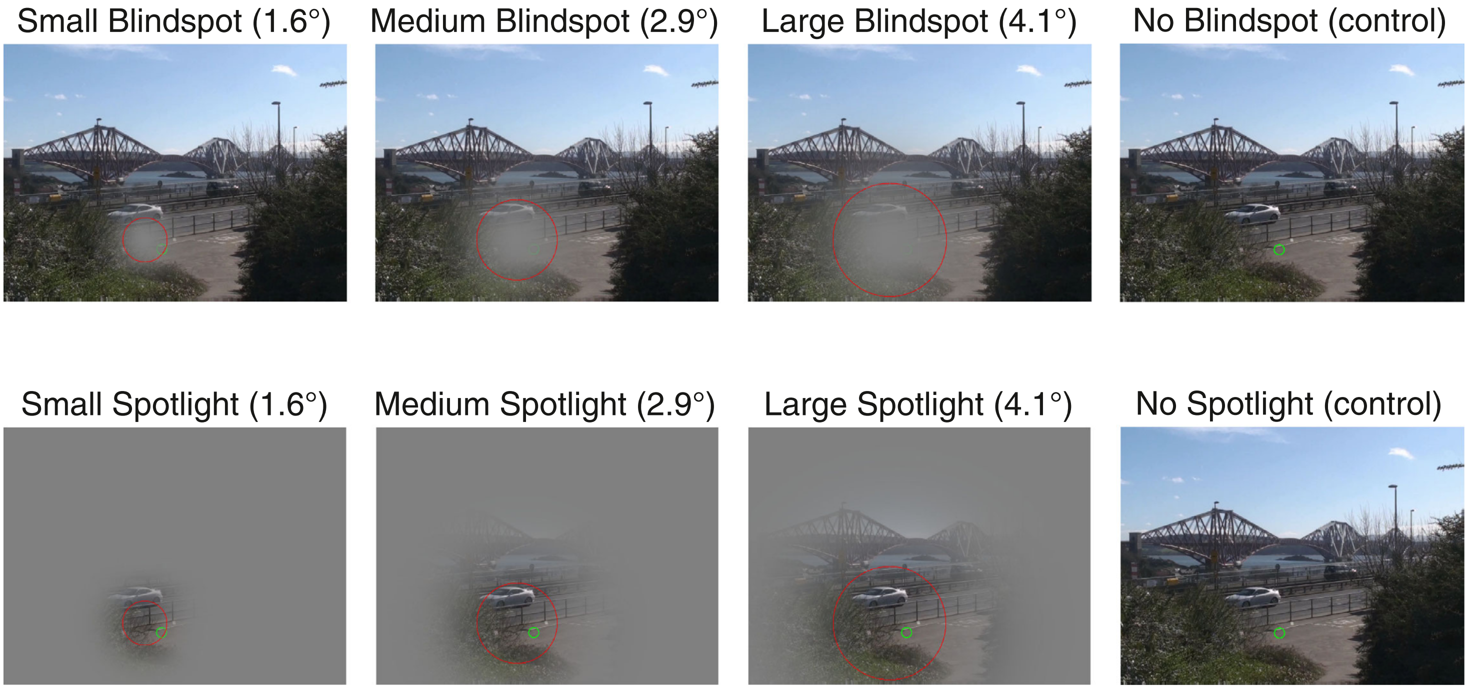


Figure 3. Still images from dynamic scenes included in Experiments 1 and 2. For eight dynamic scenes, the frame that was used as a static image in Experiment 2 is shown. In the lower right panel, the face of the person was blurred to protect their identity.



Small Blindspot (1.6°) Medium Blindspot (2.9°) Large Blindspot (4.1°) No Blindspot (control)

Small Spotlight (1.6°) Medium Spotlight (2.9°) Large Spotlight (4.1°) No Spotlight (control)

Figure 4. Experimental design for Experiment 1. In Experiment 1, observers were shown dynamic scenes only. For visualization purposes, the first frame from one of the dynamic scenes was used. In the top row, the *Blindspots* are shown along with the normal-vision control condition. The bottom row shows the *Spotlight* conditions and their control condition. The size of the *Blindspots* and *Spotlights* increases from left to right, resulting in different types of scotomas. The red circles, which were not present in the experiment, denote the standard deviation of the two-dimensional Gaussian distribution that was used to manipulate the radius of the moving *Blindspots* and *Spotlights*; the radius value is provided in the panel title. The search target is located inside the *Spotlight*. To increase its visibility to the reader, it is highlighted in green; in the experiment, it was presented in gray.

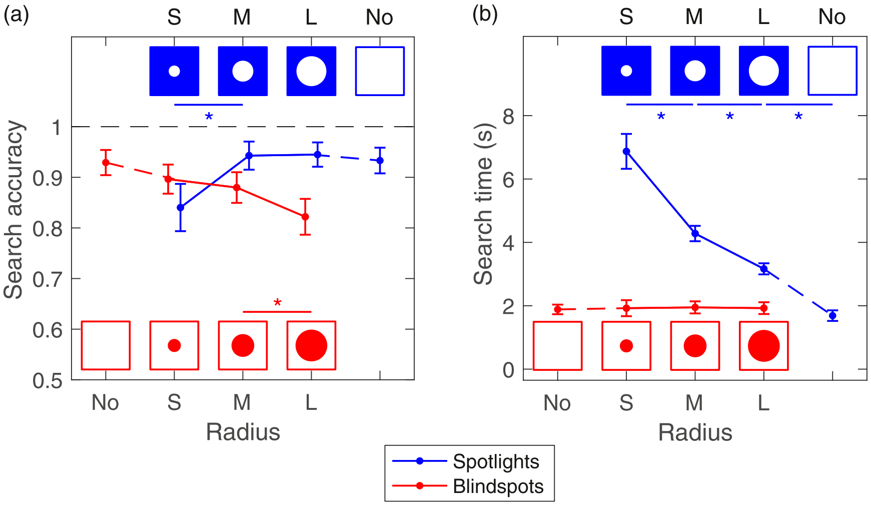


Figure 5. Mean search accuracy (a) and search time (b) for Experiment 1. The boxes represent the different experimental conditions, that is, the sizes of the gaze-contingent moving *Blindspots* (red) and *Spotlights* (blue). A line with an asterisk indicates a significant difference between adjacent conditions. Error bars are 95% within-subjects confidence intervals, using the Cousineau-Morey method (Cousineau, 2005; Morey, 2008).

from left to right), whereas for *Spotlights*, it was No-Large-Medium-Small (see *upper* legend in Figure 5, from right to left).

Behavioral measures: search accuracy and search times

Search accuracy was analyzed as the probability of correctly locating and accepting the target. A response was scored as correct ("hit") if the participant indicated to have located the target by button press and his or her gaze was within the area of interest (AOI) comprising the target (see Figure 2a). To accommodate the spatial accuracy and precision of the eye tracker, a margin was added to the search target (Holmqvist & Andersson, 2017), with the resulting AOI being 2.2° × 2.2° in size. If the participant had not responded within 20 s, the trial was coded as a "timeout." If the participant responded, but his or her gaze was not within the AOI, the response was scored as a "miss." For a given experimental condition, hits, timeouts, and misses add up to 100%. Figure 5a shows the hit probabilities for Experiment 1; the complementary timeout and miss probabilities are depicted in Supplementary Figure S1.

Overall, search accuracy was high (Figure 5a). For *Blindspots*, there was neither a significant difference between S-*Blindspots* and the control condition, $b = -0.44$, $SE = 0.24$, $z = -1.83$, $p = 0.067$, nor was there a significant difference between M-*Blindspots* and S-*Blindspots*, $b = -0.18$, $SE = 0.21$, $z = -0.87$, $p = 0.386$. However, search accuracy was significantly reduced for L-*Blindspots* as compared to M-*Blindspots*, $b = -0.51$, $SE = 0.19$, $z = -2.63$, $p = 0.008$. The reduced search accuracy for L-*Blindspots* was accompanied by more miss trials and more timed-out trials than in the control condition (Supplementary Figure S1).

For *Spotlights*, there was no significant difference between L-*Spotlights* and the control condition, $b = 0.23$, $SE = 0.28$, $z = 0.83$, $p = 0.409$. There was also no significant difference between M-*Spotlights* and L-*Spotlights*, $b = -0.04$, $SE = 0.29$, $z = -0.13$, $p = 0.896$. However, search accuracy was significantly reduced for S-*Spotlights* compared to M-*Spotlights*, $b = -1.26$, $SE = 0.25$, $z = -5.07$, $p < 0.001$. This was due to an increase in timed-out trials (Supplementary Figure S1b).

Search time was defined as the time from scene onset to successful target acquisition as indicated by the button press. Only trials with correct responses were analyzed. For *Blindspots*, none of the repeated contrasts were significant (Table 1), which means that search proficiency did not suffer when foveal (S-*Blindspot*) or even central vision (L-*Blindspot*) was not available (Figure 5b). For *Spotlights*, all of the repeated contrasts were significant, that is, search time was significantly longer for L-*Spotlights* compared to the control condition, for M-*Spotlights* compared to L-*Spotlights*, and for S-*Spotlights* compared to M-*Spotlights* (Table 1). Thus, any reduction in extrafoveal search space, even the L-*Spotlight*, increased search times, and systematically so as *Spotlights* became smaller.

Oculomotor measures: saccade amplitudes and fixation durations

Saccade amplitudes and fixation durations were analyzed as global indicators of eye-movement behavior during target acquisition (Figure 6). Relative to the respective control condition, saccade amplitudes were longer when the smallest *Blindspot* was present, S–No: $b = 0.36$, $SE = 0.13$, $t = 2.73$, and shorter when the largest *Spotlight* was present, L–No: $b = -1.14$, $SE = 0.11$, $t = -10.65$. Saccade amplitudes were significantly longer for M-*Blindspots* compared to S-*Blindspots*, $b = 0.38$, $SE = 0.13$, $t = 2.81$, and for L-*Blindspots* compared to M-*Blindspots*, $b = 0.49$, $SE = 0.14$, $t = 3.64$. Thus, the lengthening of saccade amplitudes increased with the size of the *Blindspot*. Conversely, smaller *Spotlights* were associated with progressively shorter saccade amplitudes (Table 1).

Fixation durations were significantly increased for S-*Blindspots* as compared to the control condition, $b = 16.17$, $SE = 4.66$, $t = 3.47$. There were no significant differences between M- and S-*Blindspots* and L- and M-*Blindspots*, respectively (Table 1). For *Spotlights*, all of the repeated contrasts were significant, that is, fixation durations were significantly longer for L-*Spotlights* compared to the control condition, for M-*Spotlights* compared to L-*Spotlights*, and for S-*Spotlights* compared to M-*Spotlights* (Table 1). Thus, fixation durations systematically increased as Spotlights became smaller.

Discussion

In Experiment 1, we adopted the design from Nuthmann (2014) but used dynamic rather than static scenes and different search targets. Perhaps the most striking and novel finding from Experiment 1 arose from the *Blindspot* conditions where the search times suggest that neither foveal nor central vision was necessary to attain normal search proficiency. To provide some context for the present results, we proceed by informally comparing them to the results reported by Nuthmann (2014), focusing on search times for correct trials (see Supplementary Figure S2). In the no-scotoma control condition, search times were shorter in our Experiment 1 than in the experiment by Nuthmann (2014). Compared with the modest positive slope of the *Blindspot* function observed in Nuthmann (2014), the current *Blindspot* function effectively had a slope of 0. On the other hand, the *Spotlight* function had a steeper slope for dynamic than for static scenes. For static scenes and objects as targets, the large scotomas formed a crossover point where both *Spotlights* and *Blindspots* produced similar search times (Nuthmann, 2014); this was not the case

Blindspots

Dependent variable	Parameter	Intercept	Scotoma size S–no	Scotoma size M–S	Scotoma size L–M	Random effects
Probability correct	*b*	2.34	−0.44	−0.18	−0.51	1
	SE	0.18	0.24	0.21	0.19	
	z	12.8	−1.83	−0.87	−2.63	
	p	<0.001	**0.067**	**0.386**	0.008	
Search time	*b*	2,135.9	104.56	126.48	91.46	3
	SE	180.16	99.33	125	103.47	
	t	11.86	**1.05**	**1.01**	**0.88**	
Saccade amplitude	*b*	5.47	0.36	0.38	0.49	2
	SE	0.12	0.13	0.13	0.14	
	t	43.89	2.73	2.81	3.64	
Fixation duration	*b*	207.17	16.17	2.19	2.66	4
	SE	4.88	4.66	6.56	4.88	
	t	42.49	3.47	**0.33**	**0.55**	

Spotlights

Dependent variable	Parameter	Intercept	Scotoma size L–no	Scotoma size M–L	Scotoma size S–M	Random effects
Probability correct	*b*	2.95	0.23	−0.04	−1.26	2
	SE	0.23	0.28	0.29	0.25	
	z	13.07	0.83	−0.13	−5.07	
	p	<0.001	**0.409**	**0.896**	<0.001	
Search time	*b*	3,989.51	1,448.04	1,106.45	2,544.99	4
	SE	122.88	166.2	166.07	299.28	
	t	32.47	8.71	6.66	8.5	
Saccade amplitude	*b*	3.91	−1.14	−0.35	−0.69	4
	SE	0.08	0.11	0.06	0.04	
	t	49.04	−10.65	−6.38	−17.83	
Fixation duration	*b*	220.99	30.25	10.67	29.73	7
	SE	4.82	4.24	3.07	3.67	
	t	45.82	7.13	3.48	8.09	

Table 1. Linear and generalized linear mixed models (LLM and GLMM, respectively) for Experiment 1: means (*b*), standard errors (*SE*), and test statistics (LLMs: *t* values; GLMMs: *z* values and *p* values) for fixed effects. *Note:* Nonsignificant coefficients are set in bold (LLMs: $|t| < 1.96$; GLMMs: $p > .05$). See text for further details.

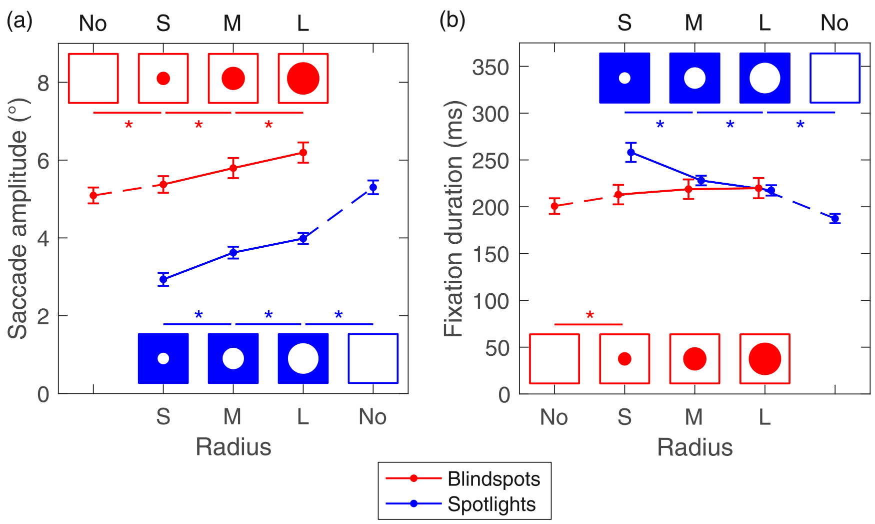


Figure 6. Mean saccade amplitudes (a) and fixation durations (b) for Experiment 1. Asterisks indicate statistically significant differences. Error bars are 95% within-subjects confidence intervals.

in Experiment 1. In summary, the availability of foveal vision (S-*Blindspot*) was equally unimportant in both experiments. Central vision was just as unimportant when searching for a contextually irrelevant target in dynamic scenes. By contrast, when searching for a contextually relevant object in static scenes without central vision being available, search times were prolonged due to longer verification times (Nuthmann, 2014). Finally, the availability of peripheral vision (L-*Spotlight*) and/or extrafoveal vision (S-*Spotlight*) was more important in the present experiment than in the one reported in Nuthmann (2014). A limiting factor of this comparison is that our observers searched for gray annuli rather than contextually relevant objects (Nuthmann, 2014) and that the scene stimuli differed. Therefore, the data from Experiment 1 alone cannot tell us conclusively whether any of the differences described above are indeed related to scene motion.

The present data indicate that the “windowing effect” on saccade amplitudes extends from visual search in static scenes (e.g., Loschky & McConkie, 2002; Miellet et al., 2010; Nuthmann, 2014) to dynamic scenes, that is, shrinking the moving *Spotlight* led to shorter saccade amplitudes, whereas increasing the moving *Blindspot* led to longer saccade amplitudes. In addition, the presence of a moving *Blindspot* was associated with elevated fixation durations, whereas this increase was not influenced by the size of the *Blindspot* (cf. Nuthmann, 2014, for static scenes). The presence of a moving *Spotlight* was also associated with increased fixation durations, and fixation durations increased as the size of the *Spotlight* decreased, thereby replicating previous results obtained with static scenes (e.g., Loschky & McConkie, 2002; Nuthmann, 2013, 2014).

Experiment 2: direct comparison of dynamic and static scenes

A limitation of Experiment 1 was that only dynamic scenes were used. The goal of Experiment 2 was therefore to directly compare visual search in dynamic and static scenes. Adding a manipulation of scene type to the design used in Experiment 1 would increase the number of experimental conditions considerably. Therefore, we dropped the large *Blindspots* and *Spotlights* for two reasons. On the one hand, the data from Experiment 1 revealed that they did not form a crossover point where both conditions produced similar search performance. On the other hand, we wanted to establish design similarities to the experiments in Nuthmann et al. (2021) where observers searched for contextually irrelevant targets in static scenes.

Methods: Participants, stimulus materials, and design

Forty-two participants were tested. The data from two participants were excluded because of their poor task performance across all experimental conditions (< 50% search accuracy, high number of timed-out trials). Included participants ($n = 40$) had a mean age of 25.5 years. The same 120 dynamic scenes as in Experiment 1 were used. The within-participant manipulations in Experiment 2 were as follows: 2 (scotoma type: *Blindspot* vs. *Spotlight*) × 2 (size: small vs. medium) × 2 (scene type: dynamic vs. static) + 2 control conditions (dynamic vs. static). Thus, Experiment 2 allowed for a direct comparison between dynamic and static scenes. The size of the scotoma was defined as the radius of a circle (small: 1.25°, medium: 2.5°). For one of the scenes used in the experiment, Figure 7 provides a visualization of the five experimental conditions that were tested for a given scene type. A demo visualizing the gaze data for an exemplary trial from search with a medium *Blindspot* is available on https://youtu.be/zOhgan5ah5w.

The scene-type manipulation was blocked such that participants completed one block of trials with dynamic scenes and another one with static scenes. The order of blocks was counterbalanced across subjects. Within each scene-type block, the visual-field manipulation was presented in three blocks. To minimize the impact of order effects, the subblock with the control trials was always presented in the middle, whereas the two scotoma subblocks were presented first or last, using counterbalancing. Within each subblock, scenes with different scotoma sizes were presented randomly. Each subblock started with a practice trial. Participants completed 12 experimental trials in each of the 10 experimental conditions.

Results

We begin by reporting measures of search efficiency and global oculomotor measures. The three *Blindspot* conditions and the three *Spotlight* conditions were analyzed separately. Unless otherwise stated, the contrast coding was as follows: For the two-level factor scene type, simple coding (−0.5/+0.5) with “dynamic scenes” as reference level was used, and for the three-level factor scotoma size, backward difference coding was used. For the dynamic-scene conditions, we additionally present analyses of fixation-based motion.

Search accuracy

One scene was excluded from the analysis of search accuracy because none of the participants were able to find the target in any of the experimental

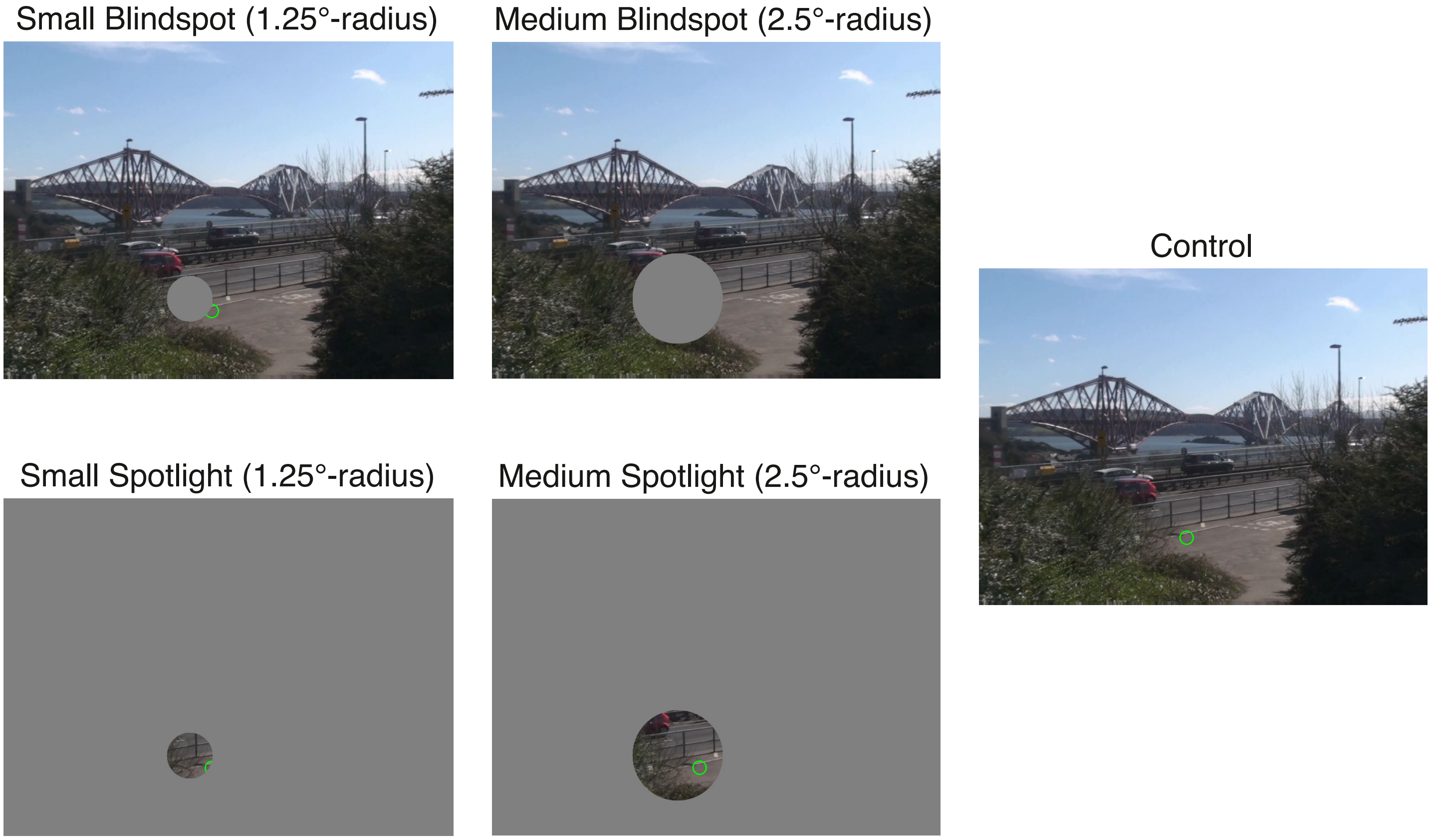

Small Blindspot (1.25°-radius)
Medium Blindspot (2.5°-radius)
Control
Small Spotlight (1.25°-radius)
Medium Spotlight (2.5°-radius)

Figure 7. Experimental design for Experiment 2. In Experiment 2, both dynamic and static scenes were used as stimuli. For visualization purposes, the five experimental conditions pertaining to static scenes are shown. The example scene is the same as in Figure 4, but this time the frame that was used as a static image in Experiment 2 is shown. A demo showing the corresponding dynamic conditions is available at https://youtu.be/b5NXnaCZqeU.

conditions. Figure 8a shows the results for hit probabilities; the complementary timeout and miss probabilities are depicted in Supplementary Figure S3.

For *Blindspots*, the main effect of scene type was significant, $b = -0.23$, $SE = 0.1$, $z = -2.19$, $p = 0.029$; search accuracy was reduced for static scenes as compared to dynamic scenes (Figure 8a). With regard to blindspot size, search accuracy was significantly reduced for S-*Blindspots* compared to the control condition, $b = -0.44$, $SE = 0.14$, $z = -3.26$, $p = 0.001$. Moreover, search accuracy was significantly smaller for M-*Blindspots* than for S-*Blindspots*, $b = -1.05$, $SE = 0.12$, $z = -8.58$, $p < 0.001$. This difference was significantly larger for static than for dynamic scenes, interaction: $b = -0.57$, $SE = 0.24$, $z = -2.35$, $p = 0.019$; the reduction in search accuracy for M-*Blindspots* in static scenes was accompanied by more miss trials and also more timed-out trials (Supplementary Figure S3). The other interaction term (scene type × S–No Blindspot) was not significant (Table 2).

For *Spotlights*, the significant main effect of scene type, $b = 0.23$, $SE = 0.1$, $z = 2.32$, $p = 0.02$, was in the opposite direction such that search accuracy was increased for static scenes as compared to dynamic scenes (Figure 8a). With regard to spotlight size, search accuracy was significantly reduced for M-*Spotlights* compared to the control condition, $b = -0.54$, $SE = 0.13$, $z = -4.29$, $p < 0.001$. Moreover, search accuracy was significantly lower for S-*Spotlights* than for M-*Spotlights*, $b = -1.49$, $SE = 0.11$, $z = -13.22$, $p < 0.001$. The low search accuracy for S-*Spotlights* was mirrored by a massive increase in timed-out trials (Supplementary Figure S3b); when only foveal vision was available, visual search proved to be very difficult. Scene type and spotlight size did not interact (Table 2).

Search time and its epochs: Blindspots

The results for search times are depicted in Figure 8b. For *Blindspots*, the data are suggestive of a crossover interaction: For static scenes, search times were numerically shorter in the control condition and longer when a M-*Blindspot* was present. To test this explicitly, we specified a mixed model using dummy coding and "dynamic scenes" and "no scotoma" as reference levels. The simple effect for scene type was

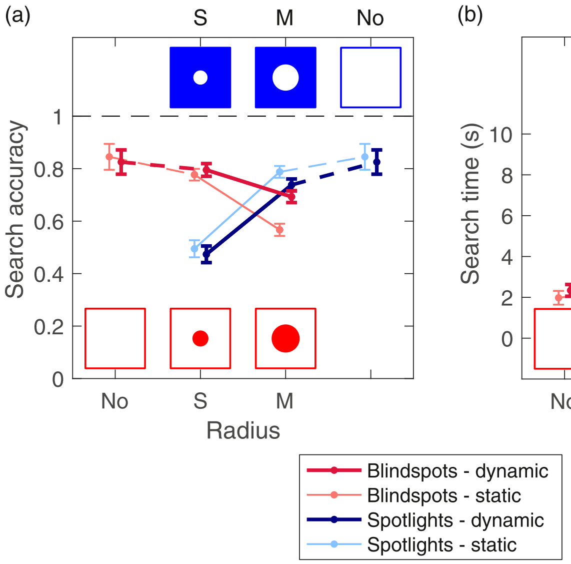


Figure 8. Mean search accuracy (a) and search time (b) for Experiment 2. Two *Blindspot* sizes (red tones) and two *Spotlight* sizes (blue tones) are compared with a no-scotoma control condition for both dynamic scenes (bold lines, darker colors) and static scenes (normal-weight lines, lighter colors). Error bars are 95% within-subjects confidence intervals.

Blindspots

Dependent variable	Parameter	Intercept	Scene type	Scotoma size S–No	Scotoma size M–S	Scene type × Scotoma size S–No	Scene type × Scotoma size M–S	Random effects
Probability correct	*b*	1.49	−0.23	−0.44	−1.05	−0.43	−0.57	2
	SE	0.18	0.1	0.14	0.12	0.27	0.24	
	z	8.28	−2.19	−3.26	−8.58	−1.6	−2.35	
	p	<0.001	0.029	0.001	<0.001	**0.109**	0.019	
Search time	*b*	2,909.33	60.39	453.77	323.08	439.25	357.36	8
	SE	243.15	134.04	116.19	156.71	193.52	281.91	
	t	11.97	**0.45**	3.91	2.06	2.27	**1.27**	
Saccade amplitude	*b*	5.95	0.17	0.9	0.93	0.15	0.72	7
	SE	0.12	0.11	0.1	0.13	0.18	0.24	
	t	51.55	**1.53**	8.63	7.04	**0.85**	3.06	
Fixation duration	*b*	219.53	−6.34	22.35	−4.68	6.71	−6.17	8
	SE	4.42	3.81	4.81	3.45	7.73	6.07	
	t	49.66	**−1.66**	4.65	**−1.36**	**0.87**	**−1.02**	

Spotlights

Dependent variable	Parameter	Intercept	Scene type	Scotoma size M–No	Scotoma size S–M	Scene type × Scotoma size M–No	Scene type × Scotoma size S–M	Random effects
Probability correct	*b*	1.07	0.23	−0.54	−1.49	0.12	−0.23	2
	SE	0.13	0.1	0.13	0.11	0.25	0.22	
	z	8.4	2.32	−4.29	−13.22	0.49	−1.03	
	p	<0.001	0.02	<0.001	<0.001	**0.622**	**0.304**	
Search time	*b*	5,597.18	−237.58	3,434.56	3,109.09	−94.24	420.99	6
	SE	163.81	217.92	169.52	337.78	339.77	520.17	
	t	34.17	**−1.09**	20.26	9.2	**−0.28**	**0.81**	
Saccade amplitude	*b*	3.45	−0.03	−2.27	−0.34	0.16	−0.03	8
	SE	0.08	0.09	0.13	0.05	0.14	0.06	
	t	40.88	**−0.35**	−17.33	−7.27	**1.15**	**−0.55**	
Fixation duration	*b*	243.57	−19.46	34.4	45	−15.74	−1.03	11
	SE	3.47	4	5.09	2.69	5.18	5.41	
	t	70.13	−4.87	6.76	16.72	−3.04	**−0.19**	

Table 2. Linear and generalized linear mixed models (LLM and GLMM, respectively) for Experiment 2: means (*b*), standard errors (*SE*), and test statistics (LLMs: *t* values; GLMMs: *z* values and *p* values) for fixed effects. *Note:* Nonsignificant coefficients are set in bold (LLMs: $|t| < 1.96$; GLMMs: $p > .05$). See text for further details.

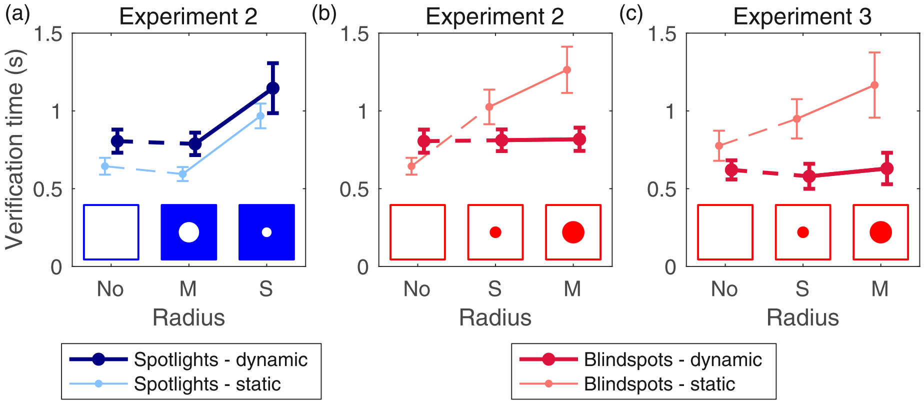


Figure 9. Mean verification times in Experiment 2 (*Spotlights*: panel a, *Blindspots*: panel b) and Experiment 3 (c). Experimental conditions are ordered according to expected task difficulty. The data points for the no-scotoma conditions are the same in panels (a) and (b) as they represent the shared control condition. Error bars are within-subject standard errors.

significant, $b = -335.67$, $SE = 134.27$, $t = -2.5$, indicating that search times were shorter for static than for dynamic scenes when no scotoma was present. Importantly, for dynamic scenes, there was neither a significant difference between S-*Blindspots* and the control condition, $b = 217.54$, $SE = 136.7$, $t = 1.59$, nor between M-*Blindspots* and the control condition, $b = 343.39$, $SE = 182.55$, $t = 1.88$. The scene type × S–No Blindspot was not significant, $b = 433.6$, $SE = 238.73$, $t = 1.82$, but the scene type × M–No Blindspot interaction was, $b = 792.27$, $SE = 282.19$, $t = 2.81$. To directly test the effect blindspot size had on search times for static scenes, we reran the analysis with static rather than dynamic scenes as the reference level. For static scenes, the search-time difference between S-*Blindspots* and the control condition was significant, $b = 662.96$, $SE = 150.67$, $t = 4.4$, and so was the difference between M-*Blindspots* and the control condition, $b = 1{,}112.26$, $SE = 197.81$, $t = 5.62$. Taken together, *Blindspots* had no significant effect on search times for dynamic scenes, but they did for static scenes.

In additional analyses, we tested whether these search-time costs were due to increased verification times, as suggested by previous research (Clayden et al., 2020; Nuthmann, 2014; Nuthmann et al., 2021). To this end, search time was split into three subcomponents: search initiation time, scanning time, and verification time (e.g., Clayden et al., 2020; Malcolm & Henderson, 2009; Nuthmann, 2014; Nuthmann & Malcolm, 2016); see https://youtu.be/zOhgan5ah5w for a demo.

Search initiation time is the latency of the first saccade, which equates to the duration of observers' initial fixation at the center of the screen (Malcolm & Henderson, 2009). This epoch reflects the time it took to process the gist of the scene and to prepare the first saccade. Scanning time is the time between the first saccade and the first fixation on the target. Verification time indexes the elapsed time between the beginning of the first fixation on the target and search termination. This component of search reflects the time needed to decide that the target is in fact the target. Short verification times typically arise if subjects make no more than one fixation within the AOI comprising the target before pressing the button (see demo). In cases where the initial fixation is followed by one or more immediate refixations, verification time is equivalent to first-pass gaze duration[1] on the target. Particularly long verification times include instances in which observers fixated the target but then continued searching before returning to it (Castelhano et al., 2008; Clayden et al., 2020). If foveal or central vision is masked, it can happen that the eyes move off the target to unmask it and then process it in parafoveal or peripheral vision, thereby increasing verification time (Clayden et al., 2020; Nuthmann, 2014).

Figure 9b depicts the pattern of verification times for the *Blindspot* and control conditions. Verification times were subjected to a mixed model using dummy coding and "static scenes" and "no scotoma" as reference levels. The simple effect for scene type was not significant, $b = 165.27$, $SE = 94.98$, $t = 1.74$. For the transformed data, however, the effect was significant, $b = 0.15$, $SE = 0.06$, $t = 2.43$, suggesting that verification times were longer for dynamic than for static scenes when no scotoma was present. For static scenes, the difference in verification times between S-*Blindspots* and the control condition was significant, $b = 432.97$, $SE = 122.47$, $t = 3.54$, and so was the difference between M-*Blindspots* and the control condition, $b = 703.12$, $SE = 148.99$, $t = 4.72$. The scene type × S–No Blindspot interaction was significant, $b = -380.66$, $SE = 135.93$, $t = -2.8$, and so was the scene type × M–No Blindspot interaction, $b = -675.13$, $SE = 146.93$, $t = -4.59$.

For completeness, Supplementary Figure S4 and Supplementary Table S1 additionally provide the results for search initiation and scanning times, using the

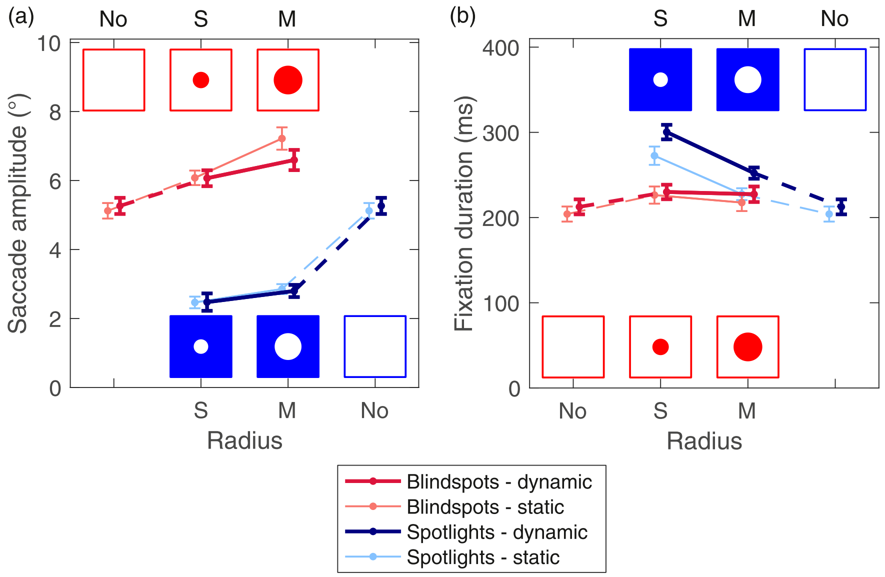


Figure 10. Mean saccade amplitudes (a) and fixation durations (b) for Experiment 2. Two *Blindspot* sizes (red tones) and two *Spotlight* sizes (blue tones) are compared with a no-scotoma control condition for both dynamic scenes (bold lines, darker colors) and static scenes (normal-weight lines, lighter colors). Error bars are 95% within-subjects confidence intervals.

default contrast coding. In brief, search initiation times were significantly shorter for static as compared to dynamic scenes, $b = -60.22$, $SE = 7.77$, $t = -7.75$. For scanning times, there was no significant difference between dynamic and static scenes, $b = -69.95$, $SE = 90.68$, $t = -0.77$.

Search time and its epochs: Spotlights

For *Spotlights*, the data depicted in Figure 8b indicate numerically shorter search times for static as compared to dynamic scenes. This difference was not significant for the untransformed data, $b = -237.58$, $SE = 217.92$, $t = -1.09$; for the transformed data, however, it was significant, $b = -0.01$, $SE = 0.004$, $t = -2.61$. Moreover, search time was significantly longer for M-*Spotlights* compared to the control condition and for S-*Spotlights* compared to M-*Spotlights* (Table 2). Spotlight size and scene type did not interact (Table 2).

A decomposition of search times into three epochs revealed that the numerically shorter search times for static as compared to dynamic scenes were due to shorter search initiation times, $b = -53.27$, $SE = 7.22$, $t = -7.38$, and shorter verification times, $b = -132.13$, $SE = 43.69$, $t = -3.02$; scanning times did not differ for dynamic and static scenes, $b = -49.5$, $SE = 200.62$, $t = -0.25$. Moreover, smaller *Spotlights* were associated with progressively longer search initiation and scanning times. The complete results are provided in Supplementary Figure S4 and Supplementary Table S1. Figure 9a depicts the pattern of verification times for the *Spotlight* and control conditions.

In the above analyses, the normal-vision control conditions were included as part of the experimental manipulations of scotoma size. Therefore, additional LMMs were run on the search-time data from the control conditions only. The maximal model included the fixed effect scene type (simple coding) along with four random effects. When no scotoma was present, search times were significantly shorter for static than for dynamic scenes, $b = -424.52$, $SE = 153.05$, $t = -2.77$.

Saccade amplitudes and fixation durations

Mean saccade amplitudes for the experimental conditions tested in Experiment 2 are depicted in Figure 10a. For *Blindspots*, the main effect of scene type was not significant, $b = 0.17$, $SE = 0.11$, $t = 1.53$. With regard to blindspot size, saccade amplitudes were significantly larger for S-*Blindspots* compared to the control condition, $b = 0.9$, $SE = 0.1$, $t = 8.63$. Moreover, saccade amplitudes were significantly larger for M-*Blindspots* than for S-*Blindspots*, $b = 0.93$, $SE = 0.13$, $t = 7.04$. This difference was significantly larger for static than for dynamic scenes, interaction: $b = 0.72$, $SE = 0.24$, $t = 3.06$. The other interaction term (scene type × S–No Blindspot) was not significant (Table 2).

For *Spotlights*, the main effect of scene type was not significant, $b = -0.03$, $SE = 0.09$, $t = -0.35$. However, saccade amplitudes were significantly shorter for M-*Spotlights* compared to the control condition, $b = -2.27$, $SE = 0.13$, $t = -17.33$, and significantly shorter for S-*Spotlights* compared to M-*Spotlights*, $b = -0.34$, $SE = 0.05$, $t = -7.27$. There were no significant interactions between scene type and spotlight size (Table 2).

For *Blindspots*, the data depicted in Figure 10b indicate somewhat shorter fixation durations for static as compared to dynamic scenes. This difference was not significant for the untransformed data, $b = -6.34$, $SE = 3.81$, $t = -1.66$; for the transformed data, however, it was significant, $b = -0.13$, $SE = 0.05$, $t = -2.39$. Fixation durations were significantly longer for S-*Blindspots* as compared to the control condition, $b = 22.35$, $SE = 4.81$, $t = 4.65$, but not for M-*Blindspots* as compared to S-*Blindspots*, $b = -4.68$, $SE = 3.45$, $t = -1.36$. There were no significant interactions between scene type and blindspot size (Table 2).

For *Spotlights*, fixation durations were significantly shorter for static as compared to dynamic scenes, $b = -19.46$, $SE = 4$, $t = -4.87$. Fixation durations were significantly longer for M-*Spotlights* as compared to the control condition, $b = 34.4$, $SE = 5.09$, $t = 6.76$, whereby this difference was significantly smaller for static as compared to dynamic scenes (scene type × M–No interaction: $b = -15.74$, $SE = 5.18$, $t = -3.04$). Moreover, fixation durations were significantly longer for S-*Spotlights* as compared to M-*Spotlights*, $b = 45$, $SE = 2.69$, $t = 16.72$; the scene type × S–M interaction was not significant (Table 2).

Analysis of fixation-based motion

If scene motion had an influence on saccade target selection in our task, the search-time costs for dynamic scenes that we observed in the control condition should be reduced in the *Spotlight* conditions, in which much less motion was visible. However, this does not appear to be the case as the search-time analysis for *Spotlights* did not reveal an interaction between scene type and scotoma size. Note that this does not imply that motion did not play any role at all as effects of scene motion could be fast-acting and short-lived.

To test this, each video was transformed to a motion energy video by computing optical flow by means of the Lucas–Kanade derivative of Gaussian method (Lucas & Kanade, 1981). The implementation offered by MATLAB's Computer Vision Toolbox was used with default settings. The algorithm estimates the displacement of each pixel between the current frame n and the previous frame $n - 1$. Each video consisted of 500 individual fames (25 fps * 20 s), yielding 499 pairwise comparisons. The motion between two consecutive video frames is described by four 1,024 × 768 matrices describing (a) the x component of velocity, (b) the y component of velocity, (c) the orientation of optical flow, and (d) the magnitude of optical flow. For our purposes, we used the magnitude of optical flow. For illustration, Figure 11a shows a frame from a high-motion video in which trees and plants were seen to move during windy Scottish weather. The corresponding motion map is depicted in Figure 11b, where brighter colors correspond to higher optical flow magnitudes. The actual video and the corresponding motion magnitude video are available on https://youtu.be/cgIGYINIKOU.

To determine motion around fixations, circular patches were centered on the fixation points. The patch radius was equal to the scotoma radius; for the no-scotoma control condition, both radii (1.25 and 2.5°) were used. Fixation-based motion was quantified as the sum of optical flow magnitudes within the patch. Three types of fixations were considered separately: (a) the first fixation following search initiation (i.e., the first fixation during the scanning epoch), (b) all remaining fixations during the scanning epoch, and (c) all fixation during the verification epoch.

In the experiments, observers started searching the scene from a central fixation position. At the end of the search initiation epoch, the eyes moved to the location of the first fixation. For illustration, the yellow arrows in the top-row panels of Figure 11 represent an imaginary saccade from the central fixation (yellow circle with dashed perimeter) to the first fixation (yellow circle with solid perimeter) in the scene.

Logically, the location for the first fixation must have been selected toward the end of the central fixation already. For our calculations, we assumed that the decision about where to fixate next was made 100 ms prior to the start of the first fixation, thereby taking the duration of the saccade as well as the duration of an assumed nonlabile stage of saccade programming into account (Walshe & Nuthmann, 2015). Accordingly, we evaluated motion at 100 ms prior to the start of the first fixation (right-pointing triangle symbols in Figure 11). Interestingly, Vig et al. (2011) reported that when participants free viewed dynamic natural scenes, there was a near-zero average lag between changes in the scene and the responding eye movements, suggesting that gaze anticipated motion peaks. To account for this possibility, we also evaluated motion at the start of the first fixation by using the first frame that appeared in this fixation (circle symbols in Figure 11).

Using a patch radius that equals the scotoma radius yields two informative special cases. For the *Spotlight* conditions, the motion at 100 ms prior to the start of the fixation is invisible due to the scotoma (Figure 11c). Therefore, the *Spotlight* conditions provide a baseline to which to compare the motion values from conditions where extrafoveal and peripheral vision are intact (i.e., the control condition and the *Blindspot* conditions). Conversely, for the *Blindspot* conditions, the motion measured at the start of fixation is invisible due to the scotoma, while the motion at the selected location was visible during the previous fixation (Figure 11d).[2]

First-fixation analyses included all trials, whether target acquisition was successful or not. Fixations during the scanning and verification epochs were from

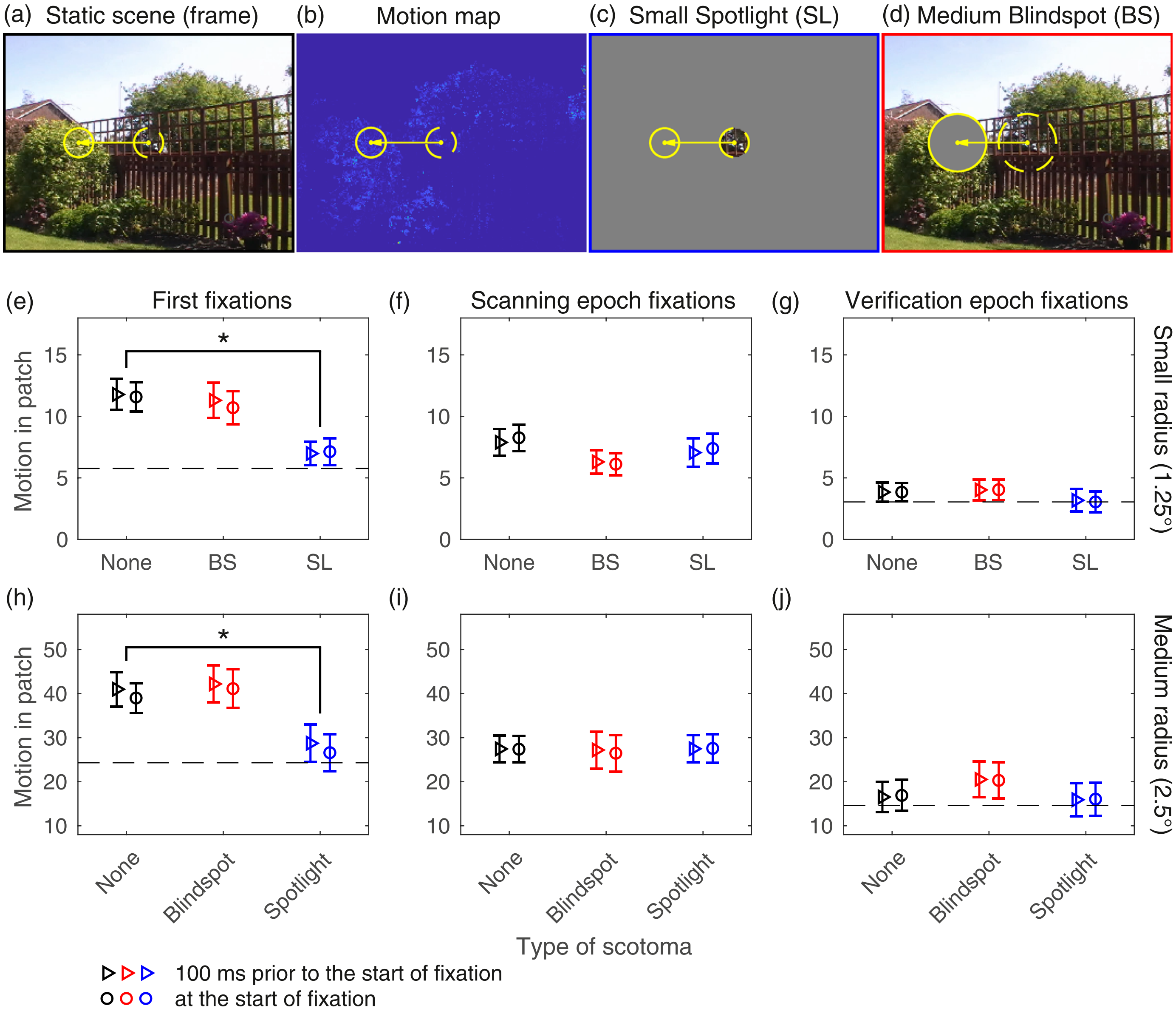


Figure 11. Analysis of motion around fixation. For a high-motion video used in Experiment 2, panel (a) shows the frame that was used as the static scene, and panel (b) shows the map depicting the motion between this frame and the previous one. In the top-row panels, the left-pointing yellow arrow represents an imaginary saccade from the central fixation (yellow circle with dashed perimeter) to the first fixation (yellow circle with solid perimeter) in the scene. Panel (c) depicts a central fixation with a small *Spotlight* (radius: 1.25°), whereas panel (d) depicts a first fixation with a medium *Blindspot* (radius: 2.5°). Fixation-based motion was determined for fixation patches that had a radius equal to the scotoma radius (middle row: radius 1.25°, bottom row: radius: 2.5°), whereby both scotoma radii were applied to the no-scotoma control condition. A given row depicts results for three different types of fixations; see text for details. Motion was calculated as the summed magnitudes of optical flow in the patch. This was done 100 ms prior to the start of fixation (right-pointing triangle symbols) and at the start of fixation (circle symbols) for the no-scotoma control condition (black), *Blindspots* (BS, red), and *Spotlights* (SL, blue). For a given radius, the *y*-axis was scaled to be the same for all three data panels. Error bars are within-subject standard errors.

correct trials only. Fixations that were contaminated by blinks and fixations for which the previous fixation was too short (i.e., sum of its duration and the duration of the incoming saccade ≤ 100 ms) were excluded from analyses. Regarding the results, we emphasize overall data patterns rather than significance of individual data points.

If motion guides early fixations in particular, motion values should be largest for the first fixation (Carmi & Itti, 2006). Moreover, we expected motion values during the verification epoch to be relatively low because there was little to no motion in the target region. Indeed, for the control condition, motion values were largest for first fixations (Figure 11e,h) and lowest for fixations made during the verification epoch (Figure 11g,j).

The critical question was whether motion values for moving *Spotlights* and *Blindspots* were significantly different from the control condition. To account for all combinations of radii (2), types of fixations (3), and temporal reference points (2), 12 separate mixed-model analyses were conducted. For the three-level factor scotoma type, simple coding was used. The no-scotoma control condition served as the reference level, which allowed us to test whether there were any differences between *Blindspots* and the control condition or between *Spotlights* and the control condition. The LMMs were random intercept models that included random intercepts for subjects and scene items. To facilitate the interpretation of results, we additionally computed mean motion values (a) within circular patches at the center of the video during the first couple of frames (horizontal dashed line in Figure 11e,h) and (b) for patches centered on the target throughout the entire duration of the video (horizontal dashed line in Figure 11g,j).

For the first fixation, motion was significantly reduced when searching with a moving *Spotlight*, both at the beginning of the fixation (radius 1.25°: $b = -4.61$, $SE = 1.26$, $t = -3.67$, see Figure 11e; radius 2.5°: $b = -10.31$, $SE = 3.38$, $t = -3.05$, see Figure 11h) and 100 ms before (radius 1.25°: $b = -4.83$, $SE = 1.36$, $t = -3.55$, radius 2.5°: $b = -9.46$, $SE = 3.75$, $t = -2.52$). By contrast, there were no significant differences between *Blindspot* fixations and control fixations in any of the first-fixation analyses.

For scanning-epoch fixations with the 2.5° radius, there were no significant differences between *Blindspot* fixations and control fixations or *Spotlight* fixations and control fixations (see Figure 11i). For the 1.25° radius, however, compared to the control condition, motion was significantly reduced for both *Blindspots* and *Spotlights*, both 100 ms prior to the start of fixation (*Blindspots*: $b = -2.11$, $SE = 0.81$, $t = -2.59$, *Spotlights*: $b = -1.58$, $SE = 0.76$, $t = -2.08$) and at the start of fixation (*Blindspots*: $b = -2.54$, $SE = 0.80$, $t = -3.17$, *Spotlights*: $b = -1.60$, $SE = 0.75$, $t = -2.13$). Interestingly, the effects were not significant when the mixed model included by-subject random intercepts only. For verification-epoch fixations, all analyses yielded nonsignificant fixed effects contrasts (see Figure 11g,j).

Discussion

For the no-scotoma control condition for dynamic scenes, search accuracy was lower in Experiment 2 than in Experiment 1, and search times for correct trials were longer. For the no-scotoma conditions in Experiment 2, search accuracy was at comparable levels for dynamic and static scenes. In the scotoma conditions, however, there were subtle differences in search accuracy for dynamic and static scenes: For *Blindspots*, search was more accurate for dynamic than for static scenes, whereas for *Spotlights*, it was less accurate for dynamic than for static scenes.

For the natural vision control conditions, there was evidence for longer search times for dynamic than for static scenes. This was due to significantly longer initiation times along with longer verification times for dynamic scenes. Search times were not elevated when searching dynamic scenes with S- and M-*Blindspots*, which agrees with the results from Experiment 1. For static scenes, however, search times systematically increased as less information was available in the center of vision. A gaze-based decomposition of search times suggested that the observed search-time costs were due to increased verification times. Finally, the *Spotlight* analyses revealed both longer search initiation and verification times for dynamic scenes, leading to numerically longer search times for dynamic than for static scenes.

The saccade-amplitude data showed the well-known “windowing effect” for both types of scenes. For both dynamic and static scenes, fixation durations were prolonged when a moving *Blindspot* was present compared with the no-scotoma control condition. For *Spotlights*, fixation durations were longer for dynamic than for static scenes. Moreover, for both types of scenes, fixation durations increased as *Spotlights* decreased.

Motion in extrafoveal and peripheral vision affected the selection of the target for the very first saccade. For the subsequent scanning epoch, however, the data suggest that the task-irrelevant motion could be overridden.

Experiment 3: partial replication of Experiment 2 with different scenes

The results obtained in Experiment 2 showed subtle differences between dynamic and static scenes, in particular for the *Blindspot* conditions. For further investigation, we conducted a third experiment that included the *Blindspot* and control conditions but not the *Spotlight* conditions from Experiment 2. Going beyond a pure replication, Experiment 3 included scenes with more motion and greater visual complexity than in the previous experiments.

More motion could lead to more peripheral capture and therefore exacerbate the differences between dynamic and static scenes. Scenes with more motion tended to be more visually complex too (see below).

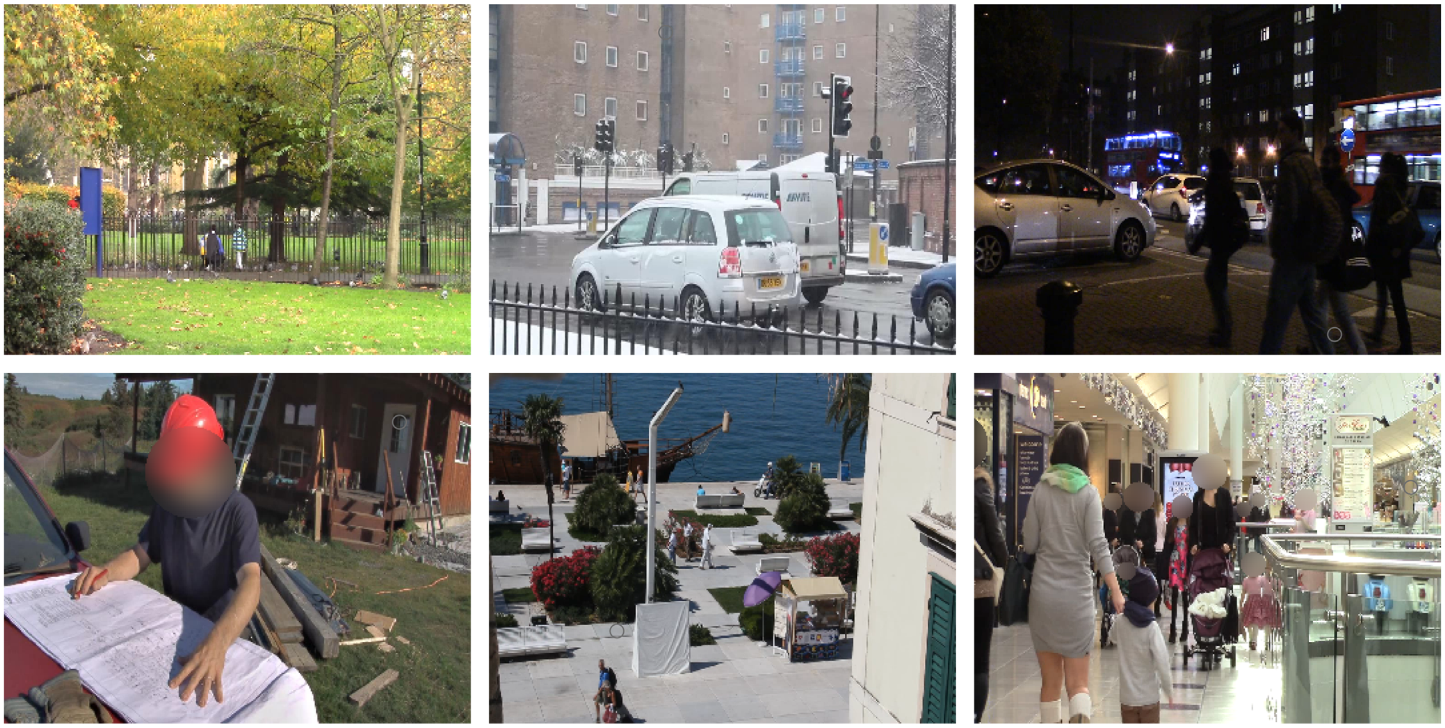


Figure 12. Still images from dynamic scenes used in Experiment 3. For six of the new scenes, the frame that was used as a static image is shown. In the figure, faces were blurred for identity protection.

For a difficult letter-in-scene search task, Henderson et al. (2009) found that more cluttered static scenes were associated with longer search times. Moreover, visual complexity generally worsens peripheral vision (see Rosenholtz, 2016, for a review). Therefore, and alternatively, search in more complex scenes could reduce or negate the differences between dynamic and static scenes.

Methods: Participants, stimulus materials, and design

Twenty-four participants (15 women, 9 men, $M = 26$ years, age range: 18–40 years) were tested in Experiment 3. The new scenes were 45 real-world videos that were recorded in London (England, United Kingdom). Most of the scenes were shot in outdoor environments. For six of the new dynamic real-world scenes, Figure 12 shows the frame that was used as a static image in the experiment.

For the static images, we assessed their visual complexity by using the Feature Congestion measure of visual clutter introduced by Rosenholtz, Li, and Nakano (2007). For each image, a scalar representing the clutter of the entire image was computed, with larger values indicating more visual clutter. For the static images from the London videos, mean Feature Congestion clutter was higher ($M = 4.37$, $SD = 1$) than for the videos used in the other experiments ($M = 2.74$, $SD = 0.64$), $t(54.84) = 9.93$, $p < .001$. In a second analysis, we did not consider the entire image but only the AOI comprising the target. Again, the clutter values were significantly higher for the London images ($M = 5.3$, $SD = 2.23$) than for the other images ($M = 2.82$, $SD = 1.23$), $t(51.58) = 6.94$, $p < .001$.

To estimate motion in the videos, we used the optical flow estimation method described above. For each original video, a motion energy video was created by using the magnitude of optical flow (cf. Mayer et al., 2015). For a given frame (or matrix) of the motion video, the optical flow magnitudes were summed up. Then, for a given video, the mean across the summed magnitudes was calculated, yielding one aggregate motion value per video. These motion values were significantly higher for the 43 London videos as compared to the 119 videos used in the previous experiments, $t(65.31) = 4.55$, $p < .001$. For the AOI comprising the target, motion values were also significantly higher for the London videos than for the other videos, $t(52.27) = 2.85$, $p = .006$.

To increase the number of stimuli, 15 scenes from Experiments 1 and 2 were additionally included. Therefore, 58 scenes entered the analysis stage. For a given scene, the motion in the video and the visual clutter in the static frame were correlated such that scenes with more motion were also more visually complex, $r(56) = 0.594$, $p < .001$.

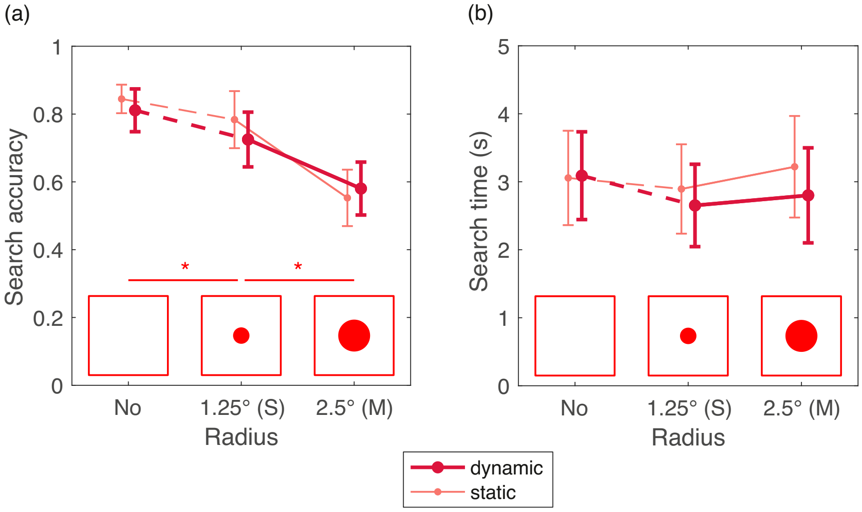


Figure 13. Mean search accuracy (a) and search time (b) for Experiment 3. Two different *Blindspot* sizes are compared with a no-scotoma control condition for both dynamic scenes (color: crimson, bold line, large marker) and static scenes (color: salmon, normal-weight line, normal-size marker). Error bars are 95% within-subjects confidence intervals. A line with an asterisk indicates a significant difference between adjacent conditions.

In Experiment 3, there were no *Spotlight* conditions. Apart from that, the design was identical to Experiment 2. Thus, we implemented a 2 (scene type: static vs. dynamic) × 3 (Blindspot size: none vs. small vs. medium) design. There were 10 trials in each of the six conditions.

Results

Unless otherwise stated, the contrast coding was as follows: For the two-level factor scene type, simple coding with the reference level “dynamic scenes” was used; for the three-level factor blindspot size, backward difference coding was used.

Search accuracy

For four of the scenes, none of the participants were able to find the target in any of the experimental conditions because it was not sufficiently salient. These scenes were excluded from the analysis of search accuracy. Figure 13a shows the hit probabilities; the complementary timeout and miss probabilities are depicted in Supplementary Figure S5. For hits, there was no significant difference between dynamic and static scenes, $b = 0.16$, $SE = 0.16$, $z = 1.01$, $p = 0.312$. With regard to blindspot size, search accuracy was significantly reduced for S-*Blindspots* compared to the control condition, $b = -0.73$, $SE = 0.21$, $z = -3.47$, $p = 0.001$. Moreover, search accuracy was significantly smaller for M-*Blindspots* than for S-*Blindspots*, $b = -1.21$, $SE = 0.18$, $z = -6.7$, $p < 0.001$. Scene type and blindspot size did not interact (Table 3). The reduction in search accuracy for M-*Blindspots* was accompanied by more timed-out trials and more miss trials (Supplementary Figure S5).

Search time and its epochs

For search time (Figure 13b), there were no significant effects (Table 3). However, effects on verification time, which we observed in Experiment 2, operate on a smaller time scale, with the critical question being whether they are large enough to drive corresponding effects on overall search time (Clayden et al., 2020). Therefore, we investigated subprocesses of search, with a focus on verification times (Figure 9c). Verification times were subjected to a mixed model using dummy coding and “static scenes” and “no scotoma” as reference levels. The simple effect for scene type was not significant, $b = -161.04$, $SE = 120.3$, $t = -1.34$. For static scenes, the difference in verification times between S-*Blindspots* and the control condition was not significant, $b = 212.47$, $SE = 141.08$, $t = 1.51$. However, the difference between M-*Blindspots* and the control condition was significant for the untransformed data, $b = 509.82$, $SE = 236.23$, $t = 2.16$, but not for the transformed data, $b = 0.02$, $SE = 0.02$, $t = 0.98$. The scene type × S–No Blindspot was not significant, $b = -211.2$, $SE = 176.43$, $t = -1.2$. The scene type × M–No Blindspot interaction was not significant for the untransformed data, $b = -369.53$, $SE = 218.63$, $t = -1.69$; for the transformed data, however, it was significant, $b = -0.07$, $SE = 0.03$, $t = -2.23$.

Dependent variable	Prameter	Intercept	Scene type	Scotoma size S–No	Scotoma size M–S	Scene type × Scotoma size S–No	Scene type × Scotoma size M–S	Random effects
Probability correct	*b*	1.35	0.16	−0.73	−1.21	−0.05	−0.59	2
	SE	0.23	0.16	0.21	0.18	0.42	0.36	
	z	5.95	1.01	−3.47	−6.7	−0.11	−1.66	
	p	<0.001	**0.312**	0.001	<0.001	**0.909**	**0.097**	
Search time	*b*	3,517.88	131.69	22.45	340.39	81.18	230.38	3
	SE	363.62	207.25	187.38	207.81	376.35	416.71	
	t	9.67	**0.64**	**0.12**	**1.64**	**0.22**	**0.55**	
Saccade amplitude	*b*	5.67	−0.28	0.94	0.81	0.13	−0.27	5
	SE	0.15	0.13	0.14	0.12	0.22	0.32	
	t	38.09	−2.09	6.92	6.63	**0.6**	**−0.83**	
Fixation duration	*b*	215.26	−3.55	4.5	2.38	5.62	0.46	4
	SE	6.68	2.46	5.04	3.07	5.78	6.29	
	t	32.24	**−1.44**	**0.89**	**0.78**	**0.97**	**0.07**	

Table 3. Linear and generalized linear mixed models (LLM and GLMM, respectively) for Experiment 3: means (*b*), standard errors (*SE*), and test statistics (LLMs: *t* values; GLMMs: *z* values and *p* values) for fixed effects. *Note:* Nonsignificant coefficients are set in bold (LLMs: $|t| < 1.96$; GLMMs: $p > .05$). See text for further details.

For completeness, Supplementary Figure S6 and Supplementary Table S2 additionally provide the results for search initiation and scanning times, using the default contrast coding. As in Experiment 2, search initiation times were significantly shorter for static as compared to dynamic scenes, $b = -79.64$, $SE = 7.72$, $t = -10.31$. For scanning times, there was no significant difference between dynamic and static scenes, $b = -129.15$, $SE = 229.32$, $t = -0.56$.

Effect of scene motion and visual complexity on search time

Different from Experiment 2, search times were not prolonged for dynamic scenes in any of the experimental conditions tested. Therefore, additional analyses were performed to explore the effects of visual complexity and motion on search times. For a given type of scene, data were collapsed across the blindspot and control conditions. Visual complexity and motion are continuous variables, which were z transformed for the LMM analyses. The LMMs were random intercept models.

The first model considered the data for static scenes and included visual complexity as fixed effect. The effect of complexity was significant, $b = 1{,}120.61$, $SE = 274.79$, $t = 4.08$, and its direction indicated that scenes with higher visual complexity were associated with longer search times.

The second model considered the data for dynamic scenes and included both motion and visual complexity as fixed effects. Individual scores were calculated for each trial to acknowledge that the video was only shown until the subject's button press. For each frame of the video the subject had seen, the optical flow magnitudes were summed up. To create the motion score, these values were added up and divided by the number of frames. Similarly, the visual complexity score was derived as the mean of the Feature Congestion clutter values for all video frames the subject had seen during the trial. (Of course, these values are highly correlated with the corresponding values for the static images.) There was no significant effect of motion on search times, $b = 58.02$, $SE = 369.32$, $t = 0.16$. However, there was a significant effect of visual complexity, $b = 1{,}004.10$, $SE = 412.30$, $t = 2.44$. To facilitate the interpretation of these results, we conducted the same analyses for the data from Experiment 2 and found no significant effects.

Saccade amplitudes and fixation durations

The main effect of scene type was significant, $b = -0.28$, $SE = 0.13$, $t = -2.09$; saccade amplitudes were shorter for static as compared to dynamic scenes (Figure 14a). Moreover, saccade amplitudes were significantly longer for S-*Blindspots* as compared to the control condition, $b = 0.94$, $SE = 0.14$, $t = 6.92$, and for M-*Blindspots* as compared to S-*Blindspots*, $b = 0.81$, $SE = 0.12$, $t = 6.63$. Thus, larger *Blindspots* were associated with progressively longer saccade amplitudes. There were no significant interactions between blindspot size and scene type (Table 3). In this experiment, neither scene type nor the presence of a *Blindspot*, nor blindspot size, had significant effects on fixation durations (Table 3, Figure 14b).

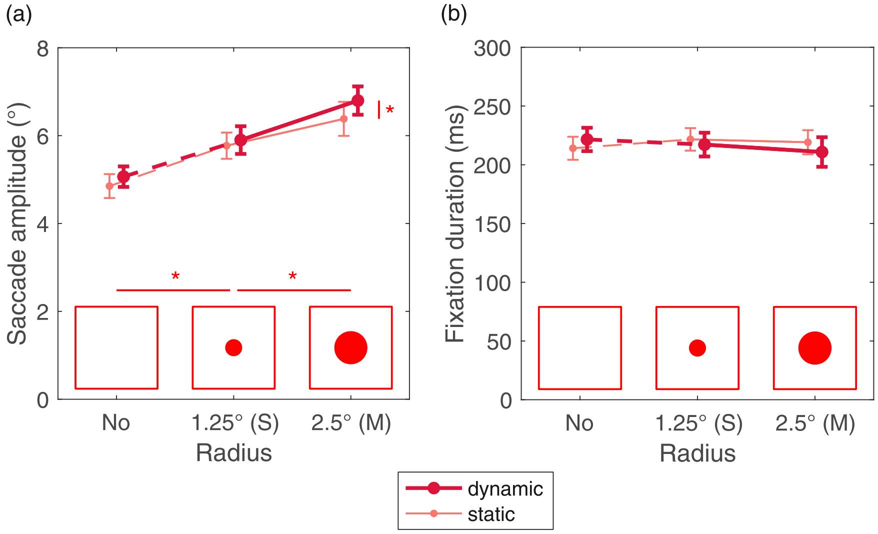


Figure 14. Mean saccade amplitudes (a) and fixation durations (b) for Experiment 3. Two different *Blindspot* sizes are compared with a no-scotoma control condition for both dynamic scenes (color: crimson, bold line, large marker) and static scenes (color: salmon, normal-weight line, normal-size marker). Error bars are 95% within-subjects confidence intervals. Lines with asterisks indicate a significant difference between condition means.

Discussion

In Experiment 3, there were more timed-out trials and longer search times for successful trials than in the corresponding conditions of Experiment 2. Saccade amplitudes were longer when searching dynamic as compared to static scenes (Smith & Mital, 2013). For fixation durations and search times, there were no differences between dynamic and static scenes. Taken together, the results suggest that the increased visual complexity of the scenes used in Experiment 3 made the target more difficult to acquire than in Experiment 2. Using scenes with a larger range of visual complexities, we also found an effect of visual complexity on search times, for both static and dynamic scenes (cf. Henderson et al., 2009, for static scenes). One interpretation of these results is that search in more complex scenes negated the differences between dynamic and static scenes that we had observed in Experiment 2.

General discussion

Using eye tracking, we combined two approaches to illuminate commonalities and differences in visual search for a static target embedded in complex dynamic and static real-world scenes. First, we used gaze-contingent *Blindspots* and *Spotlights* to investigate the importance of the different regions of the visual field to the search process (Nuthmann, 2014). Second, we used observers' gaze data to decompose their button-press search times into temporal epochs that are associated with particular subprocesses of search (Castelhano et al., 2008; Malcolm & Henderson, 2009).

The logic of the gaze-contingent scotoma manipulations is that information processing during the task will be disrupted to the extent that the missing information is needed for the task (Larson & Loschky, 2009; Nuthmann, 2014). For example, the extrafoveal scotoma (S-*Spotlight*) blocks out extrafoveal vision and keeps foveal vision intact, which means that we can assess observers' ability to do the task with foveal vision only. Conversely, the foveal scotoma (S-*Blindspot*) blocks out foveal vision and keeps extrafoveal vision intact, which means that we can assess observers' ability to do the task with extrafoveal vision only. Together, the S-*Spotlight* and the S-*Blindspot* inform us about the relative importance of foveal and extrafoveal vision to the task. Let us assume that search with an S-*Spotlight* leads to a drop in performance, whereas search with an S-*Blindspot* does not (Nuthmann, 2014). Such a pattern of results indicates that foveal vision is neither sufficient (*S-Spotlight*) nor necessary (*S-Blindspot*) for normal task performance. That is, extrafoveal vision is sufficient (*S-Blindspot*) and necessary (*S-Spotlight*). In summary, we would consider foveal vision to be unimportant and extrafoveal vision to be important to the task.

Under the assumption that motion attracts attention and gaze, we would expect search times to be longer for dynamic than for static scenes. In our natural vision control conditions, search times were prolonged for dynamic scenes in Experiment 2 but not in Experiment 3. In both experiments, search initiation times were elevated for dynamic scenes. Apparently, the potential or actual presence of motion in the dynamic scene increased the time needed to select a first search target candidate. Moreover, verification times were prolonged for dynamic scenes in Experiment 2 but not in Experiment 3. Given that the two experiments differed with regard to the materials, we suggest that scenes with greater visual complexity and more motion (Experiment 3) made the acquisition of the target equally difficult for dynamic and static scenes.

For our main Experiment 2, comparing the data from the no-scotoma control conditions to data obtained with different types of scotomas enabled us to explore the role motion played in foveal and central vision as opposed to extrafoveal and peripheral vision during different epochs of search. Our analyses yielded three key results. First, in conditions in which it was possible to process motion in extrafoveal and peripheral vision, observers directed their very first saccade to locations with higher motion values than when motion could not be preprocessed (Figure 11e,h). Thus, the first saccade was guided by motion in the scene (cf. Carmi & Itti, 2006, for free viewing). However, and secondly, this was not the case for subsequent saccades made during the scanning epoch, representing the actual search process (Figure 11f,i). These data suggest that task-irrelevant motion in extrafoveal and peripheral vision did not affect the guidance component of search, which is consistent with the notion that task demands can provide a cognitive override of stimulus-based salience (Einhäuser et al., 2008; Underwood et al., 2006). Still, finding a short-lived effect of scene motion has practical relevance in that losing the first saccade to a task-irrelevant moving element in the scene could be the difference between a safe and an unsafe traffic situation, for example. Third, when motion was potentially present (*Spotlights*) or absent (*Blindspots*) in foveal or central vision only, we observed differences in verification times for dynamic and static scenes (Figure 9); see below.

Previous research on visual search in static scenes suggested that foveal vision was surprisingly unimportant to the task. For object-in-scene search, S- and M-*Blindspots* neither affected search accuracy nor search time (Nuthmann, 2014). Only for L-*Blindspots*, degrading central vision, was there both a reduction in search accuracy as well as a significant increase in search times for correct trials due to prolonged verification times. Results were slightly different when the target was a contextually irrelevant black letter (T or L) of variable size. In this case, a foveal scotoma (circular with a radius of 1°) was associated with significantly increased verification times, which in turn boosted search times (Clayden et al., 2020). For letter search with an M-*Blindspot* (circle radius 2.5°, as in our Experiments 2 and 3), search times were significantly prolonged, which was (mainly) driven by an increase in verification times (Nuthmann et al., 2021).

Overall, the present findings for *static* scenes were consistent with these prior results. To guide search, participants could rely on a precise template of the target's shape, size, and color (grayish). S-*Blindspots* blocked foveal vision, whereas M-*Blindspots* blocked foveal and part of parafoveal vision. In Experiment 2, we observed search-time costs in these conditions, which originated from increased verification times. In Experiment 3, however, the presence of *Blindspots* was not associated with significant increases in search times.

For *dynamic* scenes, the search-time results for correct trials from three experiments converge on the novel finding that neither foveal vision nor central vision was necessary to attain normal search proficiency. Importantly, verification times were not increased during visual search with moving *Blindspots* in dynamic scenes. Participants knew that the target was static. Moreover, due to implicit learning, they were most likely aware that the target almost never (Experiments 1 and 2) or rarely (Experiment 3) interacted with moving elements in the scene. Therefore, we suggest that the low prevalence of motion in the target region helped to reduce the uncertainty about the presence of the target, thereby counteracting any increases in verification times, as observed for static scenes.

However, for both static and dynamic scenes, there were fewer hits during search with moving *Blindspots*, in particular for larger *Blindspots*. Instead, there were more timeouts, which indicates that visual search *can* be more difficult with a *Blindspot* than without (see also Clayden et al., 2020; Nuthmann et al., 2021). There were also more miss trials, but their interpretation is less straightforward due to the difficulty in distinguishing between incidents where the target was not located from trials where observers' eyes did not fixate on a correctly located target when the overt response was made (Clayden et al., 2020; Nuthmann, 2014).

Moreover, the results from Experiments 1 and 2 emphasize the importance of extrafoveal and peripheral to visual search, which is consistent with previous research using static scenes (Loschky & McConkie, 2002; Nuthmann, 2014). In Experiment 1, in which only dynamic scenes were used, search accuracy was only reduced when extrafoveal vision was masked (S-*Spotlights*). In Experiment 2, in which both dynamic and static scenes were used, we observed a small but signification reduction in search accuracy for M-*Spotlights* and a considerable decrease in search accuracy for S-*Spotlights*.

For correct trials, we observed massive search-time costs in both experiments. Specifically, any reduction in extrafoveal search space increased search times and systematically so as *Spotlights* became smaller. The gaze-based decomposition of search times revealed that search initiation times and scanning times also increased as *Spotlights* became smaller (cf. Nuthmann, 2014).

These results support the view that peripheral input provides information (a) for processing the gist of the scene and (b) for guiding the search process and eye-movement planning, thereby supporting core assumptions of a recent theory of visual information acquisition in tasks such as driving (B. Wolfe et al., 2020).

In the *Spotlight* conditions of Experiment 2, search accuracy was lower for dynamic than for static scenes, which was an unexpected result. For correct trials, both search initiation and verification times were prolonged for dynamic scenes, leading to numerically longer search times. These results are in agreement with data reported by Reingold and Loschky (2002), who found that degrading the scene outside a gaze-contingent moving window delayed target acquisition more strongly for dynamic than for static scenes.

Why, then, were verification times during search with a moving *Spotlight* longer for dynamic scenes as compared to static scenes (Figure 9a)? We speculate that the sudden appearance of motion within a *Spotlight* on some verification-epoch fixations may have made target verification more difficult. Interestingly, for the two types of scotomas, the verification-time differences between dynamic and static scenes had opposite signs, that is, for dynamic scenes, we observed longer verification times during *Spotlight* search but shorter verification times during *Blindspot* search (both in Experiments 2 and 3); see Figure 9. This dissociation warrants further research.

In previous research, both mean saccade amplitudes and fixation durations were found to be longer for dynamic as compared to static scenes, both for free-viewing and spot-the-location tasks (Smith & Mital, 2013). Using a target acquisition task, we observed longer saccade amplitudes for dynamic scenes in Experiment 3 but not in Experiment 2. In Experiment 3, mean fixation durations did not differ for dynamic and static scenes, whereas in Experiment 2, there was some evidence for longer fixation durations in dynamic scenes.

As a reasonable first step, we used nonmoving targets. It would of course be interesting for future studies to use moving targets (cf. Reingold & Loschky, 2002). Future studies should also consider using different types of targets and/or manipulate their properties. To reduce the variability in target salience between scenes, during stimulus generation, we positioned the target at an individually determined location in each scene. Still, a few targets were “invisible” to the observers, in particular for the London scenes, for which target color was not individually adjusted. It may be advantageous for future studies to better control for target salience by placing targets algorithmically, for example, by extending the Target Embedding Algorithm that we developed for static scenes (Clayden et al., 2020) to dynamic scenes.

Moreover, the present research could be productively extended by using objects that are naturally part of the scenes as search targets. Research on visual search has used a range of tasks, from looking for arbitrary targets within random arrays, through to searching for contextually relevant objects within naturalistic scenes. For basic laboratory search tasks, guidance by basic features like color and motion plays an important role in constraining the deployment of attention (J. M. Wolfe & Horowitz, 2017). During object-in-scene search, however, feature-based guidance is complemented or overridden by scene-based guidance (e.g., Underwood et al., 2006). According to the Surface Guidance Framework by Castelhano and colleagues, attention and gaze are directed to surfaces in the scene that are most associated with the target (Pereira & Castelhano, 2019). Moreover, Võ and colleagues have fleshed out the idea that scenes, like language, have a grammar, comprising both scene semantics and syntax (Võ et al., 2019). The notion that observers use various forms of scene guidance when searching naturalistic scenes has found empirical support (Castelhano & Krzyś, 2020; Võ, 2021, for reviews).

In our experiments, observers searched for a ring target that was superimposed on the scene stimulus. This situation bears similarities to real-world searches for which there is minimal or no guidance by scene context (e.g., search for a fly). But note that ring targets tend to violate the scene syntax, that is, the physical rules of the scene environment in which they appear (cf. Biederman et al., 1982). So far, few studies have investigated how scene-based guidance (beyond processing of the scene’s gist) interacts with information acquisition outside the fovea (see Pereira & Castelhano, 2014, for an exception). Future research on this topic may indicate whether the present results for dynamic versus static scenes generalize from arbitrary search targets to different types of contextually relevant search objects.

Conclusions

In the real world, we often search for a nonmoving object amid an environment that contains some moving elements. When having observers search for a static, contextually irrelevant target in videos depicting naturalistic scenes, we found that task-irrelevant motion in extrafoveal and peripheral vision had a fast but transient effect on saccade target selection. Regarding the importance of foveal, parafoveal, and peripheral vision to the search process, any differences found between dynamic and static scenes were relatively subtle. Moreover, the results of this multiexperiment study highlight the importance of using different sets of scenes to test the generalizability of results.

Keywords: visual search, eye movements, dynamic scenes, motion, moving windows, scotomas

Acknowledgments

The authors thank Lesley M. Deas for creating the videos recorded in Scotland and for collecting the data for Experiment 1 as part of her Masters project at the University of Edinburgh. We also thank Tim J. Smith for making the London videos available and for his helpful feedback on this research.

Commercial relationships: none.
Correspondence author: Antje Nuthmann.
Email: nuthmann@psychologie.uni-kiel.de.
Address: Kiel University, Institute of Psychology, Olshausenstr. 62, 24118 Kiel, Germany.

Footnotes

[1] First-pass gaze duration is defined as the sum of all fixations on the target during the first pass, that is, before the eyes left the target for the first time (Liversedge & Findlay, 2000).

[2] Strictly speaking, this only holds for saccade lengths that are larger than or equal to the sum of the patch radius and the scotoma radius. For shorter saccade lengths between fixations with a moving *Spotlight*, motion around fixation *n* partially overlaps with motion around fixation *n* – 1. For fixations with a moving *Blindspot*, shorter saccade lengths imply that not all motion within the patch centered on fixation *n* could be processed during fixation *n* – 1. We did not remove these cases, as this would considerably reduce the number of observations, in particular in the *Spotlight* conditions.

References

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science, 14*(5), 427–432, https://doi.org/10.1111/1467-9280.01458.

Abrams, R. A., & Christ, S. E. (2006). Motion onset captures attention: A rejoinder to Franconeri and Simons (2005). *Perception & Psychophysics, 68*(1), 114–117, https://doi.org/10.3758/BF03193661.

Açik, A., Bartel, A., & König, P. (2014). Real and implied motion at the center of gaze. *Journal of Vision, 14*(1), 2, https://doi.org/10.1167/14.1.2.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412, https://doi.org/10.1016/j.jml.2007.12.005.

Ball, K. K., Beard, B. L., Roenker, D. L., Miller, R. L., & Griggs, D. S. (1988). Age and visual search: Expanding the useful field of view. *Journal of the Optical Society of America A, 5*(12), 2210–2219, https://doi.org/10.1364/JOSAA.5.002210.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278, https://doi.org/10.1016/j.jml.2012.11.001.

Basler, A. (1906). Über das Sehen von Bewegungen. *Archiv für die gesamte Physiologie des Menschen und der Tiere, 115*(11), 582–601, https://doi.org/10.1007/BF01677292.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48, https://doi.org/10.18637/jss.v067.i01.

Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of the effective stimulus in visual search. *Perception & Psychophysics, 62*(3), 576–585, https://doi.org/10.3758/BF03212109.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*(2), 143–177, https://doi.org/10.1016/0010-0285(82)90007-X.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 26*(2), 211–252, https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

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

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the "*Spotlight*" reveals universal information use for face recognition. *PLoS ONE, 5*(3), e9708, https://doi.org/10.1371/journal.pone.0009708.

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research, 46*(26), 4333–4345, https://doi.org/10.1016/j.visres.2006.08.019.

Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception, & Psychophysics, 72*(5), 1283–1297, https://doi.org/10.3758/APP.72.5.1283.

Castelhano, M. S., & Krzyś, K. (2020). Rethinking space: A review of perception, attention, and memory in scene processing. *Annual Review of Vision Science, 6*, 563–586, https://doi.org/10.1146/annurev-vision-121219-081745.

Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review, 15*(4), 795–801, https://doi.org/10.3758/PBR.15.4.795.

Clayden, A. C., Fisher, R. B., & Nuthmann, A. (2020). On the relative (un)importance of foveal vision during letter search in naturalistic scenes. *Vision Research, 177*, 41–55, https://doi.org/10.1016/j.visres.2020.07.005.

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.

Cornelissen, T. H. W., & Võ, M. L.-H. (2017). Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception, & Psychophysics, 79*(1), 154–168, https://doi.org/10.3758/s13414-016-1203-7.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 42–45, https://doi.org/10.20982/tqmp.01.1.p042.

Cristino, F., & Baddeley, R. (2009). The nature of the visual representations involved in eye movements when walking down the street. *Visual Cognition, 17*(6–7), 880–903, https://doi.org/10.1080/13506280902834696.

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision, 10*(10), 28, https://doi.org/10.1167/10.10.28.

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision, 8*(2), 2, https://doi.org/10.1167/8.2.2.

Finlay, D. (1982). Motion perception in the peripheral visual field. *Perception, 11*(4), 457–462, https://doi.org/10.1068/p110457.

Foulsham, T., & Underwood, G. (2011). If visual saliency predicts search, then why? Evidence from normal and gaze-contingent search tasks in natural scenes. *Cognitive Computation, 3*(1), 48–63, https://doi.org/10.1007/s12559-010-9069-9.

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65*(7), 999–1010, https://doi.org/10.3758/BF03194829.

Franconeri, S. L., & Simons, D. J. (2005). The dynamic events that capture visual attention: A reply to Abrams and Christ (2005). *Perception & Psychophysics, 67*(6), 962–966, https://doi.org/10.3758/BF03193623.

Glaholt, M. G., Rayner, K., & Reingold, E. M. (2012). The mask-onset delay paradigm and the availability of central and peripheral visual information during scene viewing. *Journal of Vision, 12*(1), 9, https://doi.org/10.1167/12.1.9.

Goldstein, R. B., Woods, R. L., & Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place? *Computers in Biology and Medicine, 37*(7), 957–964, https://doi.org/10.1016/j.compbiomed.2006.08.018.

Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision, 9*(1), 32, https://doi.org/10.1167/9.1.32.

Hillstrom, A. P., & Yantis, S. (1994). Visual motion and attentional capture. *Perception & Psychophysics, 55*(4), 399–411, https://doi.org/10.3758/BF03205298.

Hinde, S. J., Smith, T. J., & Gilchrist, I. D. (2017). In search of oculomotor capture during film viewing: Implications for the balance of top-down and bottom-up control in the saccadic system. *Vision Research, 134*, 7–17, https://doi.org/10.1016/j.visres.2017.01.007.

Holmqvist, K., & Andersson, R. (2017). *Eye tracking: A comprehensive guide to methods, paradigms and measures*. Lund Eye-Tracking Research Institute.

Hulleman, J., & Olivers, C. N. L. (2017). The impending demise of the item in visual search. *Behavioral and Brain Sciences, 40*, e132, https://doi.org/10.1017/S0140525X15002794.

Hutson, J. P., Smith, T. J., Magliano, J. P., & Loschky, L. C. (2017). What is the role of the film viewer? The effects of narrative comprehension and viewing task on gaze control in film. *Cognitive Research: Principles and Implications, 2*(1), 46, https://doi.org/10.1186/s41235-017-0080-5.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–53). Elsevier, https://doi.org/10.1016/B978-008043361-5/50003-1.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12*(6), 1093–1123, https://doi.org/10.1080/13506280444000661.

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

Kunar, M. A., & Watson, D. G. (2014). When are abrupt onsets found efficiently in complex visual search? Evidence from multielement asynchronous dynamic search. *Journal of Experimental Psychology:*

Human Perception and Performance, 40(1), 232–252, https://doi.org/10.1037/a0033544.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26, https://doi.org/10.18637/jss.v082.i13.

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision, 9*(10), 6, https://doi.org/10.1167/9.10.6.

Larsson, L., Nyström, M., Ardö, H., Åstrom, K., & Stridh, M. (2016). Smooth pursuit detection in binocular eye-tracking data with automatic video-based performance evaluation. *Journal of Vision, 16*(15), 20, https://doi.org/10.1167/16.15.20.

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, https://doi.org/10.1167/13.12.11.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences, 4*(1), 6–14, https://doi.org/10.1016/S1364-6613(99)01418-7.

Loschky, L. C., Larson, A. M., Magliano, J. P., & Smith, T. J. (2015). What would Jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PLoS ONE, 10*(11), e0142474, https://doi.org/10.1371/journal.pone.0142474.

Loschky, L. C., & McConkie, G. W. (2002). Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied, 8*(2), 99–117, https://doi.org/10.1037/1076-898X.8.2.99.

Loschky, L. C., McConkie, G. W., Yang, H., & Miller, M. E. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition, 12*(6), 1057–1092, https://doi.org/10.1080/13506280444000652.

Loschky, L. C., Szaffarczyk, S., Beugnet, C., Young, M. E., & Boucart, M. (2019). The contributions of central and peripheral vision to scene-gist recognition with a 180° visual field. *Journal of Vision, 19*(5), 15, https://doi.org/10.1167/19.5.15.

Loschky, L. C., & Wolverton, G. S. (2007). How late can you update gaze-contingent multiresolutional displays without detection? *ACM Transactions on Multimedia Computing, Communications, and Applications, 3*(4), 25, https://doi.org/10.1145/1314303.1314310.

Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 674–679.

Mackworth, N. H. (1965). Visual noise causes tunnel vision. *Psychonomic Science, 3*, 67–68, https://doi.org/10.3758/BF03343023.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9*(11), 8, https://doi.org/10.1167/9.11.8.

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., & Hubel, D. H. (2009). Microsaccades: A neurophysiological analysis. *Trends in Neurosciences, 32*(9), 463–475, https://doi.org/10.1016/j.tins.2009.05.006.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315, https://doi.org/10.1016/j.jml.2017.01.001.

Mayer, K. M., Vuong, Q. C., & Thornton, I. M. (2015). Do people "pop out"? *PLoS ONE, 10*(10), e0139618, https://doi.org/10.1371/journal.pone.0139618.

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.

McIlreavy, L., Fiser, J., & Bex, P. J. (2012). Impact of simulated central scotomas on visual search in natural scenes. *Optometry and Vision Science, 89*(9), 1385–1394, https://doi.org/10.1097/OPX.0b013e318267a914.

Miellet, S., Zhou, X., He, L., Rodger, H., & Caldara, R. (2010). Investigating cultural diversity for extrafoveal information use in visual scenes. *Journal of Vision, 10*(6), 21, https://doi.org/10.1167/10.6.21.

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation, 3*(1), 5–24, https://doi.org/10.1007/s12559-010-9074-z.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4*(2), 61–64, https://doi.org/10.20982/tqmp.04.2.p061.

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., Clayden, A. C., & Fisher, R. B. (2021). The effect of target salience and size in visual search within naturalistic scenes under degraded vision. *Journal of Vision, 21*(4), 2, https://doi.org/10.1167/jov.21.4.2.

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, https://doi.org/10.1167/16.2.3.

Orban de Xivry, J.-J., & Lefèvre, P. (2007). Saccades and pursuit: two outcomes of a single sensorimotor process. *Journal of Physiology, 584*(1), 11–23, https://doi.org/10.1113/jphysiol.2007.139881.

Parkhurst, D., Culurciello, E., & Niebur, E. (2000). Evaluating variable resolution displays with visual search: Task performance and eye movements. In *Proceedings of the Eye Tracking Research & Applications symposium* (pp. 105–109). Association for Computing Machinery, https://doi.org/10.1145/355017.355033.

Pereira, E. J., & Castelhano, M. S. (2014). Peripheral guidance in scenes: The interaction of scene context and object content. *Journal of Experimental Psychology: Human Perception and Performance, 40*(5), 2056–2072, https://doi.org/10.1037/a0037524.

Pereira, E. J., & Castelhano, M. S. (2019). Attentional capture is contingent on scene region: Using surface guidance framework to explore attentional mechanisms during search. *Psychonomic Bulletin & Review, 26*(4), 1273–1281, https://doi.org/10.3758/s13423-019-01610-z.

Pinto, Y., Olivers, C. N. L., & Theeuwes, J. (2006). When is search for a static target among dynamic distractors efficient? *Journal of Experimental Psychology: Human Perception and Performance, 32*(1), 59–72, https://doi.org/10.1037/0096-1523.32.1.59.

Pinto, Y., Olivers, C. N. L., & Theeuwes, J. (2008). Static items are automatically prioritized in a dynamic environment. *Visual Cognition, 16*(7), 916–932, https://doi.org/10.1080/13506280701575375.

Post, R. B., & Johnson, C. A. (1986). Motion sensitivity in central and peripheral vision. *American Journal of Optometry and Physiological Optics, 63*(2), 104–107, https://doi.org/10.1097/00006324-198602000-00004.

Rayner, K. (2014). The gaze-contingent moving window in reading: Development and review. *Visual Cognition, 22*(3–4), 242–258, https://doi.org/10.1080/13506285.2013.879084.

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

Reder, S. M. (1973). On-line monitoring of eye-position signals in contingent and noncontingent paradigms. *Behavior Research Methods & Instrumentation, 5*(2), 218–228, https://doi.org/10.3758/BF03200168.

Reingold, E. M., & Loschky, L. C. (2002). Saliency of peripheral targets in gaze-contingent multiresolutional displays. *Behavior Research Methods, Instruments, & Computers, 34*(4), 491–499, https://doi.org/10.3758/BF03195478.

Reingold, E. M., Loschky, L. C., McConkie, G. W., & Stampe, D. M. (2003). Gaze-contingent multiresolutional displays: An integrative review. *Human Factors, 45*(2), 307–328, https://doi.org/10.1518/hfes.45.2.307.27235.

Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Review of Vision Science, 2*, 437–457, https://doi.org/10.1146/annurev-vision-082114-035733.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7*(2), 17, https://doi.org/10.1167/7.2.17.

Saida, S., & Ikeda, M. (1979). Useful visual field size for pattern perception. *Perception & Psychophysics, 25*(2), 119–125, https://doi.org/10.3758/BF03198797.

Sanders, A. F. (1970). Some aspects of the selective process in the functional visual field. *Ergonomics, 13*(1), 101–117, https://doi.org/10.1080/00140137008931124.

Saunders, D. R., & Woods, R. L. (2014). Direct measurement of the system latency of gaze-contingent displays. *Behavior Research Methods, 46*(2), 439–447, https://doi.org/10.3758/s13428-013-0375-5.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language, 110*, 104038, https://doi.org/10.1016/j.jml.2019.104038.

Seedorff, M., Oleson, J., & McMurray, B. (2019). Maybe maximal: Good enough mixed models optimize power while controlling Type I error. *PsyArXiv*, https://doi.org/10.31234/osf.io/xmhfr.

Shioiri, S., & Ikeda, M. (1989). Useful resolution for picture perception as a function of eccentricity. *Perception, 18*(3), 347–361, https://doi.org/10.1068/p180347.

Smith, T. J. (2013). Watching you watch movies: Using eye tracking to inform cognitive film theory. In A. P. Shimamura (Ed.), *Psychocinematics: Exploring cognition at the movies* (pp. 165–191). Oxford

University Press, https://doi.org/10.1093/acprof:oso/9780199862139.003.0009.

Smith, T. J., & Henderson, J. M. (2008). Edit Blindness: The relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research, 2*(2), 6, https://doi.org/10.16910/jemr.2.2.6.

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision, 13*(8), 16, https://doi.org/10.1167/13.8.16.

Startsev, M., Agtzidis, I., & Dorr, M. (2019). Characterizing and automatically detecting smooth pursuit in a large-scale ground-truth data set of dynamic natural scenes. *Journal of Vision, 19*(14), 10, https://doi.org/10.1167/19.14.10.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*(4), 766–786, https://doi.org/10.1037/0033-295X.113.4.766.

Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology, 18*(3), 321–342, https://doi.org/10.1080/09541440500236661.

Valuch, C., & Ansorge, U. (2015). The influence of color during continuity cuts in edited movies: An eye-tracking study. *Multimedia Tools and Applications, 74*(22), 10161–10176, https://doi.org/10.1007/s11042-015-2806-z.

van Diepen, P. M. J., De Graef, P., & Van Rensbergen, J. (1994). On-line control of moving masks and windows on a complex background using the ATVista videographics adapter. *Behavior Research Methods, Instruments, & Computers, 26*(4), 454–460, https://doi.org/10.3758/BF03204665.

van Diepen, P. M. J., Wampers, M., & d'Ydewalle, G. (1998). Functional division of the visual field: Moving masks and moving windows. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 337–355). Elsevier, https://doi.org/10.1016/B978-008043361-5/50016-X.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer, https://doi.org/10.1007/978-0-387-21706-2.

Vig, E., Dorr, M., Martinetz, T., & Barth, E. (2011). Eye movements show optimal average anticipation with natural dynamic scenes. *Cognitive Computation, 3*(1), 79–88, https://doi.org/10.1007/s12559-010-9061-4.

Võ, M. L.-H. (2021). The meaning and structure of scenes. *Vision Research, 181*, 10–20, https://doi.org/10.1016/j.visres.2020.11.003.

Võ, M. L.-H., Boettcher, S. E. P., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology, 29*, 205–210, https://doi.org/10.1016/j.copsyc.2019.03.009.

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

Wolfe, B., Dobres, J., Rosenholtz, R., & Reimer, B. (2017). More than the Useful Field: Considering peripheral vision in driving. *Applied Ergonomics, 65*, 316–325, https://doi.org/10.1016/j.apergo.2017.07.009.

Wolfe, B., Sawyer, B. D., & Rosenholtz, R. (2020). Toward a theory of visual information acquisition in driving. *Human Factors*. Advance online publication, https://doi.org/10.1177/0018720820939693.

Wolfe, J. M. (2015). Visual search. In A. Kingstone, J. M. Fawcett, & E. F. Risko (Eds.), *The handbook of attention* (pp. 27–56). MIT CogNet.

Wolfe, J. M. (2020). Forty years after feature integration theory: An introduction to the special issue in honor of the contributions of Anne Treisman. *Attention, Perception, & Psychophysics, 82*(1), 1–6, https://doi.org/10.3758/s13414-019-01966-3.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour, 1*, 0058, https://doi.org/10.1038/s41562-017-0058.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115*(4), 787–835, https://doi.org/10.1037/a0013118.

Zhaoping, L. (2019). A new framework for understanding vision from the perspective of the primary visual cortex. *Current Opinion in Neurobiology, 58*, 1–10, https://doi.org/10.1016/j.conb.2019.06.001.