THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The effect of target salience and size in visual search within naturalistic scenes under degraded vision

OPEN ACCESS

The effect of target salience and size in visual search within naturalistic scenes under degraded vision

Antje Nuthmann

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

Adam C. Clayden

School of Engineering, Arts, Science and Technology,
University of Suffolk, UK
Psychology Department, School of Philosophy,
Psychology and Language Sciences, University of
Edinburgh, UK

Robert B. Fisher

School of Informatics, University of Edinburgh, UK

We address two questions concerning eye guidance during visual search in naturalistic scenes. First, search has been described as a task in which visual salience is unimportant. Here, we revisit this question by using a letter-in-scene search task that minimizes any confounding effects that may arise from scene guidance. Second, we investigate how important the different regions of the visual field are for different subprocesses of search (target localization, verification). In Experiment 1, we manipulated both the salience (low vs. high) and the size (small vs. large) of the target letter (a "T"), and we implemented a foveal scotoma (radius: 1°) in half of the trials. In Experiment 2, observers searched for high- and low-salience targets either with full vision or with a central or peripheral scotoma (radius: 2.5°). In both experiments, we found main effects of salience with better performance for high-salience targets. In Experiment 1, search was faster for large than for small targets, and high-salience helped more for small targets. When searching with a foveal scotoma, performance was relatively unimpaired regardless of the target's salience and size. In Experiment 2, both visual-field manipulations led to search time costs, but the peripheral scotoma was much more detrimental than the central scotoma. Peripheral vision proved to be important for target localization, and central vision for target verification. Salience affected eye movement guidance to the target in both central and peripheral vision. Collectively, the results lend support for search models that incorporate salience for predicting eye-movement behavior.

Introduction

In search for a specific target object in a naturalistic scene, we use selective attention to deploy our limited attentional resources, as well as our eyes, to candidate targets. This deployment is guided by knowledge of the basic features of the target and, when possible, by the rules that govern the placement of that target in a scene (Wolfe, 2015). Here, we investigate the causal influence of bottom-up visual salience on gaze guidance during scene search. To this end, we manipulate the salience and size of context-free targets within scenes. Moreover, we explore the importance of foveal vision (Experiment 1) and central versus peripheral vision (Experiment 2) for the task. We found that search was more efficient for high-salience than for low-salience targets. Salience affected eye movement guidance to the target in both central and peripheral vision.

It is widely agreed that eye movements in naturalistic scenes are controlled by both bottom-up (stimulus-driven) and top-down (task-driven, context-driven, or goal-driven) factors (Malcolm, Groen, & Baker, 2016). Research on bottom-up control has been dominated by salience-driven approaches, in which a saliency map is computed using low-level image features to guide task independent gaze allocation (Borji & Itti, 2013; Borji, Sihite, & Itti, 2013a for reviews). The first computational model of this kind was Itti, Koch, and Niebur's (1998) implementation of Koch and Ullman's (1985) computational architecture based on the feature integration theory (Treisman & Gelade,

1980). The feature integration theory explains human behavior in visual search tasks involving covert shifts of attention. Extending this research, the saliency model was introduced as a model of covert and overt orienting in search (Itti & Koch, 2000; Itti et al., 1998). According to simulations by Itti and Koch (2000), the saliency model performed similarly to, or better than, human searchers looking for oriented lines among distractor lines or for a camouflaged tank in a natural environment. Still, when observers are given a visual search task (or a task altogether), top-down influences on attention and eye guidance are often believed to dominate (Koehler, Guo, Zhang, & Eckstein, 2014).

Few empirical studies have investigated the role of target salience in search within natural scenes. Whereas some studies manipulated the salience of the target object (Foulsham & Underwood, 2007; Underwood, Templeman, Lamming, & Foulsham, 2008), others used low-salience targets that were presented along with high-salience distractors (Henderson, Malcolm, & Schandl, 2009; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006) or distractors that were either high or low in salience (Underwood & Foulsham, 2006).

One of these search tasks required observers to indicate whether there was a piece of fruit in the scene (Underwood et al., 2006). If present, the piece of fruit was always a low-salience object, according to the saliency model by Itti and Koch (2000). Some of the scenes also included a high-salience object, which served as a distractor. There was little attentional capture by the distractor. However, when there was a high-salience distractor present, then the low-salience target was fixated later than when it was absent, and near distractors were more disruptive than those furthest from the target. The authors concluded that the purpose of inspection can provide a cognitive override that renders visual salience secondary. The key finding that the most salient region is neglected in favor of a completely nonsalient target was replicated in a subsequent study by different authors (Henderson et al., 2009).

Underwood and Foulsham (2006) had subjects search for a small gray rubber ball, which was inserted into half of the scenes. This target was of very low visual salience. Beyond that, the visual salience and semantic congruency of two nontarget objects were manipulated. The authors summarized that search was unaffected by salience or congruency. On closer inspection, the data showed an unexpected interaction. When both nontarget objects were congruent with the overall meaning of the scene, fixation of the more salient of them was slow, rather than fast. Presumably, the inspection of a bright object had low priority when the task required the detection of a small dark target (Underwood & Foulsham, 2006).

Foulsham and Underwood (2007) manipulated the visual salience of the target directly by comparing medium- and low-salience target objects; objects were again chosen on the basis of their saliency model ranks. The authors excluded high-salience targets based on the argument that natural search is often performed in situations where the target is not the most salient object. There was little evidence that visual salience was important in eye guidance during either category or instance search. Underwood et al. (2008) used a comparative visual search task in which target objects were manipulated regarding their visual conspicuity (i.e., salience) and semantic congruency. Manual reaction times and eye movement guidance to the target were not affected by visual salience.

Foulsham and Underwood (2011) used a slightly different approach: rather than manipulating scenes and objects, they used the predictions of the saliency model by Itti and Koch (2000) to select target regions that were either salient or nonsalient. As would be predicted by the saliency model, behavioral search times were shorter for highly salient regions than either low-salient regions or control regions. Control regions and low-salient regions did not differ reliably. Interestingly, salience did not affect the process of localizing the target region in space, as indexed by the latency to first fixation on the region. This implies that the subsequent verification process (is this the target?) took longer when the region was low in salience and that this effect was large enough to affect total search time. In a second experiment, peripheral filtering of low-level features was expected to modify the effect of target saliency on search, but this was not the case (Foulsham & Underwood, 2011).

The main problem with identifying the causal contribution of visual salience to gaze guidance is an inherent correlation with higher-order factors such as objects and semantics (Henderson, Brockmole, Castelhano, & Mack, 2007; Nuthmann & Henderson, 2010; Stoll, Thrun, Nuthmann, & Einhäuser, 2015). In the studies reviewed above, effects of salience were assessed between different objects or scenes, which potentially introduces additional confounds. To address these issues, we used context-free letter targets rather than contextually relevant search targets. In two experiments, observers searched for a black letter "T" embedded in grayscale photographs of real-world scenes. We used our target embedding algorithm (T.E.A., Clayden, Fisher, & Nuthmann, 2020)[1] to generate within-scene manipulations of target salience (low vs. high) and—in Experiment 1—also target size (small vs. large). Our approach minimizes any confounding effects that may arise from various forms of scene guidance (semantic, syntactic, and episodic guidance; Biederman, Mezzanotte, & Rabinowitz, 1982; Henderson & Ferreira, 2004). Specifically, using context-free targets prevents observers from using their knowledge about the likely positions of targets to

guide their attention and eye movements. Moreover, by inserting the targets in an algorithmic manner via image processing techniques, we also minimized artefacts that may otherwise occur due to post hoc editing of scenes.

Saliency maps translate physical properties of the stimulus such as luminance, orientation, color, and size into saliency values. Since these stimulus dimensions have different characteristics, combining them is a nontrivial problem (Itti & Koch, 1999). The size feature is typically accounted for in an implicit manner by incorporating multiple spatial scales of processing. In this way, saliency models attempt to account for size over image regions and not over objects, which is a limiting factor of this approach (Borji, Sihite, & Itti, 2013b). Borji et al. (2013b) addressed this issue by asking observers to choose which object (out of two in a given image) stands out the most based on its low-level features. Both saliency and object size were important for selecting the object. Observers' judgments were well described by a linear combination of the two variables in an integrated model of saliency and object size. Moreover, previous investigations of object-based selection in scenes found independent effects of object size and object-based salience on fixation probability, with large objects and highly salient objects being more frequently selected for fixation (Nuthmann, Schütz, & Einhäuser, 2020; Stoll et al., 2015). Regarding visual search, in previous work we manipulated target size while controlling for target salience by probing the scene for locations of median salience (Clayden et al., 2020). In these experiments, we observed better search performance for larger targets. Extending this research, we designed Experiment 1 to assess the independent contributions of target salience and target size, as well as their interaction.

If our vision was the same throughout the visual field, visual search would be easy most of the time. However, foveal and extrafoveal vision differ, owing to our foveated visual systems (Rosenholtz, 2016). Saliency models, as well as theories of search, oftentimes ignore that visual acuity declines systematically from the fovea into the periphery. Of course, there are notable exceptions. For example, Itti (2006) added a gaze-contingent foveation filter to a variant of the saliency model, and the target acquisition model (Zelinsky, 2008), as well as the MASC model (Adeli, Vitu, & Zelinsky, 2017) implement a fixation-by-fixation retina transformation of the search image. Previous research has shown that foveal vision is less important and peripheral vision is more important for scene search than previously thought (Clayden et al., 2020; McIlreavy, Fiser, & Bex, 2012; Nuthmann, 2014). Here, we extend this research by assessing the role target salience plays in foveal vision (Experiment 1) and central versus peripheral vision (Experiment 2).

In visual search, guidance by basic features can be bottom-up or top-down (Wolfe, 2015). Bottom-up guidance is stimulus-driven, based on local differences. Here, we tested the independent and combined effects of target salience and size during active eye-movement search. Top-down guidance is user-driven, based on the observer's understanding of the task. In our experiments, on each trial participants were asked to look for the letter "T." Given that letters are overlearned categories, observers were expected to use top-down guidance to deploy attention to the target.

Any model where salience is combined with target knowledge would predict that search should be more efficient for high-salience than for low-salience targets. Clearly, results from most of the studies reviewed above did not lend support to this hypothesis. Here, we revisit the question by using a task that emphasizes feature guidance and minimizes the role of scene guidance. Moreover, Experiment 1 allowed us to assess the independent effects of target salience and size.

In our experiments, search with normal, nondegraded vision was compared to search with a foveal scotoma (radius: 1°) in Experiment 1, and to central and peripheral scotomas (radius: 2.5°) in Experiment 2. When searching with a foveal scotoma, we have found performance to be relatively unimpaired regardless of the target's size (Clayden et al., 2020). In Experiment 1, we explored whether foveal vision would gain a more prominent role if the target's salience was reduced, along with its size. In Experiment 2, we expect the peripheral scotoma to be more detrimental than the central scotoma (cf. Nuthmann, 2014). Analyzing subprocesses of search will allow us to test the assumption of a central-peripheral dichotomy according to which central vision is mainly for seeing (decoding or recognizing) and peripheral vision is mainly for looking (selecting) (Zhaoping, 2019). Applied to the target acquisition task that we used, we should find peripheral vision to be important for target localization and central vision for verification. Thus we expect the peripheral scotoma to selectively impair target localization and the central scotoma to impair target verification only (cf. Nuthmann, 2014). Beyond that, the simulated scotomas allow us to assess the effect of target salience in peripheral and central vision.

General method

Participants

Thirty-two participants (10 males) between the ages of 18 and 27 (mean age 21 years) participated in Experiment 1. Thirty-six participants (seven males) between the ages of 18 and 27 (mean age 21 years) participated in Experiment 2. All participants had normal or corrected-to-normal vision by self-report. They gave their written consent before the experiment

and either received study credit or were paid at a rate of £7 per hour for their participation, which lasted about one hour. The experiments were approved by the Psychology Research Ethics Committee of the University of Edinburgh and conformed to the Declaration of Helsinki.

Apparatus

Working with gaze-contingent displays requires minimizing the latency of the system (Loschky & Wolverton, 2007; Saunders & Woods, 2014). Moreover, gaze-contingent manipulations of foveal vision call for eye-tracking equipment with high spatial accuracy and precision (Geringswald, Baumgartner, & Pollmann, 2013). Participants' eye movements were recorded binocularly with an EyeLink 1000 Desktop mount system (SR Research, Ottawa, ON, Canada) with high accuracy (0.15° best, 0.25° to 0.5° typical) and high precision (0.01° root-mean-square [RMS]). The Eyelink 1000 was equipped with the 2000 Hz camera upgrade, allowing for binocular recordings at a sampling rate of 1000 Hz per eye. Stimuli were presented on a 21-inch CRT monitor with a refresh rate of 140 Hz at a viewing distance of 90 cm, 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.

The experiments were programmed in MATLAB 2013a (The MathWorks, Natick, MA, USA) using the OpenGL-based Psychophysics Toolbox 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007), which incorporates the EyeLink Toolbox extensions (F. W. Cornelissen, Peters, & Palmer, 2002). A game controller was used to record participants' behavioral responses.

Stimulus materials

In both experiments, we used 120 grayscale images of naturalistic scenes (800 × 600 pixels), which came from a variety of categories; 98 of these photographs were previously used as colored images in Nuthmann (2014). Additional images were used as practice scenes.

The search target was always the letter "T," which was inserted into the scene by using the T.E.A. introduced by Clayden et al. (2020). Specifically, the T was inserted in sans-serif style; that is, consisting of two bars. For
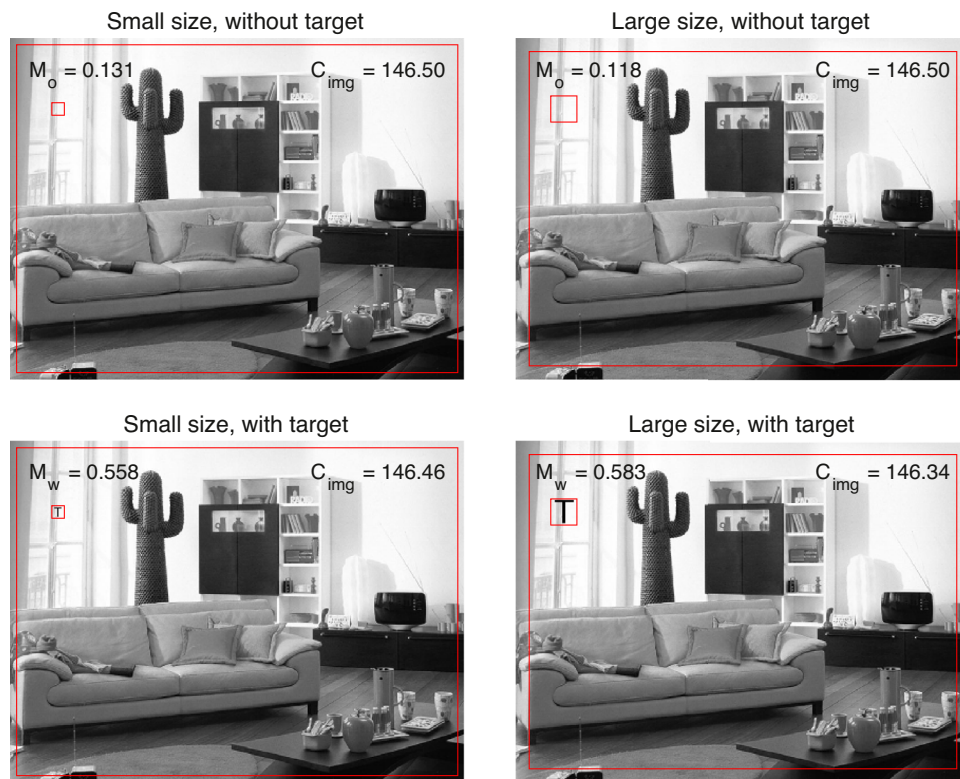
Figure 1. Target embedding algorithm. In this example, the squared evaluation box (in red) is positioned at $(r,c) = (125, 85)$ in all panels. The local RMS contrast is calculated both without the target letter (M_o, top row) and with the target letter inserted (M_w, bottom row), for both the small target (left column) and the large target (right column). C_{img} denotes the mean luminance of the image, without the target letter (top row) or with (bottom row). The outer rectangle (in red) marks the region of the image border that was not considered for target insertion.
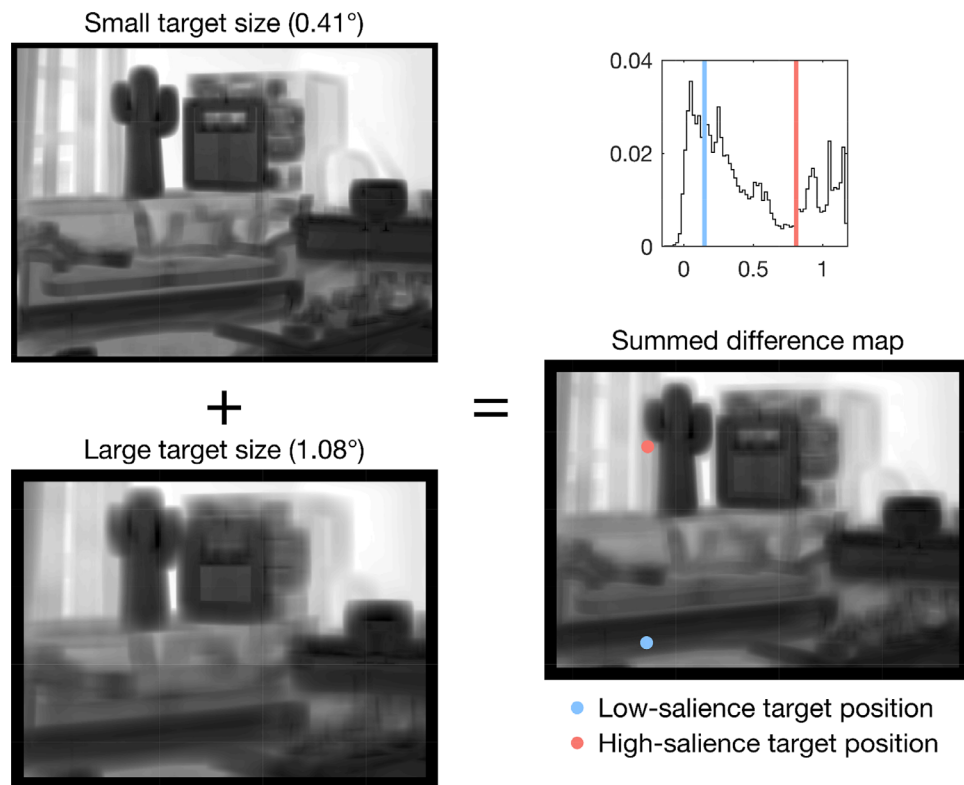
Figure 2. Algorithmic target placement at low and high salient regions within the scene. Left: Contrast difference maps for the small and large target size used in Experiment 1, for the example scene used in Figure 1 and Figure 7. Right: Summed contrast difference map (bottom) and the distribution of the map's values (top). Two vertical lines were added to the histogram to mark the lower (light blue) and upper (salmon) quartiles of the distribution. These values were used to determine suitable positions for low- and high-salience targets in the scene image. For the example image, the final target positions are marked with colored dots in the summed contrast difference map. For visualization purposes, the values of a given map were scaled to the same range (i.e., to [0,1]).

the small target letter, the horizontal bar was 13 pixels in length and two pixels in width, whereas the vertical bar was 16 pixels in length and three pixels in width. For the large target letter, the horizontal bar was 33 pixels in length and four pixels in width, whereas the vertical bar was 40 pixels in length and five pixels in width.

To determine suitable positions for low- and high-salience targets, we inserted the T into every possible location of the original scene image and calculated how much it would stand out from the scene background. To this end, a rectangular region that was slightly larger than the target was moved pixel-by-pixel through the image. Using the larger dimension of the target letter (i.e., its height) as a reference, the region's size was determined by adding a constant buffer of three pixels to either side (plus one pixel to center the region on the current position). As a result, the region size was 23 × 23 pixels for small target letters and 47 × 47 pixels for large target letters.

As a measure of visual salience, we used a version of RMS contrast: the standard deviation of luminance values of all pixels in the evaluated region was divided by the mean luminance of the image (Bex & Makous,

2002; Nuthmann & Einhäuser, 2015; Reinagel & Zador, 1999). First, the RMS contrast M_o was calculated for the evaluation box at each position in the image, see Appendix A for the mathematical details of the calculations. Next, the black target letter was inserted at a given position by replacing pixel values of the original image by the pixel values of the target. After target insertion, the RMS contrast M_w was computed for the evaluation box comprising the T. Afterward, the contrast change value $\Delta C = M_w - M_o$ was computed to quantify the visual salience of the target letter at a given location within the scene.

To provide an example, in Figure 1 the evaluation box is centered on image position $(r,c) = (125, 85)$, with (r,c) denoting the rows and columns of the image. For the large target, we obtain $M_o = 0.118$ and $M_w = 0.583$, with $\Delta C = 0.464$. Thus adding a black T to a relatively bright region of the image leads to a relatively large change in local contrast. For the example image used in Figures 1 and 2, our GitHub page (see footnote 1) shows a dynamic visualization of the contrast calculations for all possible target positions.

Calculating ΔC at each pixel in the image yields a map comprising the contrast difference values within the image. The contrast difference map was calculated separately for small and large targets (Figure 2). Afterward, the two resultant maps were summed together. This allowed us to compute a single location for both target sizes, because the values of the two difference maps varied slightly. The summed difference map was then probed by our algorithm to locate suitable pixel (i.e., potential target) positions. The criteria for choosing the low- and high-salience regions were the lower and upper quartile changes in local contrast when inserting the letter into the scene. If the exact value for the lower or upper quartile of the distribution did not exist in the summed contrast difference map, the closest existing value was used. Candidate locations were tested against two exclusion criteria (Clayden et al., 2020). In the experiments, participants started their search at the center of the scene, with a foveal or central scotoma blocking their view on many of the trials. Therefore locations within 3° from the center were excluded. To avoid truncation of the letter, locations at image boundaries were also excluded (Figures 1 and 2). If there was more than one possible target location left, one was selected at random as the location of the target for that scene image and salience condition. The resulting distributions of target positions reveal broad coverage (Figure B1). For further validation, each target's eccentricity was calculated as the Euclidean distance between target position and image center. Mean eccentricities did not differ for low- and high-salience targets, $t(119) = -0.32$, $p = 0.746$.

Creation of gaze-contingent scotomas

In Experiment 1, we implemented a *foveal* scotoma; in Experiment 2, we contrasted a *central* scotoma with a *peripheral* scotoma. For the foveal and central scotomas, we used a gaze-contingent technique that was originally introduced by Rayner and Bertera (1979) for sentence reading. The authors referred to their implementation as *moving mask*; other terms include *simulated scotoma* (Bertera, 1988). 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, whereas information outside the window is unaltered (Miellet, Zhou, He, Rodger, & Caldara, 2010; Nuthmann, 2014). As in our previous study (Clayden et al., 2020), the foveal scotoma in Experiment 1 was a symmetric circular gray mask with a radius of 1° to completely obscure foveal vision (see Figure 3 below). The central scotoma (Experiment 2) had a radius of 2.5°, thereby eliminating both foveal and part of parafoveal vision (Figure 7b below). For the peripheral scotoma (Experiment 2), 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, Zhou, & Miellet, 2010; Nuthmann, 2014). Our central and peripheral scotomas had equal radii (2.5°) and so were inverse manipulations of one another (Figure 7 below).

The general idea underlying our scotoma implementation is to mix a foreground image and a background image via a mask image (van Diepen, De Graef, & Van Rensbergen, 1994). The foreground image is formed by the experimental stimulus; that is, by the current scene image. 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). The mask image defines the type, shape, and size of the gaze-contingent scotoma. It was a normalized grayscale image, where pixel values of 255 (white) represent portions of the foreground image that show through while values of 0 (black) are masked and therefore replaced by the corresponding background image pixels. For the foveal and the central scotoma, a circular 0-center, 255-surround map formed the mask. For the peripheral scotoma, an inverted mask was used; that is, a circular 255-center, 0-surround map. To avoid sharp-boundary scotomas, the perimeter of the circular mask or window was slightly faded through low-pass filtering (Clayden et al., 2020).

To minimize the latency of the measurement system, we used an eye tracker with a binocular sampling rate of 1000 Hz and fast online access of new gaze samples. Specifically, the eye tracker computed a new gaze position every millisecond and made it available in less than 2 ms. Moreover, the Psychophysics Toolbox 3 for MATLAB offers fast creation of gaze-contingent scotomas using texture-mapping and OpenGL (Open Graphics Library). This technique provides various blending operations that enable image combinations to take place via an image's alpha channel (see Duchowski & Çöltekin, 2007, for details on the general technique). The mask image served as the alpha mask for blending of the foreground and background images. To obtain a composite rendering of the scene image with the scotoma, three textures were created—for the foreground image, background image, and mask image, respectively. During the search trial, the center of the mask texture was translated to the coordinates of the current gaze position. Thus, gaze contingency was realized by moving the mask across the stimulus. This solution avoids the need for computationally expensive real-time image synthesis.

Because scene images typically occupy the entire monitor space, a full refresh cycle is required to
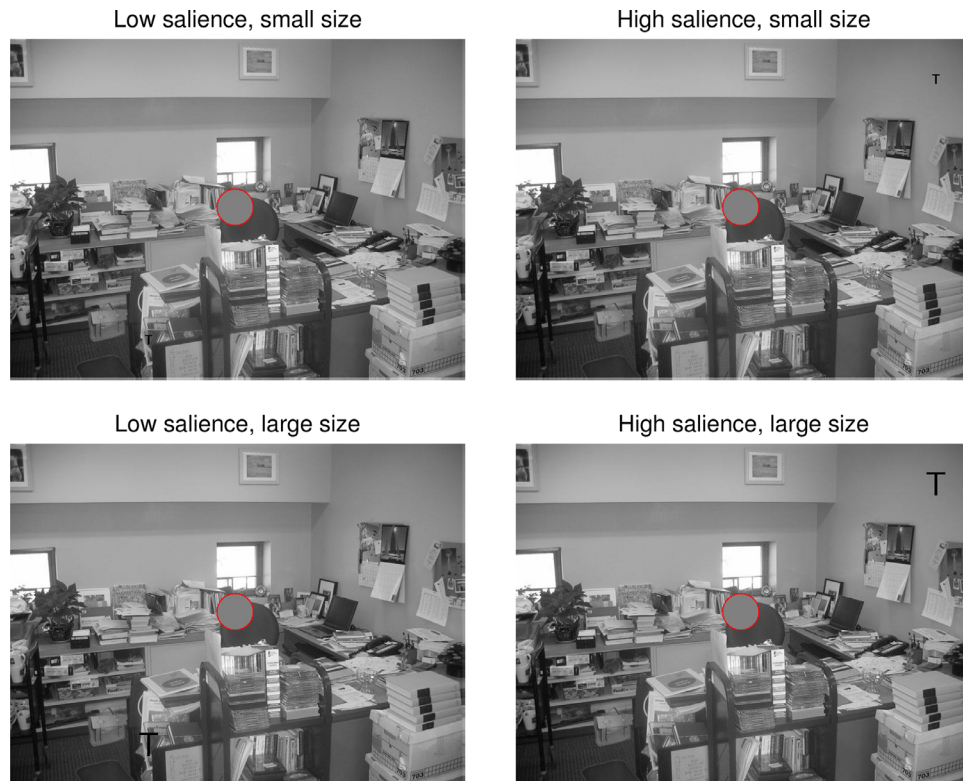
Figure 3. Four foveal-scotoma conditions for one of the scenes used in Experiment 1. Left column: low-salience targets, right column: high-salience targets; top row: small targets, bottom row: large targets. The gray disk in the center of the image is the foveal mask that moved concomitantly with the participant's gaze. In the figure, the foveal scotoma is highlighted with a red circle. In the experiment, each observer searched each scene in one of the size × salience conditions only, either with or without a simulated foveal scotoma.

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 re-aligned with the average horizontal and vertical position of the two eyes (Nuthmann, 2013, for discussion). Even with a state-of-the-art system, small temporal delays in updating the display contingent on the participant's gaze are unavoidable. Any mismatch between gaze position and scotoma position that may result should be largest during a saccade and right after a saccade. However, observers are blind to mismatches during this period, due to saccadic suppression and the time needed for perception to be restored (McConkie & Loschky, 2002).

Procedure

At the beginning of the experiment, the eye tracker was calibrated using a series of nine fixed targets distributed around the display, followed by a 9-point accuracy test. At the start of each trial, a fixation cross was presented at the center of the screen for 600 ms and acted as a fixation check. The fixation check was judged successful if gaze position, averaged across both eyes, consistently remained within an area of 40 × 40 pixels (1.24° × 1.24°) 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 appeared on the screen. Once subjects had found the target letter, 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, Rayner, & Reingold, 2012; Nuthmann, 2014). Trials timed-out 15 seconds after stimulus presentation if no response was made. There was an intertrial interval of one second before the next fixation cross was presented.

Data analysis

The SR Research Data Viewer software with default settings was used to convert the raw data obtained by the eye tracker into a fixation sequence matrix. Data from the right eye were analyzed. The behavioral and eye-movement data were further processed and

analyzed using the R system for statistical computing (R Development Core Team). Figures were created using MATLAB (Figures 1 to 3 and 7) or the *ggplot2* package (version 3.2.1; Wickham, 2016) supplied in R (remaining figures). The T.E.A. was programmed in MATLAB.

Analyses of fixation durations and saccade lengths excluded fixations that were interrupted with blinks. Analysis of fixation durations disregarded the initial, central fixation in a trial. However, its duration was analyzed separately as search initiation time. The button press terminating the search took place during the last fixation in a trial. Therefore, the last fixation was also excluded from analysis of fixation durations. However, its duration contributed to the measurement of verification time. 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). In the present study, this precaution was not followed because the presence of a foveal 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, the optimal λ-coefficient for the Box-Cox power transformation (Box & Cox, 1964) was estimated 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

We used linear mixed-effects models (LMM) for analyzing continuous response variables, specifically search time and its three subcomponents, saccade amplitude, and fixation duration. Search accuracy was analyzed using binomial generalized linear mixed-effects models (GLMM). A technical introduction to both types of mixed models is provided by Demidenko (2013). The analyses were conducted with the R package *lme4* (version 1.1.-23; Bates, Maechler, Bolker, & Walker, 2015). Separate (G)LMMs were estimated for each dependent variable.

Search accuracy was assessed through a binary variable; in a given trial, the search target was correctly located (1) or not (0). In the GLMM, the resulting probabilities were modeled through a link function (Bolker et al., 2009). For binary data, there are three common choices for link functions: logit, probit, and complementary log-log (Demidenko, 2013). For our analyses we used the logit transformation of the probability, which is the default for the *glmer* function in the R package *lme4*. Thus, in a binomial GLMM, parameter estimates are obtained on the log-odds or logit scale, which is symmetric around zero, corresponding to a probability of 0.5, and ranges from negative to positive infinity (Jaeger, 2008).

A mixed-effects model contains both fixed-effects and random-effects terms (Bates et al., 2015). Because mixed models are regression techniques, factors of the experimental design usually enter the model as contrasts (Schad, Vasishth, Hohenstein, & Kliegl, 2020). For Experiment 1, to specify the contrasts simple coding (also known as deviation coding or effects coding) was used for all three factors of the experimental design ($-0.5/+0.5$). The reference levels were small size, low-salience, and no scotoma. The mixed-model equation is provided in Appendix C.

For Experiment 2, simple coding was used for the two-level factor target salience. For the three-level factor scotoma type, contrasts were chosen such that they tested hypotheses about the expected pattern of means. More generally, the different scotomas were expected to affect overall task difficulty, which may lead to differences in search performance and global eye movement measures. For example, search times were expected to be longest for search with a peripheral scotoma. In this case, factor levels were ordered accordingly (no scotoma, central scotoma, peripheral scotoma), and 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). Moreover, we reasoned that a specific type of scotoma may selectively impair a specific subprocess of search. To test these more specific hypotheses, 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 the central scotoma and the control condition or between the peripheral scotoma and the control condition. Simple coding and backward difference coding yield centered contrasts, in which case the model intercept reflects the grand mean of the dependent variable.

The mixed models included subjects and scene items 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 (Barr, Levy, Scheepers, & Tily, 2013). Random slopes estimate the degree to which each fixed effect varies across subjects and/or scene items. For example, the by-item random slope for salience captures whether scene items vary in the extent to which target salience affects search performance or eye-movement parameters (see Nuthmann, Einhäuser, & Schütz, 2017, for an example).

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). For Experiment 1, where the same contrast coding was used for all dependent variables, the maximal variance-covariance matrix of the random effects is provided in Appendix C. Across experiments, none of these maximal models converged (maximal number of iterations: 10^6). For LMMs, the maximal random-effects structure was backward-reduced using the *step* function of the R package *lmerTest* (version 3.1-2; Kuznetsova, Brockhoff, & Christensen, 2017). If the final fitted model returned by the algorithm had convergence issues, we proceeded to fit zero-correlation parameter models in which the random slopes are retained but the correlation parameters are set to zero (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017; Seedorff, Oleson, & McMurray, 2019). The full random-effects structure of the zero-correlation parameter LMM required 16 (Experiment 1) or 12 (Experiment 2) variance components to be estimated. This random-effects structure was evaluated and backward-reduced to arrive at the model that was justified by the data.

Model nonconvergence tends to be a much larger issue with GLMMs than with LMMs (Seedorff et al., 2019). Indeed, the GLMMs we report are random intercept models because random slope models did not converge.

For parameter optimization, the bobyqa optimizer was used for LMMs, and a combination of Nelder-Mead and bobyqa for GLMMs. 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 0.05 (Baayen, Davidson, & Bates, 2008).

In the (G)LMM analyses, data from individual trials (subject–item combinations) were considered. For the data depicted in Figures 5, 6, and 9 to 11, means were calculated for each subject, and these were then averaged across subjects. Result figures display the data

on their original scale. When using the T.E.A. to prepare the stimulus material, for one of the photographs the different versions were not saved into the correct folders on the lab computer because of human error. For three additional scenes, participants had difficulty finding the low-salience target. These four scenes were therefore excluded from analysis.

Experiment 1

Design

Experiment 1 had a 2 × 2 × 2 within-subjects design with two-level factor target size (small vs. large), two-level factor target salience (low vs. high) and two-level factor foveal scotoma (absent vs. present), see Figure 3. Small targets were 0.41° in size (letter width), and large targets were 1.08°². Scene locations for low- and high-salience targets were algorithmically determined, as described above, at the lower and upper quartile level of salience change. The factor scotoma refers to the implementation of a visual field manipulation. In the scotoma condition, foveal vision was blocked by a gaze-contingent moving mask. This was contrasted with a normal-vision control condition.

The 120 scenes used in the experiment were assigned to eight lists of 15 scenes each. The scene lists were rotated over participants, such that a given participant was exposed to a list for only one of the eight experimental conditions created by the 2 × 2 × 2 design. There were eight groups of four participants, and each group of participants was exposed to unique combinations of list and experimental condition. To summarize, participants viewed each of the 120 scene items once, with 15 scenes in each of the eight experimental conditions. Across the 32 participants, each scene item appeared in each condition four times.

The visual field manipulation was blocked so that participants completed two blocks of trials in the experiment: in one block observers' foveal vision was available, in the other block it was obstructed by a simulated foveal scotoma. Each block started with four practice trials, one for each target salience × size condition. The order of blocks was counterbalanced across subjects. Within a block, scenes were presented randomly.

Results

In a first step, we analyzed different measures of search accuracy as indicators of search efficiency. For correct trials, we then analyzed search time and its subcomponents. Finally, we examined saccade
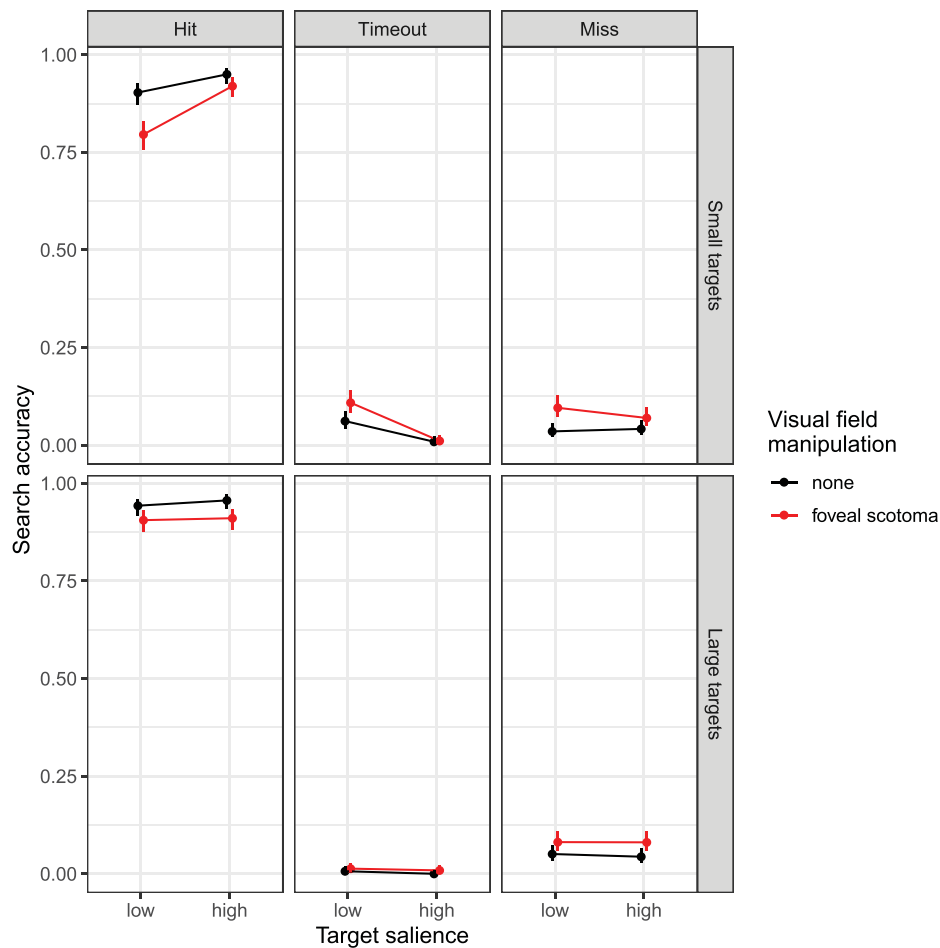
Figure 4. Measures of search accuracy for Experiment 1. Top row: small targets, bottom row: large targets. Each column presents means obtained for a designated dependent variable (see text for definitions). In each panel, data are shown for low- and high-salience targets during visual search with a simulated foveal scotoma (red) or without one (black). Data points are binomial proportions; error bars are 95% binomial proportion confidence intervals (Wilson, 1927).

amplitude and fixation duration across the viewing period.

Search accuracy

The first set of analyses examined the likelihood of finding the target letter in the scene. Performance for each experimental condition was divided into probabilities of "hit," "miss," and "timeout" cases (Clayden et al., 2020; Nuthmann, 2014). If the participant had not responded within 15 seconds, the trial was coded as a "timeout." A response was scored as a "hit" if the participant indicated to have located the target by button press and their gaze was within the rectangular area of interest comprising the target; otherwise, the response was scored as a "miss." The area of interest was 2.9° × 2.9° in size (Clayden et al., 2020). It was the same for both target sizes and included a buffer, following recommendations by Holmqvist and Andersson (2017).

There was a significant effect of scotoma on the probability of "hitting" the target such that participants were less likely to correctly locate and accept the target when foveal vision was not available, $b = -0.70$, $SE = 0.13$, $z = -5.49$, $p < 0.001$ (Figure 4, left column). Moreover, search accuracy was significantly higher for large as compared to small targets ($b = 0.41$, $SE = 0.13$, $z = 3.22$, $p = 0.001$), and it was higher for high-salience compared to low-salience targets ($b = 0.56$, $SE = 0.13$, $z = 4.43$, $p < .001$). Only one of the interactions was significant (Table 1). Specifically, there was a significant size × salience interaction ($b = -0.73$, $SE = 0.25$, $z = -2.87$, $p = 0.004$), indicating that the salience effect was smaller for large as compared to small targets. As a matter of fact, the data displayed in Figure 4 suggest that the effect of one variable was absent for the easier condition of the other variable. To test this explicitly, we specified a posthoc GLMM using dummy-coded variables with the following reference levels: large targets, high-salience targets,

		Intercept	Target size	Target salience	Foveal scotoma	Size × salience	Size × scotoma	Salience × scotoma	Size × salience × scotoma
Probability correct	b	2.71	0.41	0.56	−0.7	−0.73	0.08	0.1	−0.67
	SE	0.15	0.13	0.13	0.13	0.25	0.25	0.25	0.51
	z	18.35	3.22	4.43	−5.49	−2.87	0.32	0.38	−1.32
	p	<0.001	0.001	<0.001	<0.001	0.004	**0.746**	**0.704**	**0.186**
Search time	b	2086.33	−927.79	−1230.15	149.09	958.04	−62.7	75.1	−251.91
	SE	112.75	99.45	121.26	81.54	166.65	101.36	128.17	202.45
	t	18.5	−9.33	−10.14	**1.83**	5.75	**−0.62**	**0.59**	**−1.24**
Search initiation time	b	269.41	−9.02	−19.83	38.14	−9.33	−8.04	−11.61	0.35
	SE	8.35	3.44	4.1	12.1	6.88	6.88	6.87	13.76
	t	32.26	−2.62	−4.84	3.15	**−1.35**	**−1.17**	**−1.69**	**0.03**
Scanning time	b	1127.95	−723.15	−968.64	−16.72	768.16	6.73	185.87	−296.28
	SE	71.43	80.07	104.24	46.4	135.65	92.8	92.7	185.42
	t	15.79	−9.03	−9.29	**−0.36**	5.66	**0.07**	2.01	**−1.6**
Verification time	b	677.29	−178.51	−225.95	118.25	172.78	−60.1	−92.32	42.14
	SE	69.39	36.97	37.34	45.73	64.39	50.45	57.28	100.78
	t	9.76	−4.83	−6.05	2.59	2.68	**−1.19**	**−1.61**	**0.42**
Saccade amplitude	b	5.3	−0.43	−0.57	0.4	−0.22	−0.26	0.01	−0.4
	SE	0.11	0.06	0.1	0.08	0.12	0.12	0.12	0.24
	t	49.19	−6.96	−5.99	4.89	**−1.81**	−2.11	**0.04**	**−1.63**
Fixation duration	b	204.56	−9.49	−20.01	19.85	7.68	7.85	7.09	3.63
	SE	4.1	2.55	2.86	3.88	3.52	4.09	3.52	9.69
	t	49.85	−3.72	−6.99	5.12	2.18	**1.92**	2.02	**0.37**

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. *Notes:* Nonsignificant coefficients are set in bold (LLMs: $|t| < 1.96$; GLMMs: $p > 0.05$). See text for further details.
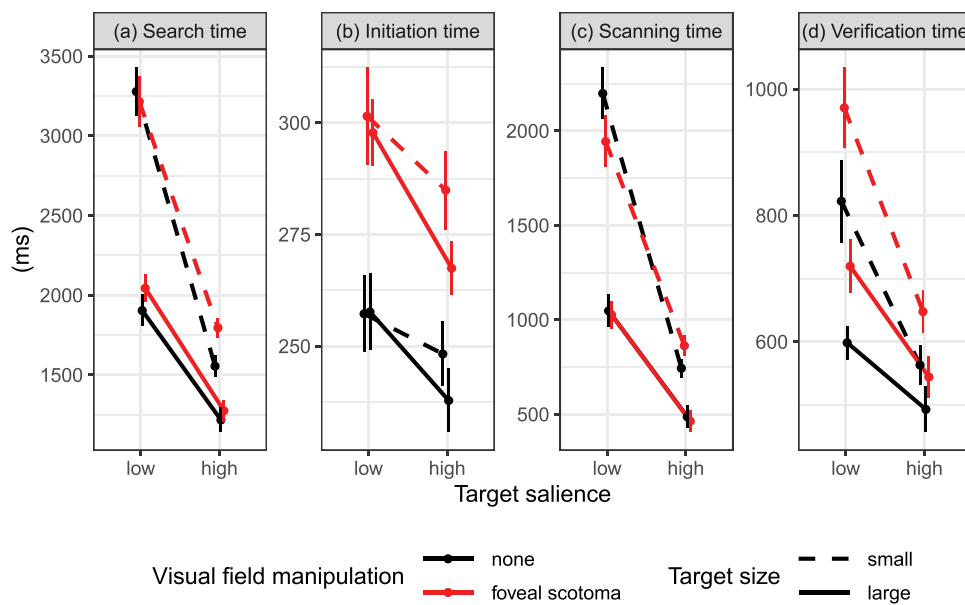
Figure 5. Search time and its three epochs for Experiment 1. Each panel displays the means for a designated dependent variable (see panel title); note the different *y*-axis scales for the different measures. Targets differed in visual salience (*x*-axis) and size (small: dashed line, large: solid line). Observers searched the scene either with a simulated foveal scotoma (red line) or without one (black line). Search times are the sum of search initiation, scanning, and verification times. Error bars are within-subjects standard errors, using the Cousineau-Morey method (Cousineau, 2005; Morey, 2008).

foveal scotoma. The simple effect for target size, representing the size effect for high-salience targets, was not significant ($b = 0.10$, $SE = 0.24$, $z = 0.41$, $p = 0.685$). The simple effect for target salience, representing the salience effect for large targets, was also not significant ($b = -0.07$, $SE = 0.23$, $z = -0.31$, $p = 0.754$). However, the size × salience interaction was significant ($b = -1.05$, $SE = 0.31$, $z = -3.37$, $p < 0.001$).

When searching with a scotoma, the probability of missing the target was increased ($b = 0.72$, $SE = 0.14$, $z = 5.00$, $p < 0.001$). Timeout probability was low, with no timeouts for large high-salience targets; no statistical analysis was performed.

Search time and its subcomponents

Search time is the time taken from scene onset to participants' button press terminating the search. Participants' gaze data were used to split search time 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). Search initiation time is the interval between scene onset and the initiation of the first saccade (i.e., initial saccade latency). Scanning time is the time from the first eye movement until the participant's gaze enters the target's area of interest. Verification time is the time from first entering the target interest area until the participant confirms their decision via button press. Whereas the scanning time measure reflects the process of localizing the target in space, verification time reflects the time needed to decide that the fixated object is the target (Malcolm & Henderson, 2009). Longer scanning times indicate weaker target guidance. Long verification times tend to include instances in which observers fixated the target but then continued searching before returning to it (Castelhano, Pollatsek, & Cave, 2008; Clayden et al., 2020; Rutishauser & Koch, 2007; Zhaoping & Frith, 2011; Zhaoping & Guyader, 2007). Moreover, in the absence of foveal or central vision, the eyes may move off the target to unmask it and then process it in parafoveal or peripheral vision (Clayden et al., 2020; Nuthmann, 2014). In both cases, there will be off-target fixations between the first and final fixation on the target, the number of which appears to depend on the difficulty of the search (Clayden et al., 2020; Rutishauser & Koch, 2007).

We manipulated both the target's size and its salience to explore how the effects combine. Specifically, if high-salience helps more for small targets, we should observe an interaction between target size and target salience. In previous letter-in-scene search experiments, in which target size was varied, we found that the verification process was slowed down when foveal vision was not available, whereas the actual search process, indexed by scanning time, remained unaffected (Clayden et al., 2020). Moreover, we tested whether the importance of foveal vision to target verification depended on the size of the target, but the data remained ambiguous (Clayden et al., 2020). With the present experiment, we wanted to test whether the availability of foveal vision during target verification was more important if the target's salience was reduced, along with its size. If that were the case, the foveal scotoma should be more detrimental for low-salience than for high-salience targets, and it should be most detrimental for targets that are small and low in salience.

The analysis of search times showed a significant effect of target size with faster searches for large as compared to small targets ($b = -927.79$, $SE = 99.45$, $t =$

-9.33). The effect of target salience was also significant, with shorter search times for high-salience as compared to low-salience targets ($b = -1230.15$, $SE = 121.26$, $t = -10.14$). There was also a significant interaction between target size and salience such that the salience effect was smaller for large targets ($b = 958.04$, $SE = 166.65$, $t = 5.75$). Analyses of the three subprocesses of search showed the same pattern of results (Table 1). The only exception was a nonsignificant target size × salience interaction for search initiation time ($b = -9.33$, $SE = 6.88$, $t = -1.35$).

The presence of a foveal scotoma had a significant effect on search initiation and verification, with both subprocesses of search being slowed down (Table 1). Importantly, scanning time was not prolonged when searching with a foveal scotoma ($b = -16.72$, $SE = 46.4$, $t = -0.36$). Button-press search times are the sum of search initiation, scanning, and verification times. For the untransformed data, the search-time difference between the foveal scotoma and the control condition was not significant ($b = 149.09$, $SE = 81.54$, $t = 1.83$). For the transformed data, however, the effect of scotoma was significant ($b = 0.003$, $SE = 0.001$, $t = 3.84$); it was qualified by a significant scotoma × salience interaction such that the detrimental effect of a foveal scotoma was larger for high-salience targets ($b = 0.002$, $SE = 0.001$, $t = 3.16$).

For none of the dependent variables was there a significant scotoma × size interaction (Table 1). There was no significant scotoma × salience interaction for search initiation, scanning, and verification times (Table 1). The three-way interaction was not significant for any of the dependent variables (Table 1).

Saccade amplitudes and fixation durations

Saccade amplitudes and fixation durations were analyzed to characterize eye-movement behavior during visual search (Figure 6). During scene search with a simulated foveal scotoma, we expected to observe larger saccade amplitudes and longer fixation durations (Clayden et al., 2020; Nuthmann, 2014). Moreover, in previous experiments we found an increase in target size to be associated with shorter saccade amplitudes and shorter fixation durations (Clayden et al., 2020).

For saccade amplitudes we observed a significant effect of scotoma, with longer saccades when searching with a foveal scotoma than without ($b = 0.40$, $SE = 0.08$, $t = 4.89$; Figure 6, top row). There was also a significant effect of target size with shorter saccade amplitudes for large as compared to small targets ($b = -0.43$, $SE = 0.06$, $t = -6.96$). In addition, there was a significant effect of target salience with shorter saccade amplitudes for high-salience as compared to low-salience targets ($b = -0.57$, $SE = 0.10$, $t = -5.99$). The interaction between target size and scotoma was
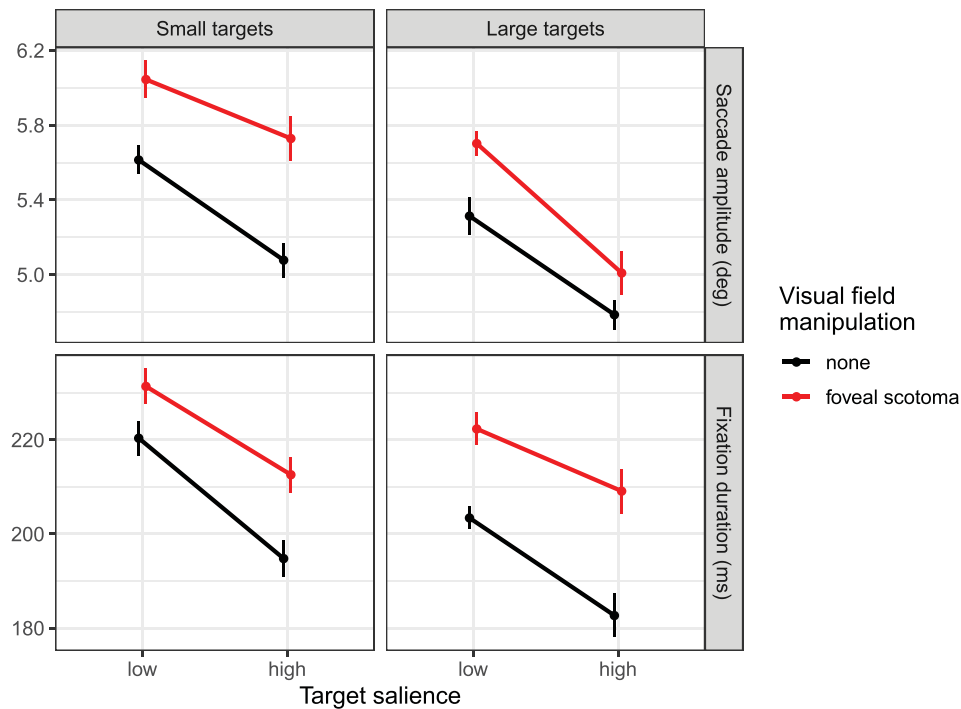
Figure 6. Mean saccade amplitudes (top row) and fixation durations (bottom row) for small targets (left column) as opposed to large targets (right column) in Experiment 1. In each panel, data are presented for low- and high-salience targets during visual search with or without a simulated foveal scotoma. Error bars are within-subjects standard errors.

significant ($b = -0.26$, $SE = 0.12$, $t = -2.11$), indicating that the size effect was larger (i.e., more negative) with a foveal scotoma than without. For the transformed data, however, this interaction was not significant ($b = -0.07$, $SE = 0.04$, $t = -1.76$). Thus the interaction was transformed away, making it noninterpretable (Loftus, 1978; Wagenmakers, Krypotos, Criss, & Iverson, 2012). None of the other interactions were significant (Table 1).

The analysis of fixation durations revealed a similar pattern of results. There was a significant effect of scotoma, with longer fixation durations when searching with a foveal scotoma than without ($b = 19.85$, $SE = 3.88$, $t = 5.12$; Figure 6, bottom row). There was also a significant effect of target size with shorter fixation durations for large as compared to small targets ($b = -9.49$, $SE = 2.55$, $t = -3.72$). In addition, there was a significant effect of target salience with shorter fixation durations for high-salience as compared to low-salience targets ($b = -20.01$, $SE = 2.86$, $t = -6.99$). Furthermore, there was a significant size × salience interaction ($b = 7.68$, $SE = 3.52$, $t = 2.18$) that was absent for the transformed data ($b = 0.08$, $SE = 0.05$, $t = 1.71$). Moreover, there was a significant scotoma × salience interaction ($b = 7.09$, $SE = 3.52$, $t = 2.02$), indicating that the salience effect was smaller with a foveal scotoma than without. None of the other interactions were significant (Table 1).

Experiment 2

Design

In Experiment 2, we dropped the manipulation of target size and instead used the small targets from Experiment 1 throughout. As in Experiment 1, we manipulated the visual salience of the target letter (low vs. high). This was crossed with another visual field manipulation: observers searched for the target with a central or peripheral scotoma, for which the normal-vision control condition provided a baseline (Figure 7). Compared with the foveal scotoma in Experiment 1 (radius: 1°), the central scotoma in Experiment 2 had a larger radius (2.5°). The central scotoma was contrasted with the inverse manipulation of a peripheral scotoma with the same radius. In the visual-cognition literature, central vision is defined as extending to about 5° from fixation, with peripheral vision being everything beyond 5° (Loschky, Szaffarczyk, Beugnet, Young, & Boucart, 2019). Technically, our central scotoma did not completely cover central vision, and our peripheral scotoma obscured more than peripheral vision.

To facilitate comparisons across experiments, we used the same scenes with the same locations for low- and high-salience targets as in Experiment 1. A given participant saw each of the 120 scene items once, with 20 scenes in each of the six experimental conditions.
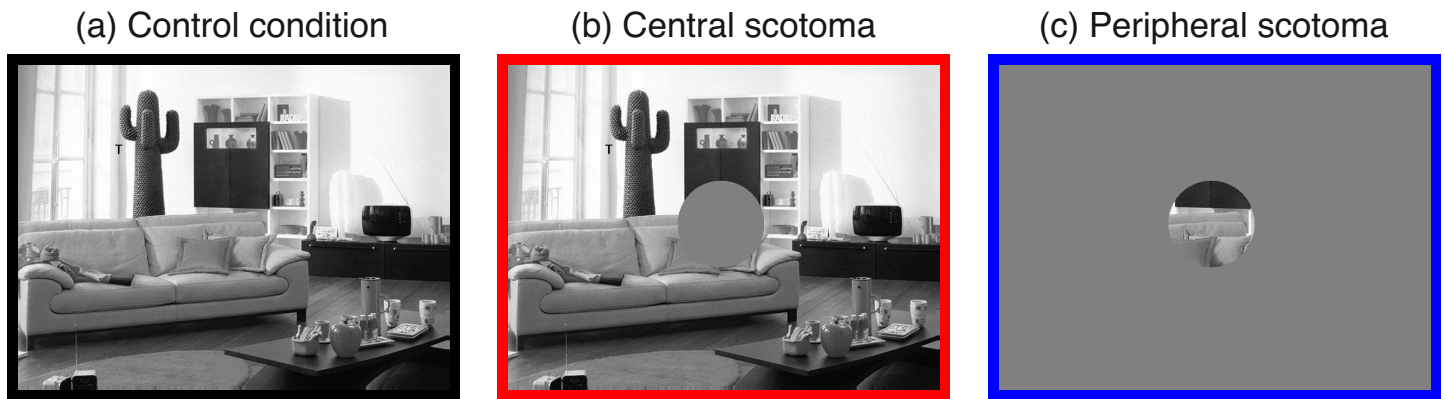
(a) Control condition (b) Central scotoma (c) Peripheral scotoma

Figure 7. Scotoma conditions used in Experiment 2. Observers searched the scene either with full vision (control condition), or with a central or peripheral scotoma (radius: 2.5°). Note that the colored borders match the colors used to distinguish the scotoma-type conditions in Figures 8 to 11. Search targets varied in visual salience; the example scene used for this figure includes the high-salience target.
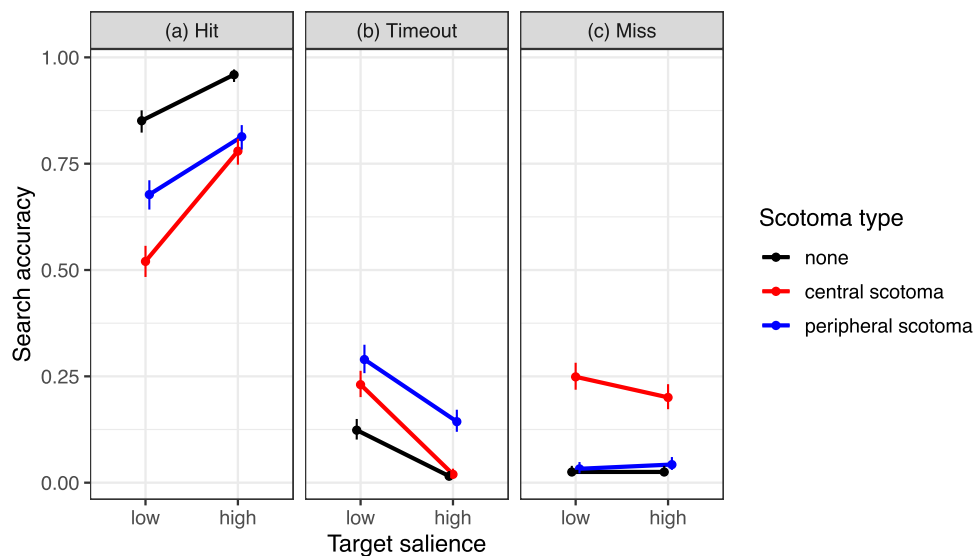
Figure 8. Measures of search accuracy for Experiment 2. Each panel presents means obtained for a designated dependent variable, which is specified in the panel title. Data are shown for low- and high-salience targets and for different scotoma types (red: central scotoma; blue: peripheral scotoma; black: no-scotoma control condition). Data points are binomial proportions, error bars are 95% binomial proportion confidence intervals (Wilson, 1927).

The visual field manipulation was blocked so that participants completed three blocks of trials in the experiment. Each block started with four practice trials, two for each target salience condition. The order of blocks was counterbalanced across subjects. Within a block, scenes were presented randomly.

Results

Search accuracy

The type of the simulated scotoma affected the probability of "hitting" the target, with highest probabilities in the no-scotoma control condition and lowest probabilities for the central scotoma (Figure 8a). The effect of scotoma type on search accuracy was tested using backward difference coding (Table 2). The GLMM results substantiated that search accuracy was significantly reduced for the peripheral scotoma condition (P) compared to the no-scotoma control condition (No) (P-No: $b = -1.44$, $SE = 0.14$, $z = -10.19$, $p < 0.001$). For the central scotoma (C), search accuracy was lower than for the peripheral scotoma (C-P: $b = -0.77$, $SE = 0.13$, $z = -6.01$, $p < 0.001$). As in Experiment 1, there was a significant main effect of target salience on search accuracy, with better performance for high-salience than for low-salience targets ($b = 1.21$, $SE = 0.10$, $z = 11.94$, $p < 0.001$).

Dependent variable	Contrast coding (scotoma type)	Reference level	Scot 1 (definition)	Scot 2 (definition)		Intercept	Target salience	Scot 1	Scot 2	Salience × Scot 1	Salience × Scot 2
Probability correct	BWD	No – P – C	P - No	C - P	b	1.76	1.21	−1.44	−0.77	−0.62	0.74
					SE	0.15	0.1	0.14	0.13	0.27	0.19
					z	11.9	11.94	−10.19	−6.01	−2.31	3.82
					p	<0.001	<0.001	<0.001	<0.001	0.021	<0.001
Search time	BWD	No – C – P	C - No	P - C	b	3932.88	−1523.04	756.36	2995.29	−707.37	1761.53
					SE	123.55	156.45	117.71	214.97	227.02	354.56
					t	31.83	−9.74	6.43	13.93	−3.12	4.97
Search initiation time	Simple	No scotoma	C - No	P - No	b	273.21	−8.62	17.09	81.99	4.06	12.85
					SE	6.83	3.77	8.18	7.43	8.91	9.45
					t	39.98	−2.29	2.09	11.03	**0.46**	**1.36**
Scanning time	Simple	No scotoma	C - No	P - No	b	2709.84	−1043.68	85.21	3725.53	6.18	978.91
					SE	92.54	124.55	108.27	187.46	209.5	304.12
					t	29.28	−8.38	**0.79**	19.87	**0.03**	3.22
Verification time	Simple	No scotoma	C - No	P - No	b	910.92	−442.75	564.89	−36.76	−660.97	54.2
					SE	59.84	63.09	98.63	64.28	172.18	84.33
					t	15.22	−7.02	5.73	**−0.57**	−3.84	**0.64**
Saccade amplitude	Simple	No scotoma	C - No	P - No	b	4.87	−0.42	1.6	−2.2	−0.18	0.43
					SE	0.1	0.06	0.15	0.1	0.16	0.08
					t	47.11	−6.5	10.63	−21.7	**−1.16**	5.27
Fixation duration	BWD	No – C – P	C - No	P - C	b	211.77	−12.24	17.09	18.79	11.58	10.29
					SE	3.82	2.12	4.5	6.33	4.07	4.23
					t	55.42	−5.78	3.8	2.97	2.85	2.43

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. *Notes:* Nonsignificant coefficients are set in bold (LLMs: |*t*| < 1.96; GLMMs: *p* > .05). BWD = backward difference. See text for further details.

The salience effect was significantly reduced for the peripheral scotoma compared to the no-scotoma control condition (salience × P-No interaction: $b = -0.62$, $SE = 0.27$, $z = -2.31$, $p = 0.021$). The salience effect was significantly increased for the central scotoma compared to the peripheral scotoma (salience × C-P interaction: $b = 0.74$, $SE = 0.19$, $z = 3.82$, $p < 0.001$).

The drop in performance for search with a peripheral scotoma was due to an increase in timed out trials (Figure 8b). The further loss in performance when searching with a central scotoma originated from two sources. On the one hand, there were more timed out trials than in the control condition but fewer than with a peripheral scotoma (Figure 8b). On the other hand, the probability of missing the target was increased (Figure 8c).

Search time and its subcomponents

Trials with correct responses were analyzed further. The type of the simulated scotoma affected button-press search times, which were shortest in the no-scotoma control condition and longest when searching with a peripheral scotoma (Figure 9a). The effect of scotoma type on search times was tested using backward difference coding. Search times were significantly longer during search with a central scotoma than during search without a scotoma (C-No: $b = 756.36$, $SE = 117.71$, $t = 6.43$). Search times were further increased for the peripheral scotoma compared to the central scotoma (P-C: $b = 2995.29$, $SE = 214.97$, $t = 13.93$). Moreover, there was a significant main effect of target salience with shorter search times for high-salience compared to low-salience targets ($b = -1523.04$, $SE = 156.45$, $t =$

-9.74). The salience effect was significantly increased for the central scotoma compared to the no-scotoma control condition (salience × C-No interaction: $b = -707.37$, $SE = 227.02$, $t = -3.12$). The salience effect was significantly reduced for the peripheral scotoma compared to the central scotoma (salience × P-C interaction: $b = 1761.53$, $SE = 354.56$, $t = 4.97$).

Based on participants' gaze data, button-press search times were decomposed into search initiation, scanning, and verification times (Figures 9b through d). To evaluate the effect of scotoma type, we used simple coding with the no-scotoma control condition as the reference level. For search with a peripheral scotoma, search initiation time was significantly increased ($b = 81.99$, $SE = 7.43$, $t = 11.03$). Search initiation times were also increased for the central scotoma; this effect was significant for the untransformed data ($b = 17.09$, $SE = 8.18$, $t = 2.09$) but not for the transformed data ($b = 2.16$, $SE = 1.16$, $t = 1.86$). Moreover, there was a significant main effect of target salience with shorter search initiation times for high-salience compared to low-salience targets ($b = -8.62$, $SE = 3.77$, $t = -2.29$). The two interactions involving salience were not significant (Table 2).

Scanning time was significantly prolonged when searching with a peripheral scotoma ($b = 3725.53$, $SE = 187.46$, $t = 19.87$). For the central scotoma, there was a numerical increase in scanning time which was not significant ($b = 85.21$, $SE = 108.27$, $t = 0.79$); for the transformed data, however, it was significant ($b = 0.14$, $SE = 0.06$, $t = 2.38$). Scanning times were shorter for high-salience compared to low-salience targets ($b = -1043.68$, $SE = 124.55$, $t = -8.38$). The effect of target salience was significantly reduced for the peripheral
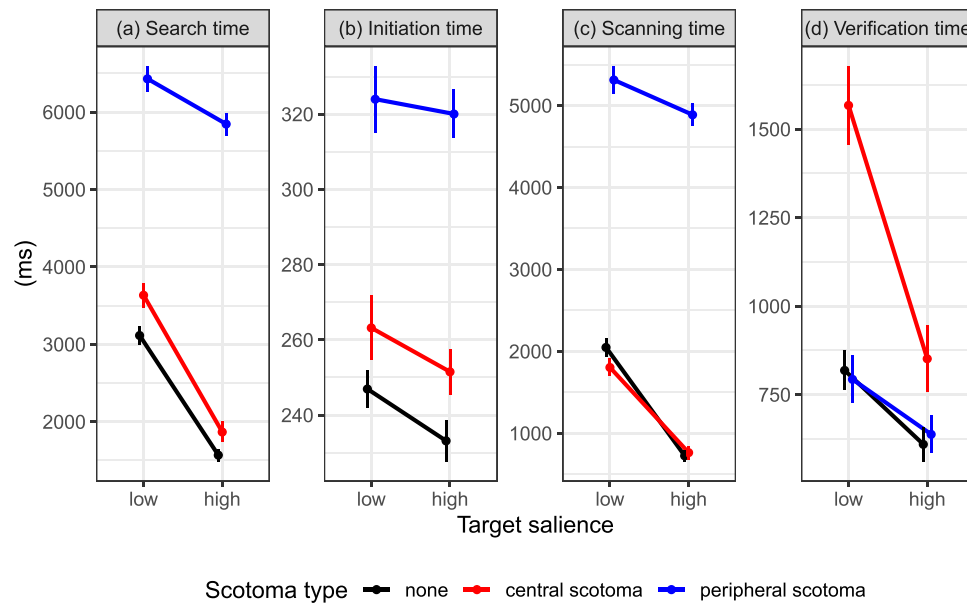
Figure 9. Search time and its three epochs for Experiment 2. Each panel displays the means for a designated dependent variable (see panel title); note the different *y*-axis scales for the different measures. Results are presented for low- and high-salience targets and for different scotoma types (red: central scotoma; blue: peripheral scotoma; black: no-scotoma control condition). Error bars are within-subjects standard errors, using the Cousineau-Morey method (Cousineau, 2005; Morey, 2008).

scotoma ($b = 978.91$, $SE = 304.12$, $t = 3.22$) but not for the central scotoma ($b = 6.18$, $SE = 209.5$, $t = 0.03$).

Verification time was significantly prolonged when searching with a central scotoma ($b = 564.89$, $SE = 98.63$, $t = 5.73$) but not when searching with a peripheral scotoma ($b = -36.76$, $SE = 64.28$, $t = -0.57$). Verification times were shorter for high-salience compared to low-salience targets ($b = -442.75$, $SE = 63.09$, $t = -7.02$). This effect was significantly increased for the central scotoma ($b = -660.97$, $SE = 172.18$, $t = -3.84$) but not for the peripheral scotoma ($b = 54.2$, $SE = 84.33$, $t = 0.64$).

Saccade amplitudes and fixation durations

Moving-window studies that implemented something akin to our peripheral scotoma have consistently reported shorter saccade amplitudes and longer fixation durations than in a normal vision control condition (e.g., Loschky & McConkie, 2002; Nuthmann, 2014). By contrast, masking or degrading central vision tends to increase both saccade amplitudes and fixation durations (Miellet et al., 2010; Nuthmann, 2014).

The present data replicate the "windowing effect" on saccade amplitudes. Compared to the no-scotoma control condition, saccade amplitudes were significantly longer when searching with a central scotoma ($b = 1.6$, $SE = 0.15$, $t = 10.63$) and significantly shorter when searching with a peripheral scotoma ($b = -2.2$, $SE = 0.1$, $t = -21.7$). Moreover, as in Experiment 1 there was a significant main effect of target salience with

shorter saccade amplitudes for high-salience compared to low-salience targets ($b = -0.42$, $SE = 0.06$, $t = -6.5$). There was also a significant salience × peripheral scotoma interaction ($b = 0.43$, $SE = 0.08$, $t = 5.27$), indicating that the effect of target salience was reduced for the peripheral scotoma. The interaction between salience and central scotoma was not significant (Table 2).

The type of the simulated scotoma also affected fixation durations, which were shortest in the no-scotoma control condition and longest when searching with a peripheral scotoma (Figure 10b). The effect of scotoma type on fixation durations was tested using backward difference coding (Table 2). The LMM results substantiated that fixation durations were significantly longer during search with a central scotoma than during search without a scotoma (C-No: $b = 17.09$, $SE = 4.5$, $t = 3.8$). For the peripheral scotoma, fixation durations were significantly increased compared to the central scotoma (P-C: $b = 18.79$, $SE = 6.33$, $t = 2.97$). As in Experiment 1, there was also a significant main effect of target salience with shorter fixation durations for high-salience compared to low-salience targets ($b = -12.24$, $SE = 2.12$, $t = -5.78$). The salience effect was significantly reduced for the central scotoma compared to the no-scotoma control condition (salience × C-No interaction: $b = 11.58$, $SE = 4.07$, $t = 2.85$). The salience effect was further reduced for the peripheral scotoma compared to the central scotoma (salience × P-C interaction: $b = 10.29$, $SE = 4.23$, $t = 2.43$).
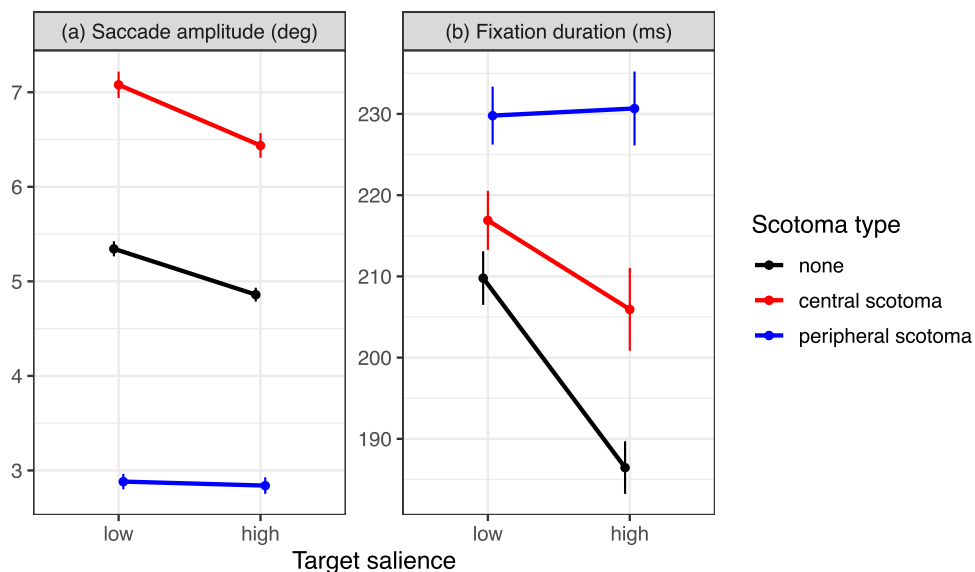
Figure 10. Mean saccade amplitudes (a) and fixation durations (b) in Experiment 2 as a function of target salience and scotoma type: red: central scotoma; blue: peripheral scotoma; black: no-scotoma control condition. Error bars are within-subjects standard errors.
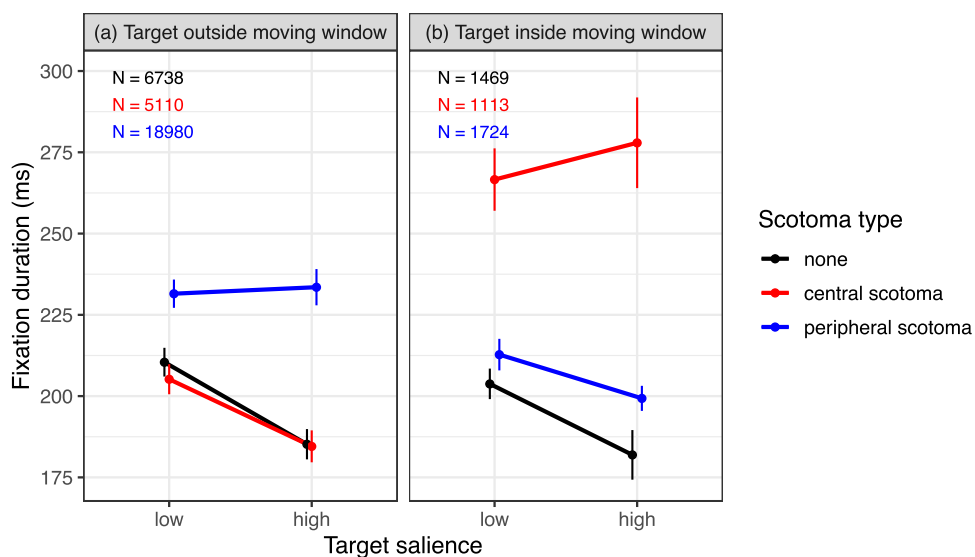
Figure 11. Mean fixation durations in Experiment 2 as a function of target salience, scotoma type, and whether the target was outside (a) or inside (b) the scotoma window that moved with the participants' eyes. Error bars are within-subjects standard errors. N = number of observations for a given scotoma-type condition.

Control analyses

With a peripheral scotoma, the target was not visible to the observer during their initial fixation at the center of the scene. During most subsequent valid fixations, the target remained invisible because it was outside the window in which scene content was available. Thus, search initiation times, saccade amplitudes, and fixation durations should be unaffected by target salience in this condition. To test this explicitly, we specified additional LMMs using dummy coding and the peripheral scotoma as reference level. In such a model, the simple effect for target

salience represents the salience effect for the peripheral scotoma. No significant salience effects were found (search initiation times: $b = -0.52$, $SE = 6.32$, $t = -0.08$; saccade amplitudes: $b = -0.03$, $SE = 0.04$, $t = -0.87$; fixation durations: $b = -0.02$, $SE = 0.01$, $t = -1.64$).

Results from existing studies suggest that visual information within both foveal, parafoveal, and peripheral vision can influence fixation duration (Einhäuser, Atzert, & Nuthmann, 2020, for review). Therefore an additional analysis explored whether effects of target salience on fixation duration arise

from both central and peripheral processing. For each individual fixation, we determined whether the target was inside or outside the circular window that was used to create the two scotomas. As an approximation, the midpoint of the target was used for this evaluation. For the central scotoma, the target was visible if it was outside the window (see Figure 7b), and invisible if it was inside the window. Conversely, for the peripheral scotoma the target was visible if it was inside the window, and invisible if it was outside the window (see Figure 7c). We expected target salience to only modulate fixation durations if the target was visible. The data are consistent with this prediction. For the central scotoma, the salience effect was present when the target was outside the window (Figure 11a), whereas it was absent when the target was inside the window (Figure 11b). For the peripheral scotoma, a salience effect emerged if the target was inside the window (Figure 11b), whereas it was absent when the target was outside the window (Figure 11a). For the no-scotoma control condition, where the target was always present, the salience effect was present for both types of fixations. Interestingly, the data also suggest that fixation durations during search with the central scotoma were not elevated when the target was visible in the periphery (Figure 11a). Given the post hoc nature of this explorative analysis, no formal statistical analyses were conducted. The number of cases in which the target was outside the window during the fixation amounted to 88% (see Figure 11 for a breakdown). This is why the analysis of all valid fixations yielded no salience effect for the peripheral scotoma and a reduced salience effect for the central scotoma (Figure 10b).

General discussion

Previous research on visual search has demonstrated that eye guidance by visual salience can be moderated, or even completely overridden by top-down guidance (Einhäuser, Rutishauser, & Koch, 2008; Henderson et al., 2009; Underwood & Foulsham, 2006). Accordingly, the role of visual salience has been marginalized in the literature on active search through eye movements. Using a letter-in-scene search task, we demonstrate in two experiments that visual salience *can* affect both the process of localizing the target in space and the process of accepting the target as the target. Moreover, in Experiment 1 we found an interaction between target salience and size, and that foveal vision was relatively unimportant even for small low-salience targets. Results from Experiment 2 showed that salience affected eye guidance during search in both central and peripheral vision.

The role visual salience plays during search was first investigated using simple displays which observers are asked to search covertly; that is, without making eye movements (Wolfe, 2015, for review). A complementary approach is to record eye movements during visual search for a target in relatively large and dense arrays (Rutishauser & Koch, 2007). Using this approach, Zhaoping and Guyader (2007) compared two efficient simple feature search tasks with two inefficient search tasks. The inefficient search tasks varied in difficulty because of differences in target-distractor similarity. Scanning times were longer for the inefficient searches than for efficient pop-out searches. For the two inefficient searches, the authors observed differences in verification time (dubbed eye-to-hand latency) but not scanning time. Thus visual salience can affect target localization and verification in densely packed arrays of simple stimuli, in a manner that is specific to the respective task (see also Zhaoping & Frith, 2011).

Investigating the causal influence of features on gaze guidance during scene search requires one to use an experimental approach in which objects or regions in natural scenes are manipulated (Foulsham & Underwood, 2007). In the studies reviewed in the Introduction, the approach has been to select targets based on the output from versions of a popular saliency map model. When manipulating properties of real-world objects in naturalistic scenes, it is impossible to exert perfect experimental control over relevant dimensions. Therefore the possibility exists that—in some existing scene sets—visual salience is confounded with other variables like object size, eccentricity, and semantic congruency. To address these issues, we used the T.E.A. (Clayden et al., 2020) to parametrically manipulate target salience and size in a letter-in-scene search task. In this task, the location of the target is not predicted by the meaning of the scene or by the identity of objects in the scene. Our task still approximates natural behavior because there are real-world searches for which there is minimal guidance by scene context (e.g., search for a fly). Moreover, scene processing and object identification are not totally suppressed when searching for a "T" overlaid onto the scene (T. H. W. Cornelissen & Võ, 2017). One caveat regarding generalizing from letter search to object search in scenes is that the letter targets tend to violate the physical rules of the scene environment in which they appear, such as gravity and surface reflectance. Additionally, although we used images of naturalistic scenes to improve the ecological validity of the search task, these scenes are still two-dimensional static representations of the environment, and so generalization to the natural world should be made with caution.

In both of our experiments, we found main effects of salience with faster search times for high-salience than for low-salience targets. Existing research has provided inconsistent results in this regard. On the one hand,

null effects were found in studies in which targets were real objects in composed scene photographs (Foulsham & Underwood, 2007; Underwood et al., 2008). On the other hand, salience did affect search times when scene cutouts were used as targets (Foulsham & Underwood, 2011, Experiment 1). In the latter study, salience affected verification time only, but not the latency to first fixation on the target (i.e., search initiation time plus scanning time). In contrast, our results demonstrate that visual salience can facilitate both eye-movement guidance to the target as well as target verification. The different results may be due to differences in the task requirements. We used a target acquisition task (Zelinsky, 2008) whereas Foulsham and Underwood (2011) required observers to decide about the presence/absence of the target. Moreover, their targets were much bigger (6° squares) than ours. These design features may also account for the fact that their mean verification times were more than twice as long as scanning times.

Using context-free targets in our experiments implied that scene context and semantic relationships could not facilitate search guidance. An alternative approach is to disrupt scene context by "scrambling" the images (Biederman, 1972). In a study by Foulsham, Alan, and Kingstone (2011), observers searched for contextually relevant targets against intact or scrambled scene backgrounds. Correlational analyses suggested that more salient targets were fixated more quickly in scrambled scenes only.

In sum, our results provide an existence proof that eye guidance by visual salience is possible during active search in naturalistic scenes. Depending on the specific task demands, this bottom-up guidance can be moderated or completely overridden by top-down guidance (Einhäuser et al., 2008; Foulsham & Underwood, 2007; Henderson et al., 2007; Henderson et al., 2009; Underwood & Foulsham, 2006; Underwood et al., 2006; Underwood et al., 2008).

In Experiment 1, we also manipulated the size of the target and found that large targets were easier and faster to find than small targets (cf. Clayden et al., 2020). As a novel result, we not only found independent effects of target salience and size, but also an interaction between the two variables. For search accuracy, the salience effect was only present for small targets, and the size effect was only present for low-salience targets. For scanning times, verification times, and search times, the interaction implied that the effect of target salience was larger for small than for large targets (Figure 5). Future work could involve testing whether the size × salience interaction generalizes from letter search to object-based fixation selection in scenes (cf. Nuthmann et al., 2020; Stoll et al., 2015). More generally, our results lend support to the view that saliency models may be enhanced by addressing the size feature more explicitly (Borji et al., 2013b).

The results for the foveal, central, and peripheral scotomas tell us how important the different regions of the visual field are for visual search and its subprocesses. During search with any type of scotoma, observers were significantly less likely to find the target than with normal vision. However, when the target was found despite the presence of a simulated foveal scotoma (Experiment 1), search times were not much elevated (Figure 5a, Table 1). In contrast, the presence of a central or peripheral scotoma (Experiment 2) led to clear search time costs (Figure 9a, Table 2). As expected, the peripheral scotoma was much more detrimental than the central scotoma, confirming that eye movements are guided by peripheral vision.

Analyzing sub-processes of search allowed for testing the assumption of a central-peripheral dichotomy according to which peripheral vision is mainly for selecting or looking, while central vision is mainly for seeing or recognizing (Zhaoping, 2019). In Experiment 1, we found that verification times, but not scanning times, were significantly prolonged when searching with a foveal scotoma (see also Clayden et al., 2020). In Experiment 2, we found that scanning times were prolonged for the peripheral but not for the central scotoma, whereas verification times were prolonged for the central scotoma but not for the peripheral scotoma (cf. Nuthmann, 2014). Collectively, the data highlight the importance of peripheral vision for target localization, and the importance of foveal and central vision for target verification. This pattern of results is consistent with the central-peripheral dichotomy (Zhaoping, 2019).

The interaction between salience and type of scotoma informs us about the role target salience plays in central and peripheral vision (Experiment 2). A central question concerned the degree to which target salience affects localization in the periphery and verification in central vision. In the normal vision baseline condition, both scanning and verification time showed a significant advantage for high-salience targets.

By comparison, the peripheral scotoma weakened the effect of target salience on scanning time (Figure 9c) and also total search time (Figure 9a). This finding is different from results by Foulsham and Underwood (2011). In their Experiment 2, the authors used a gaze-contingent 6° square window to selectively remove image features from the periphery. They tested three image features that are important for saccade target selection under the saliency map hypothesis: (1) color, (2) high-spatial frequency information, and (3) contrast (i.e., the contrast of the image was globally lowered). If saliency in peripheral vision was guiding eye movements toward the target, then peripheral filtering should eliminate or at least diminish the effect of target salience. Contrary to

these predictions, the authors found a significant salience effect on the latency to first fixate the target in all three filtering conditions. Unfortunately, their Experiment 2 lacked a full-vision control condition. Moreover, the filtering manipulations left some of the saliency map representation intact. By comparison, the peripheral scotoma in our Experiment 2 blocked out peripheral vision completely. Compared with the no-scotoma control condition, we observed a greatly diminished effect of target salience during target localization, as indexed by scanning time (Figure 9c). Our results therefore suggest that salience in peripheral vision was guiding eye movements towards the target.

The central scotoma increased the effect of target salience on verification time (Figure 9d, Table 2). Compared with the no-scotoma control condition, mean target verification times were not only elevated, but they also showed a significantly increased effect of target salience. Thus our results not only suggest that central vision benefits target verification (cf. Nuthmann, 2014), but also that this subprocess of search is influenced by target salience. The increase in the salience effect for verification times was large enough to produce an increased salience effect on total search time as well (Figure 9a, Table 2).

The data from both experiments replicate the well-known "windowing effect" on saccade amplitudes, which reflects a tendency to fixate more locations in the nondegraded scene area than the degraded area (Loschky & McConkie, 2002; Miellet et al., 2010; Nuthmann, 2014; Reingold & Loschky, 2002). Moreover, the present results replicate the finding that fixation durations are elevated in the presence of an artificial scotoma (Clayden et al., 2020; Miellet et al., 2010; Nuthmann, 2014). In our experiments, we experimentally manipulated properties of the search target, and our analysis of saccade amplitudes and fixation durations considered the entire search period. On a given fixation, the target was situated in either foveal, central, or peripheral vision, where it could be obscured by a simulated scotoma or not. In Experiment 1, global eye-movement parameters were affected by target properties such that large targets and high-salience targets were associated with shorter saccade amplitudes and shorter fixation durations; for target size, similar results were obtained by Clayden et al. (2020). Interestingly, significant effects of target size and salience were already present for the duration of the very first fixation, measured as search initiation time (Figure 5b, Table 1). Previous research has demonstrated that the "story," or gist, of a scene can be gleaned from it within around 100 ms of the onset of a scene (Oliva, 2005; Potter, 1975). Scene gist is typically perceived without recognizing any individual object. Therefore we tentatively propose that observers, during the first glance of the scene, may

form a hypothesis about the scene's search difficulty in terms of target size and salience and globally adjust their fixation durations and saccade amplitudes accordingly.

In Experiment 2, we replicated effects of target salience on saccade amplitudes, fixation durations, and search initiation times (Table 2). Interestingly, the results for the scotoma conditions in both experiments provide clues about necessary conditions for these effects to occur. The peripheral scotoma in Experiment 2 prevented observers from analyzing the scene gist and covered the target during most fixations, including the very first. In this condition, no differences for low- and high-salience targets were observed for saccade amplitudes and search initiation times; the same was true for fixation durations, as long as the target was outside the window in which scene content was visible. For saccade amplitudes, the effect of target salience was unchanged when searching with a foveal scotoma (Experiment 1) or with a central scotoma (Experiment 2). For fixation durations, the effect of target salience was reduced when searching with a foveal scotoma. For the central scotoma, the salience effect was present when the target was visible (outside the scotoma) and absent when it was not visible because of being masked by the scotoma. When all valid fixations were analyzed together, the salience effect was therefore smaller in the central-scotoma condition than in the no-scotoma control condition. Collectively, the data suggest that the salience effect on fixation durations arises from both foveal, central, and peripheral processing. Moreover, peripheral vision needs to be intact to observe effects of target salience on saccade amplitudes.

Conclusions

Methodologically, reliably disentangling stimulus-driven and task-driven influences on human behavior requires researchers to exert experimental control over relevant stimulus dimensions, which is challenging when working with images of naturalistic scenes. Here, we placed context-free targets within scenes using the T.E.A. (Clayden et al., 2020), which allowed us to manipulate their salience and size parametrically. When using these stimuli for a target acquisition task in two experiments, clear effects of target salience on search performance and eye-movement parameters were found. Moreover, the results obtained in different simulated scotoma conditions lend further support to the central-peripheral dichotomy (Zhaoping, 2019).

Acknowledgments

Footnotes

[1]The code for the T.E.A. is available at https://github.com/AdamClayden93/tea.

[2]Compared to the five target sizes that were tested in the two experiments of Clayden et al. (2020), our small targets correspond to their intermediate targets, whereas our large targets correspond to their large targets.

References

Adeli, H., Vitu, F., & Zelinsky, G. J. (2017). A model of the superior colliculus predicts fixation locations during scene viewing and visual search. *Journal of Neuroscience, 37*(6), 1453–1467, https://doi.org/10.1523/JNEUROSCI.0825-16.2016.

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.

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.

Bates, D., Maechler, 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. (1988). The effect of simulated scotomas on visual search in normal subjects. *Investigative Ophthalmology & Visual Science, 29*(3), 470–475.

Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America A-Optics Image Science and Vision, 19*(6), 1096–1106, https://doi.org/10.1364/JOSAA.19.001096.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*(4043), 77–80, https://doi.org/10.1126/science.177.4043.77.

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.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., . . . White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution, 24*(3), 127–135, https://doi.org/10.1016/j.tree.2008.10.008.

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

Borji, A., Sihite, D. N., & Itti, L. (2013a). Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *IEEE Transactions on Image Processing, 22*(1), 55–69, https://doi.org/10.1109/TIP.2012.2210727.

Borji, A., Sihite, D. N., & Itti, L. (2013b). What stands out in a scene? A study of human explicit saliency judgment. *Vision Research, 91*, 62–77, https://doi.org/10.1016/j.visres.2013.07.016.

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.

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. *Psychnomic 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.

Demidenko, E. (2013). *Mixed models: Theory and applications with R* (2d ed.). Hoboken, New Jersey: John Wiley & Sons. https://doi.org/10.1002/9781118651537.

Duchowski, A. T., & Çöltekin, A. (2007). Foveated gaze-contingent displays for peripheral LOD management, 3D visualization, and stereo imaging. *ACM Transactions on Multimedia Computing, Communications, and Applications, 3*(4):24, 1–18, https://doi.org/10.1145/1314303.1314309.

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

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, 1–19, https://doi.org/10.1167/8.2.2.

Foulsham, T., Alan, R., & Kingstone, A. (2011). Scrambled eyes? Disrupting scene structure impedes focal processing and increases bottom-up guidance. *Attention Perception & Psychophysics, 73*(7), 2008–2025, https://doi.org/10.3758/s13414-011-0158-y.

Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception, 36*(8), 1123–1138, https://doi.org/10.1068/p5659.

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.

Geringswald, F., Baumgartner, F. J., & Pollmann, S. (2013). A behavioral task for the validation of a gaze-contingent simulated scotoma. *Behavior Research Methods, 45*(4), 1313–1321, https://doi.org/10.3758/s13428-013-0321-6.

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, 1–19, https://doi.org/10.1167/12.1.9.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford: Elsevier. https://doi.org/10.1016/B978-008044980-7/50027-6.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 1–58). New York: Psychology Press.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review, 16*(5), 850–856, https://doi.org/10.3758/PBR.16.5.850.

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

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). Oxford: Elsevier Science Ltd. https://doi.org/10.1016/B978-008043361-5/50003-1.

Itti, L. (2006). Quantitative modelling of perceptual salience at human eye position. *Visual Cognition, 14*(4–8), 959–984, https://doi.org/10.1080/13506280500195672.

Itti, L., & Koch, C. (1999). Comparison of feature combination strategies for saliency-based visual attention systems. *Proc. SPIE, Vol. 3644*, pp. 473–482, https://doi.org/10.1117/12.348467.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10–12), 1489–1506, https://doi.org/10.1016/S0042-6989(99)00163-7.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254–1259, https://doi.org/10.1109/34.730558.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446, https://doi.org/10.1016/j.jml.2007.11.007.

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

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

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

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.

Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition, 6*(3), 312–319, https://doi.org/10.3758/BF03197461.

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., 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, 1–21, 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, 1–10, https://doi.org/10.1145/1314303.1314310.

Malcolm, G. L., Groen, I. I. A., & Baker, C. I. (2016). Making sense of real-world scenes. *Trends in Cognitive Sciences, 20*(11), 843–856, https://doi.org/10.1016/j.tics.2016.09.003.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9*(11):8, 1–13, https://doi.org/10.1167/9.11.8.

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.

McConkie, G. W., & Loschky, L. C. (2002). Perception onset time during fixations in free viewing. *Behavior Research Methods, Instruments, & Computers, 34*(4), 481–490, https://doi.org/10.3758/BF03195477.

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, 1–18, https://doi.org/10.1167/10.6.21.

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., & Einhäuser, W. (2015). A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences, 1339*(1), 82–96, https://doi.org/10.1111/nyas.12705.

Nuthmann, A., Einhäuser, W., & Schütz, I. (2017). How well can saliency models predict fixation selection in scenes beyond central bias? A new approach to model evaluation using generalized linear mixed models. *Frontiers in Human Neuroscience, 11*, 491, https://doi.org/10.3389/fnhum.2017.00491.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision, 10*(8):20, 1–19, https://doi.org/10.1167/10.8.20.

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

Nuthmann, A., Schütz, I., & Einhäuser, W. (2020). Salience-based object prioritization during active viewing of naturalistic scenes in young and older adults. *Scientific Reports, 10*, 22057, https://doi.org/10.1038/s41598-020-78203-7.

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier, https://doi.org/10.1016/B978-012375731-9/50045-8.

Potter, M. C. (1975). Meaning in visual search. *Science, 187*(4180), 965–966, https://doi.org/10.1126/science.1145183.

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

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems, 10*(4), 341–350, https://doi.org/10.1088/0954-898X/10/4/304.

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.

Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Review of Vision Science, 2*(1), 437–457, https://doi.org/10.1146/annurev-vision-082114-035733.

Rutishauser, U., & Koch, C. (2007). Probabilistic modeling of eye movement data during conjunction search via feature-based attention. *Journal of Vision, 7*(6):5, 1–20, https://doi.org/10.1167/7.6.5.

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.

Stoll, J., Thrun, M., Nuthmann, A., & Einhäuser, W. (2015). Overt attention in natural scenes: Objects dominate features. *Vision Research, 107*, 36–48, https://doi.org/10.1016/j.visres.2014.11.006.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136, https://doi.org/10.1016/0010-0285(80)90005-5.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology, 59*(11), 1931–1949, https://doi.org/10.1080/17470210500416342.

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.

Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition, 17*(1), 159–170, https://doi.org/10.1016/j.concog.2006.11.008.

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.

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

Wagenmakers, E.-J., Krypotos, A.-M., Criss, A. H., & Iverson, G. (2012). On the interpretation of removable interactions: A survey of the field 33 years after Loftus. *Memory & Cognition, 40*(2), 145–160, https://doi.org/10.3758/s13421-011-0158-0.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2d ed.). New York: Springer, https://doi.org/10.1007/978-3-319-24277-4.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association, 22*(158), 209–212, https://doi.org/10.1080/01621459.1927.10502953.

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

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.

Zhaoping, L., & Frith, U. (2011). A clash of bottom-up and top-down processes in visual search: The reversed letter effect revisited. *Journal of Experimental Psychology: Human Perception and Performance, 37*(4), 997–1006, https://doi.org/10.1037/a0023099.

Zhaoping, L., & Guyader, N. (2007). Interference with bottom-up feature detection by higher-level object recognition. *Current Biology, 17*(1), 26–31, https://doi.org/10.1016/j.cub.2006.10.050.

Appendix A: Mathematical definition of local contrast

As a measure of visual salience, the root-mean-square (RMS) contrast was calculated as

$$RMS(R, C) = \frac{1}{\bar{p}_{Img}} \sqrt{\frac{1}{\left((2L+1)^2 - 1\right)} \sum_{r=R-L}^{R+L} \sum_{c=C-L}^{C+L} (p(r, c) - \bar{p}(R, C))^2}$$

where L is either 11 (patch width 23) or 23 (patch width 47), $p(r,c)$ is the pixel value at row r and column c, $\bar{p}(R, C)$ is the mean of the patch calculated as

$$\bar{p}(R, C) = \frac{1}{(2L+1)^2} \sum_{r=R-L}^{R+L} \sum_{c=C-L}^{C+L} p(r, c)$$

and $\bar{p}_{Img}$ is the mean of the image.

Appendix B: Distribution of search targets in the scenes
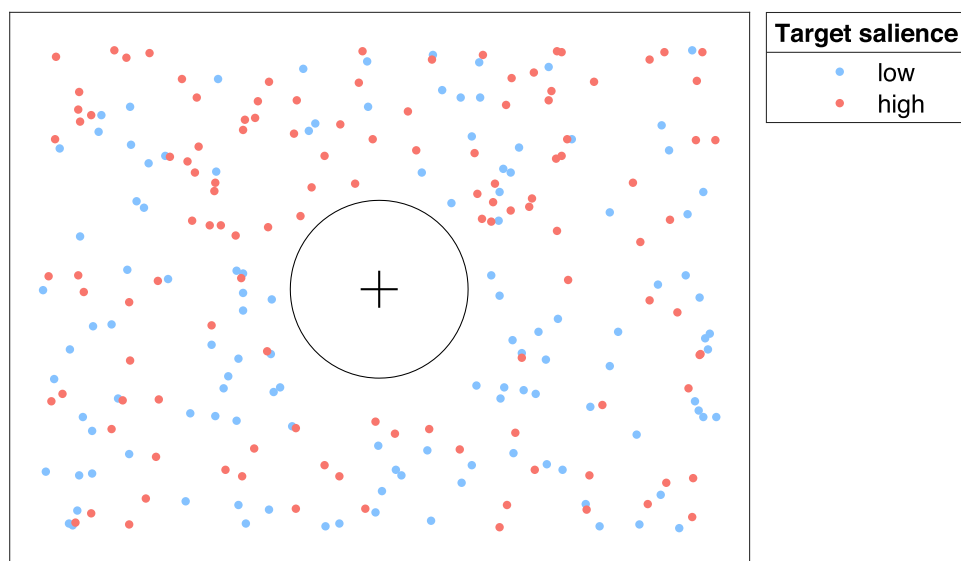

Figure B1. Positions of search targets in the 120 scenes used in Experiments 1 and 2. The light blue dots represent the positions of the low-salience targets, whereas the salmon dots represent the positions of the high-salience targets. The cross is the central fixation cross, and the circle with solid perimeter represents the central viewing area (radius 3°).

Appendix C: Mixed-model specification for Experiment 1

The mixed-model equation for the $2 \times 2 \times 2$ within-subjects design of Experiment 1 is

$$y_{si} = \beta_0 + b_{s0} + b_{i0} + \sum_{k=1}^{7} (\beta_k + b_{sk} + b_{ik}) x_{ksi} + e_{si}$$

where y denotes the dependent variable, β denotes fixed effects, b denotes random effects, and e denotes the residuals. The index s represents subjects ($1 \leq s \leq N_{sub}$), whereas the index i represents images/items ($1 \leq i \leq N_{items}$). The index k identifies the intercept and the contrasts: $k = 0$, intercept; $k = 1$, target size; $k = 2$, target salience; $k = 3$, foveal scotoma; $k = 4$, size $\times$ salience; $k = 5$, size $\times$ scotoma; $k = 6$, salience $\times$ scotoma; $k = 7$, size $\times$ salience $\times$ scotoma. For the maximal random effects structure, the variance-covariance matrix for by-subject random effects is given by

$$\Phi_s = \begin{pmatrix} var(b_{s0}) & cov(b_{s0}, b_{s1}) & cov(b_{s0}, b_{s2}) & cov(b_{s0}, b_{s3}) & cov(b_{s0}, b_{s4}) & cov(b_{s0}, b_{s5}) & cov(b_{s0}, b_{s6}) & cov(b_{s0}, b_{s7}) \\ cov(b_{s1}, b_{s0}) & var(b_{s1}) & cov(b_{s1}, b_{s2}) & cov(b_{s1}, b_{s3}) & cov(b_{s1}, b_{s4}) & cov(b_{s1}, b_{s5}) & cov(b_{s1}, b_{s6}) & cov(b_{s1}, b_{s7}) \\ cov(b_{s2}, b_{s0}) & cov(b_{s2}, b_{s1},) & var(b_{s2}) & cov(b_{s2}, b_{s3}) & cov(b_{s2}, b_{s4}) & cov(b_{s2}, b_{s5}) & cov(b_{s2}, b_{s6}) & cov(b_{s2}, b_{s7}) \\ cov(b_{s3}, b_{s0}) & cov(b_{s3}, b_{s1}) & cov(b_{s3}, b_{s2}) & var(b_{s3}) & cov(b_{s3}, b_{s4}) & cov(b_{s3}, b_{s5}) & cov(b_{s3}, b_{s6}) & cov(b_{s3}, b_{s7}) \\ cov(b_{s4}, b_{s0}) & cov(b_{s4}, b_{s1}) & cov(b_{s4}, b_{s2}) & cov(b_{s4}, b_{s3}) & var(b_{s4}) & cov(b_{s4}, b_{s5}) & cov(b_{s4}, b_{s6}) & cov(b_{s4}, b_{s7}) \\ cov(b_{s5}, b_{s0}) & cov(b_{s5}, b_{s1}) & cov(b_{s5}, b_{s2}) & cov(b_{s5}, b_{s3}) & cov(b_{s5}, b_{s4}) & var(b_{s5}) & cov(b_{s5}, b_{s6}) & cov(b_{s5}, b_{s7}) \\ cov(b_{s6}, b_{s0}) & cov(b_{s6}, b_{s1}) & cov(b_{s6}, b_{s2}) & cov(b_{s6}, b_{s3}) & cov(b_{s6}, b_{s4}) & cov(b_{s6}, b_{s5}) & var(b_{s6}) & cov(b_{s6}, b_{s7}) \\ cov(b_{s7}, b_{s0}) & cov(b_{s7}, b_{s1}) & cov(b_{s7}, b_{s2}) & cov(b_{s7}, b_{s3}) & cov(b_{s7}, b_{s4}) & cov(b_{s7}, b_{s5}) & cov(b_{s7}, b_{s6}) & var(b_{s7}) \end{pmatrix}$$

where var(.) denotes the variance and cov(.,.) the covariance matrix. Similarly, the variance-covariance matrix for by-item random effects is given by

$$\Phi_i = \begin{pmatrix} var(b_{i0}) & & & \\ cov(b_{i1}, b_{i0}) & var(b_{i1}) & & \\ \vdots & \vdots & \ddots & \\ cov(b_{i7}, b_{i0}) & cov(b_{i7}, b_{i1}) & \cdots & var(b_{i7}) \end{pmatrix}.$$

In summary, the fixed-effects structure includes seven contrasts (three main effects, three two-way interactions, one three-way interaction). Consequently, the maximal random-effects structure would require estimating 72 parameters (by subject: random intercept, 7 random slopes, 28 correlation terms; by item: same as by subject).