

1
2 **The Effect of Target Saliency and Size in Visual Search within Naturalistic Scenes**
3 **under Degraded Vision**

4
5 Antje Nuthmann^{1,2}, Adam C. Clayden^{3,2} and Robert B. Fisher⁴

6 ¹ Institute of Psychology, University of Kiel, Germany

7 ² Psychology Department, School of Philosophy, Psychology and Language Sciences,
8 University of Edinburgh, UK

9 ³ School of Engineering, Arts, Science and Technology, University of Suffolk, UK

10 ⁴ School of Informatics, University of Edinburgh, UK

11
12 **Acknowledgments**

13
14 The project was supported by the Engineering and Physical Sciences Research
15 Council (UK). Portions of this research were presented at the Bristol Vision Institute Young
16 Vision Researchers' Colloquium (Cardiff University, UK, 2016), at the 40th European
17 Conference on Visual Perception (Berlin, Germany, 2017), and at the Workshop "From
18 peripheral to transsaccadic and foveal perception" (Castle Rauischholzhausen, Germany,
19 2019).

20 Correspondence concerning this article should be addressed to Antje Nuthmann,
21 University of Kiel, Institute of Psychology, Olshausenstr. 62, 24118 Kiel, Germany. Email:
22 nuthmann@psychologie.uni-kiel.de

23 Antje Nuthmann  <http://orcid.org/0000-0003-3338-3434>

24

Abstract

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 sub-processes 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.

248 words

Keywords: naturalistic scenes; visual search; visual salience; eye movements; simulated scotomas

51 **1 Introduction**

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

62 It is widely agreed that eye movements in naturalistic scenes are controlled by both
63 bottom-up (stimulus-driven) and top-down (task-driven, context-driven, or goal-driven)
64 factors (Malcolm, Groen, & Baker, 2016). Research on bottom-up control has been
65 dominated by salience-driven approaches, in which a saliency map is computed using low-
66 level image features to guide task independent gaze allocation (Borji & Itti, 2013; Borji,
67 Sihite, & Itti, 2013a for reviews). The first computational model of this kind was Itti, Koch,
68 and Niebur's (1998) implementation of Koch and Ullman's (1985) computational
69 architecture based on the Feature Integration Theory (FIT, Treisman & Gelade, 1980). FIT
70 explains human behavior in visual search tasks involving covert shifts of attention. Extending
71 this research, the saliency model was introduced as a model of covert and overt orienting in
72 search (Itti & Koch, 2000; Itti et al., 1998). According to simulations by Itti and Koch (2000),
73 the saliency model performed similarly to, or better than, human searchers looking for
74 oriented lines amongst distractor lines or for a camouflaged tank in a natural environment.
75 Still, when observers are given a visual search task (or a task altogether), top-down
76 influences on attention and eye guidance are often believed to dominate (Koehler, Guo,
77 Zhang, & Eckstein, 2014).

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

84 One of these search tasks required observers to indicate whether or not there was a
85 piece of fruit in the scene (Underwood et al., 2006). If present, the piece of fruit was always a
86 low-saliency object, according to the saliency model by Itti and Koch (2000). Some of the
87 scenes also included a high-saliency object, which served as a distractor. There was little
88 attentional capture by the distractor. However, when there was a high-saliency distractor
89 present, then the low-saliency target was fixated later than when it was absent, and near
90 distractors were more disruptive than those furthest from the target. The authors concluded
91 that the purpose of inspection can provide a cognitive override that renders visual saliency
92 secondary. The key finding that the most salient region is neglected in favor of a completely
93 non-salient target was replicated in a subsequent study by different authors (Henderson et al.,
94 2009).

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

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

112 Foulsham and Underwood (2011) used a slightly different approach: rather than
113 manipulating scenes and objects, they used the predictions of the saliency model by Itti and
114 Koch (2000) to select target regions that were either salient or non-salient. As would be
115 predicted by the saliency model, behavioral search times were shorter for highly salient
116 regions than either low-salient regions or control regions. Control regions and low-salient
117 regions did not differ reliably. Interestingly, saliency did not affect the process of localizing

118 the target region in space, as indexed by the latency to first fixation on the region. This
119 implies that the subsequent verification process (is this the target?) took longer when the
120 region was low in salience, and that this effect was large enough to affect total search time. In
121 a second experiment, peripheral filtering of low-level features was expected to modify the
122 effect of target saliency on search, but this was not the case (Foulsham & Underwood, 2011).

123 The main problem with identifying the causal contribution of visual salience to gaze
124 guidance is an inherent correlation with higher-order factors such as objects and semantics
125 (Henderson, Brockmole, Castelano, & Mack, 2007; Nuthmann & Henderson, 2010; Stoll,
126 Thrun, Nuthmann, & Einhäuser, 2015). In the studies reviewed above, effects of salience
127 were assessed between different objects or scenes, which potentially introduces additional
128 confounds. To address these issues, we used context-free letter targets rather than
129 contextually relevant search targets. In two experiments, observers searched for a black letter
130 “T” embedded in grayscale photographs of real-world scenes. We used our Target
131 Embedding Algorithm (T.E.A., Clayden, Fisher, & Nuthmann, 2020)¹ to generate within-
132 scene manipulations of target salience (low vs. high) and—in Experiment 1—also target size
133 (small vs. large). Our approach minimizes any confounding effects that may arise from
134 various forms of scene guidance (semantic, syntactic, and episodic guidance; Biederman,
135 Mezzanotte, & Rabinowitz, 1982; Henderson & Ferreira, 2004). Specifically, using context-
136 free targets prevents observers from using their knowledge about the likely positions of
137 targets to guide their attention and eye movements. Moreover, by inserting the targets in an
138 algorithmic manner via image processing techniques, we also minimized artefacts that may
139 otherwise occur due to post hoc editing of scenes.

140 Saliency maps translate physical properties of the stimulus such as luminance,
141 orientation, color, and size into saliency values. Since these stimulus dimensions have
142 different characteristics, combining them is a non-trivial problem (Itti & Koch, 1999). The
143 size feature is typically accounted for in an implicit manner by incorporating multiple spatial
144 scales of processing. In this way, saliency models attempt to account for size over image
145 regions and not over objects, which is a limiting factor of this approach (Borji, Sihite, & Itti,
146 2013b). Borji et al. addressed this issue by asking observers to choose which object (out of
147 two in a given image) stands out the most based on its low-level features. Both saliency and
148 object size were important for selecting the object. Observers’ judgments were well described
149 by a linear combination of the two variables in an integrated model of saliency and object

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

150 size. Moreover, previous investigations of object-based selection in scenes found independent
151 effects of object size and object-based salience on fixation probability, with large objects and
152 highly salient objects being more frequently selected for fixation (Nuthmann, Schütz, &
153 Einhäuser, 2020; Stoll et al., 2015). Regarding visual search, in previous work we
154 manipulated target size whilst controlling for target salience by probing the scene for
155 locations of median salience (Clayden et al., 2020). In these experiments, we observed better
156 search performance for larger targets. Extending this research, we designed Experiment 1 to
157 assess the independent contributions of target salience and target size, as well as their
158 interaction.

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

171 In visual search, guidance by basic features can be bottom-up or top-down (Wolfe,
172 2015). Bottom-up guidance is stimulus-driven, based on local differences. Here, we tested the
173 independent and combined effects of target salience and size during active eye-movement
174 search. Top-down guidance is user-driven, based on the observer's understanding of the task.
175 In our experiments, on each trial participants were asked to look for the letter "T". Given that
176 letters are overlearned categories, observers were expected to use top-down guidance to
177 deploy attention to the target.

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

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

199 **2 General Method**

200 **2.1 Participants**

201 Thirty-two participants (10 males) between the ages of 18 and 27 (mean age 21 years)
202 participated in Experiment 1. Thirty-six participants (7 males) between the ages of 18 and 27
203 (mean age 21 years) participated in Experiment 2. All participants had normal or corrected-
204 to-normal vision by self-report. They gave their written consent prior to the experiment and
205 either received study credit or were paid at a rate of £7 per hour for their participation, which
206 lasted about one hour. The experiments were approved by the Psychology Research Ethics
207 Committee of the University of Edinburgh and conformed to the Declaration of Helsinki.

208 **2.2 Apparatus**

209 Working with gaze-contingent displays requires minimizing the latency of the system
210 (Loschky & Wolverson, 2007; Saunders & Woods, 2014). Moreover, gaze-contingent
211 manipulations of foveal vision call for eye-tracking equipment with high spatial accuracy and
212 precision (Geringswald, Baumgartner, & Pollmann, 2013). Participants' eye movements were
213 recorded binocularly with an SR Research EyeLink 1000 Desktop mount system with high
214 accuracy (0.15° best, 0.25-0.5° typical) and high precision (0.01° RMS). The EyeLink 1000
215 was equipped with the 2000 Hz camera upgrade, allowing for binocular recordings at a

216 sampling rate of 1000 Hz per eye. Stimuli were presented on a 21-inch CRT monitor with a
217 refresh rate of 140 Hz at a viewing distance of 90 cm, taking up a $24.8^\circ \times 18.6^\circ$ (width \times
218 height) field of view. A chin and forehead rest was used to keep the participants' head
219 position stable.

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

225 **2.3 Stimulus Materials**

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

229 The search target was always the letter "T", which was inserted into the scene by
230 using the Target Embedding Algorithm (T.E.A.) introduced by Clayden et al. (2020).
231 Specifically, the T was inserted in sans-serif style; that is, consisting of two bars. The
232 dimensions of these two bars are parameterized by length and width. For the small target
233 letter, the horizontal bar was 13 pixels in length and two pixels in width, whereas the vertical
234 bar was 16 pixels in length and three pixels in width. For the large target letter, the horizontal
235 bar was 33 pixels in length and four pixels in width, whereas the vertical bar was 40 pixels in
236 length and five pixels in width.

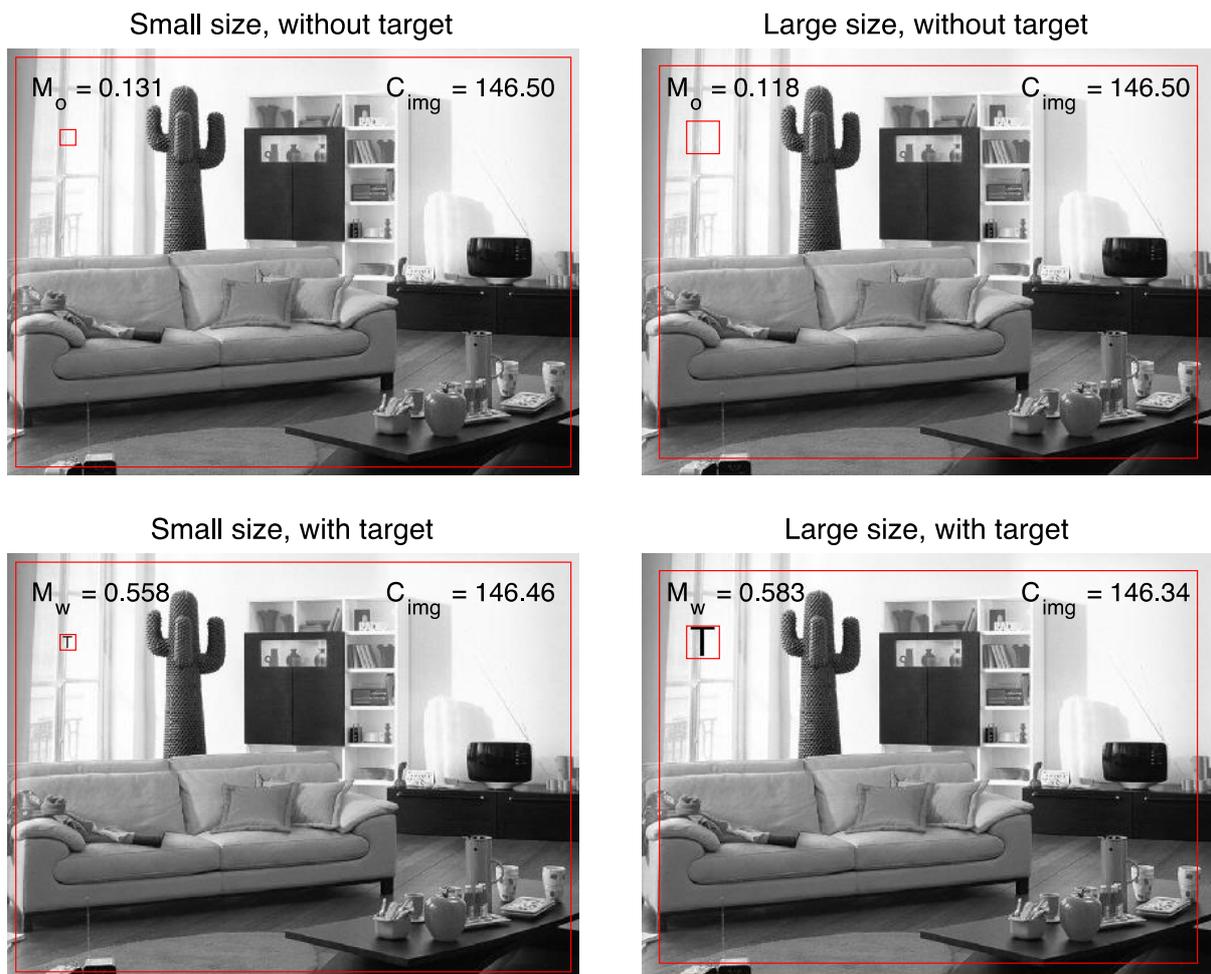
237 To determine suitable positions for low- and high-salience targets, we inserted the T
238 into every possible location of the original scene image and calculated how much it would
239 stand out from the scene background. To this end, a rectangular region that was slightly
240 larger than the target was moved pixel-by-pixel through the image. Using the larger
241 dimension of the target letter (i.e., its height) as a reference, the region's size was determined
242 by adding a constant buffer of 3 pixels to either side (plus one pixel to center the region on
243 the current position). As a result, the region size was 23×23 pixels for small target letters
244 and 47×47 pixels for large target letters.

245 As a measure of visual salience, we used a version of root-mean-square (RMS)
246 contrast: the standard deviation of luminance values of all pixels in the evaluated region was
247 divided by the mean luminance of the image (Bex & Makous, 2002; Nuthmann & Einhäuser,
248 2015; Reinagel & Zador, 1999). First, the RMS contrast M_o was calculated for the evaluation

249 box at each position in the image, see Appendix A for the mathematical details of the
 250 calculations. Next, the black target letter was inserted at a given position by replacing pixel
 251 values of the original image by the pixel values of the target. Following target insertion, the
 252 RMS contrast M_w was computed for the evaluation box comprising the T. Afterwards, the
 253 contrast change value $\Delta C = M_w - M_o$ was computed to quantify the visual salience of the
 254 target letter at a given location within the scene.

255 To provide an example, in Figure 1 the evaluation box is centered on image position
 256 $(r,c) = (125, 85)$, with (r,c) denoting the rows and columns of the image. For the large target,
 257 we obtain $M_o = 0.118$ and $M_w = 0.583$, with $\Delta C = 0.464$. Thus, adding a black T to a
 258 relatively bright region of the image leads to a relatively large change in local contrast. For
 259 the example image used in Figures 1 and 2, our GitHub page (see footnote 1) shows a
 260 dynamic visualization of the contrast calculations for all possible target positions.

261



262

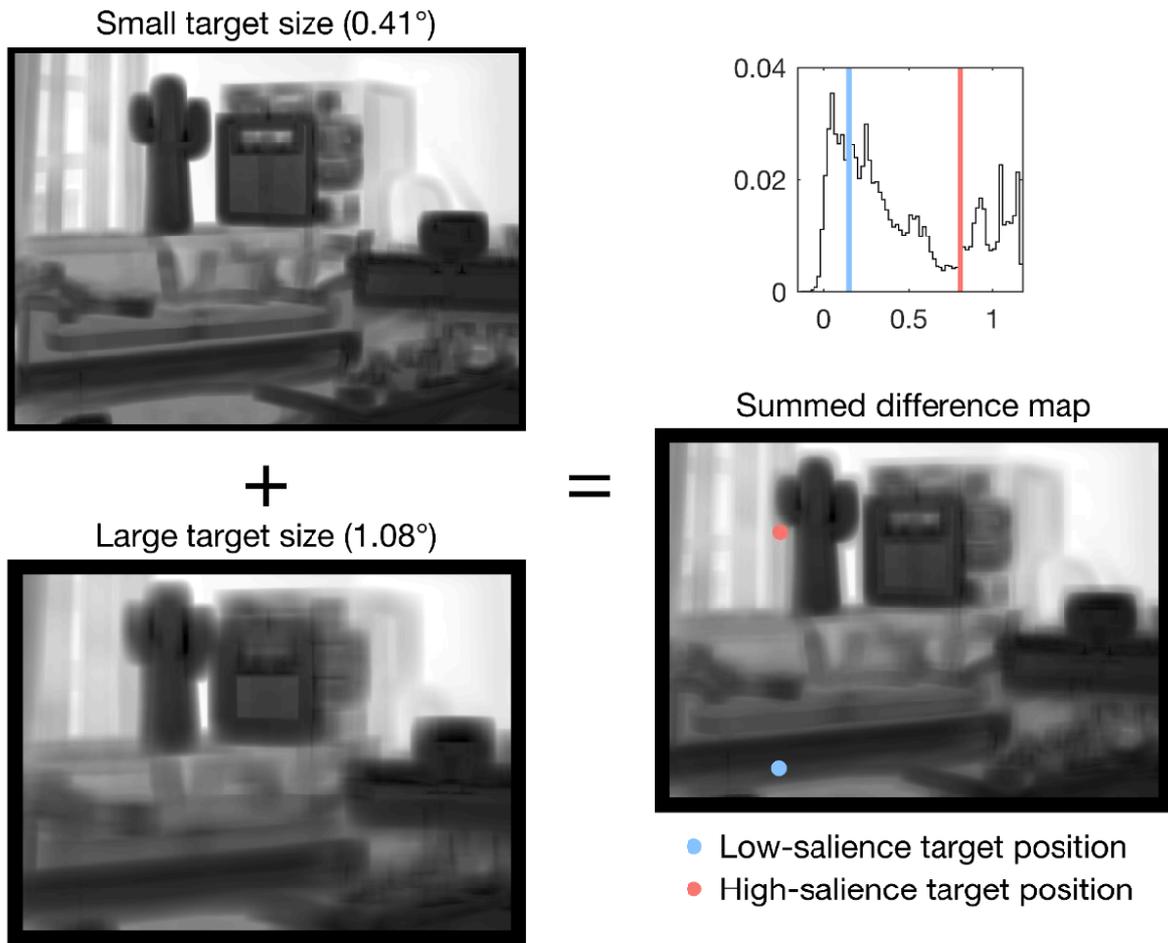
263 Figure 1. Target Embedding Algorithm. In this example, the squared evaluation box (in red)
 264 is positioned at $(r,c) = (125, 85)$ in all panels. The local RMS contrast is calculated both
 265 without the target letter (M_o , top row) and with the target letter inserted (M_w , bottom row), for

266 both the small target (left column) and the large target (right column). C_{img} denotes the mean
267 luminance of the image, without the target letter (top row) or with (bottom row). The outer
268 rectangle (in red) marks the region of the image border that was not considered for target
269 insertion.

270

271 Calculating ΔC at each pixel in the image yields a map comprising of the contrast
272 difference values within the image. The contrast difference map was calculated separately for
273 small and large targets (Figure 2). Afterwards, the two resultant maps were summed together.
274 This allowed us to compute a single location for both target sizes, as the values of the two
275 difference maps varied slightly. The summed difference map was then probed by our
276 algorithm to locate suitable pixel (i.e., potential target) positions. The criteria for choosing the
277 low and high salient regions were the lower and upper quartile changes in local contrast when
278 inserting the letter into the scene. If the exact value for the lower or upper quartile of the
279 distribution did not exist in the summed contrast difference map, the closest existing value
280 was used. Candidate locations were tested against two exclusion criteria (Clayden et al.,
281 2020). In the experiments, participants started their search at the center of the scene, with a
282 foveal or central scotoma blocking their view on many of the trials. Therefore, locations
283 within 3° from the center were excluded. To avoid truncation of the letter, locations at image
284 boundaries were also excluded (Figure 1, Figure 2). If there was more than one possible
285 target location left, one was selected at random as the location of the target for that scene
286 image and salience condition. The resulting distributions of target positions reveal broad
287 coverage (Figure B1). For further validation, each target's eccentricity was calculated as the
288 Euclidean distance between target position and image center. Mean eccentricities did not
289 differ for low- and high-salience targets, $t(119) = -0.32, p = .746$.

290



291

292 Figure 2. Algorithmic target placement at low and high salient regions within the scene. Left:

293 Contrast difference maps for the small and large target size used in Experiment 1, for the

294 example scene used in Figure 1 and Figure 7. Right: Summed contrast difference map

295 (bottom) and the distribution of the map's values (top). Two vertical lines were added to the

296 histogram to mark the lower (light blue) and upper (salmon) quartiles of the distribution.

297 These values were used to determine suitable positions for low- and high-salience targets in

298 the scene image. For the example image, the final target positions are marked with colored

299 dots in the summed contrast difference map. For visualization purposes, the values of a given

300 map were scaled to the same range (i.e., to [0,1]).

301

302 2.4 Creation of Gaze-Contingent Scotomas

303 In Experiment 1, we implemented a *foveal* scotoma; in Experiment 2, we contrasted a

304 *central* scotoma with a *peripheral* scotoma. For the foveal and central scotomas, we used a

305 gaze-contingent technique that was originally introduced by Rayner and Bertera (1979) for

306 sentence reading. The authors referred to their implementation as *moving mask*; other terms

307 include *simulated scotoma* (Bertera, 1988). When applied to scene viewing, the moving mask
308 paradigm is analogous to viewing the scene with a “blindspot”: information in the center of
309 vision is blocked from view, while information outside the window is unaltered (Miellet,
310 Zhou, He, Rodger, & Caldara, 2010; Nuthmann, 2014). As in our previous study (Clayden et
311 al., 2020), the foveal scotoma in Experiment 1 was a symmetric circular gray mask with a
312 radius of 1° to completely obscure foveal vision (see Figure 3 below). The central scotoma
313 (Experiment 2) had a radius of 2.5° , thereby eliminating both foveal and part of parafoveal
314 vision (Figure 7b below). For the peripheral scotoma (Experiment 2), we used the gaze-
315 contingent *moving window* technique (McConkie & Rayner, 1975, for reading). Applied to
316 scene viewing, the moving window paradigm is analogous to viewing the scene through a
317 “spotlight”: a defined region in the center of vision contains unaltered scene content, whilst
318 the scene content outside the window is blocked from view (Caldara, Zhou, & Miellet, 2010;
319 Nuthmann, 2014). Our central and peripheral scotomas had equal radii (2.5°) and so were
320 inverse manipulations of one another (Figure 7 below). To avoid sharp-boundary scotomas,
321 the perimeter of the gray circular mask or window was slightly faded through low-pass
322 filtering (Clayden et al., 2020).

323 The general idea underlying our scotoma implementation is to mix a foreground
324 image and a background image via a mask image (van Diepen, De Graef, & Van Rensbergen,
325 1994). The foreground image is formed by the experimental stimulus; that is, by the current
326 scene image. The background image defines the content of the masked area. In the present
327 experiments, the background image was a monochrome image (gray, RGB-value: 128, 128,
328 128), which implies that the moving scotomas were drawn in that color (Clayden et al.,
329 2020). The mask image defines the type, shape, and size of the gaze-contingent scotoma. It
330 was a normalized grayscale image, where pixel values of 255 (white) represent portions of
331 the foreground image that show through while values of 0 (black) are masked and therefore
332 replaced by the corresponding background image pixels. For the foveal and the central
333 scotoma, a circular 0-center, 255-surround map formed the mask. For the peripheral scotoma,
334 an inverted mask was used; that is, a circular 255-center, 0-surround map. To avoid sharp-
335 boundary scotomas, the perimeter of the circular mask or window was slightly faded through
336 low-pass filtering (Clayden et al., 2020).

337 To minimize the latency of the measurement system, we used an eye tracker with a
338 binocular sampling rate of 1000 Hz and fast online access of new gaze samples. Specifically,
339 the eye tracker computed a new gaze position every millisecond and made it available in less
340 than 2 ms. Moreover, PTB-3 for MATLAB offers fast creation of gaze-contingent scotomas

341 using texture-mapping and OpenGL (Open Graphics Library). This technique provides
342 various blending operations that enable image combinations to take place via an image's
343 alpha channel (see Duchowski & Çöltekin, 2007, for details on the general technique). The
344 mask image served as the alpha mask for blending of the foreground and background images.
345 To obtain a composite rendering of the scene image with the scotoma, three textures were
346 created—for the foreground image, background image, and mask image, respectively. During
347 the search trial, the center of the mask texture was translated to the coordinates of the current
348 gaze position. Thus, gaze contingency was realized by moving the mask across the stimulus.
349 This solution avoids the need for computationally expensive real-time image synthesis.

350 Since scene images typically occupy the entire monitor space, a full refresh cycle is
351 required to update the screen. In the experiments, the stimuli were displayed on a 140-Hz
352 CRT monitor, which means that it took 7.14 ms for one refresh cycle to complete.
353 Throughout the experimental trial, gaze position was continuously evaluated online. The
354 algorithm first checked whether new valid binocular gaze samples were available. If that was
355 the case, the center of the mask was re-aligned with the average horizontal and vertical
356 position of the two eyes (Nuthmann, 2013, for discussion). Even with a state-of-the-art
357 system, small temporal delays in updating the display contingent on the participant's gaze are
358 unavoidable. Any mismatch between gaze position and scotoma position that may result
359 should be largest during a saccade and right after a saccade. However, observers are blind to
360 mismatches during this period, due to saccadic suppression and the time needed for
361 perception to be restored (McConkie & Loschky, 2002).

362 **2.5 Procedure**

363 At the beginning of the experiment, the eye tracker was calibrated using a series of
364 nine fixed targets distributed around the display, followed by a 9-point accuracy test. At the
365 start of each trial, a fixation cross was presented at the center of the screen for 600 ms and
366 acted as a fixation check. The fixation check was judged successful if gaze position, averaged
367 across both eyes, consistently remained within an area of 40×40 pixels ($1.24^\circ \times 1.24^\circ$) for
368 200 ms. If this condition was not met, the fixation check timed out after 500 ms. In this case,
369 the fixation check procedure was either repeated or replaced by another calibration
370 procedure. If the fixation check was successful, the scene image appeared on the screen.
371 Once subjects had found the target letter, they were instructed to fixate their gaze on it and
372 press a button on the controller to end the trial (cf. Clayden et al., 2020; Glaholt, Rayner, &
373 Reingold, 2012; Nuthmann, 2014). Trials timed-out 15 s after stimulus presentation if no

374 response was made. There was an inter-trial interval of 1 s before the next fixation cross was
375 presented.

376 **2.6 Data Analysis**

377 The SR Research Data Viewer software with default settings was used to convert the
378 raw data obtained by the eye tracker into a fixation sequence matrix. Data from the right eye
379 were analyzed. The behavioral and eye-movement data were further processed and analyzed
380 using the R system for statistical computing (R Development Core Team). Figures were
381 created using MATLAB (Figures 1 – 3 and 7) or the *ggplot2* package (version 3.2.1;
382 Wickham, 2016) supplied in R (remaining figures). The T.E.A. was programmed in
383 MATLAB.

384 Analyses of fixation durations and saccade lengths excluded fixations that were
385 interrupted with blinks. Analysis of fixation durations disregarded the initial, central fixation
386 in a trial. However, its duration was analyzed separately as search initiation time. The button
387 press terminating the search took place during the last fixation in a trial. Therefore, the last
388 fixation was also excluded from analysis of fixation durations. However, its duration
389 contributed to the measurement of verification time. Fixation durations that are very short or
390 very long are typically discarded, based on the assumption that they are not determined by
391 on-line cognitive processes (Inhoff & Radach, 1998). In the present study, this precaution
392 was not followed because the presence of a foveal scotoma may affect eye movements (e.g.,
393 fixations were predicted to be longer than normal).

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

404 **2.7 Statistical Analysis using Mixed Models**

405 We used linear mixed-effects models (LMM) for analyzing continuous response

406 variables, specifically search time and its three subcomponents, saccade amplitude, and
407 fixation duration. Search accuracy was analyzed using binomial generalized linear mixed-
408 effects models (GLMM). A technical introduction to both types of mixed models is provided
409 by Demidenko (2013). The analyses were conducted with the R package *lme4* (version 1.1.-
410 23; Bates, Maechler, Bolker, & Walker, 2015). Separate (G)LMMs were estimated for each
411 dependent variable.

412 Search accuracy was assessed through a binary variable; in a given trial, the search
413 target was correctly located (1) or not (0). In the GLMM, the resulting probabilities were
414 modeled through a link function (Bolker et al., 2009). For binary data, there are three
415 common choices for link functions: logit, probit, and complementary log-log (Demidenko,
416 2013). For our analyses we used the logit transformation of the probability, which is the
417 default for the *glmer* function in the R package *lme4*. Thus, in a binomial GLMM parameter
418 estimates are obtained on the log-odds or logit scale, which is symmetric around zero,
419 corresponding to a probability of 0.5, and ranges from negative to positive infinity (Jaeger,
420 2008).

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

428 For Experiment 2, simple coding was used for the 2-level factor target salience. For
429 the 3-level factor scotoma type, contrasts were chosen such that they tested hypotheses about
430 the expected pattern of means. More generally, the different scotomas were expected to affect
431 overall task difficulty, which may lead to differences in search performance and global eye
432 movement measures. For example, search times were expected to be longest for search with a
433 peripheral scotoma. In this case, factor levels were ordered accordingly (no scotoma, central
434 scotoma, peripheral scotoma) and backward difference (BWD) coding (also known as sliding
435 differences or repeated contrasts) was used to compare the mean of the dependent variable
436 for one level of the ordered factor with the mean of the dependent variable for the prior
437 adjacent level (Venables & Ripley, 2002). Moreover, we reasoned that a specific type of
438 scotoma may selectively impair a specific sub-process of search. To test these more specific
439 hypotheses, simple coding was used. The no-scotoma control condition served as reference

440 level, which allowed us to test whether there were any differences between the central
441 scotoma and the control condition or between the peripheral scotoma and the control
442 condition. Simple coding and backward difference coding yield centered contrasts, in which
443 case the model intercept reflects the grand mean of the dependent variable.

444 The mixed models included subjects and scene items as crossed random factors. The
445 overall mean for each subject and scene item was estimated as a random intercept. In
446 principle, the variance-covariance matrix of the random effects not only includes random
447 intercepts but also random slopes, as well as correlations between intercepts and slopes (Barr,
448 Levy, Scheepers, & Tily, 2013). Random slopes estimate the degree to which each fixed
449 effect varies across subjects and/or scene items. For example, the by-item random slope for
450 salience captures whether scene items vary in the extent to which target salience affects
451 search performance and/or eye-movement parameters (see Nuthmann, Einhäuser, & Schütz,
452 2017, for an example).

453 To select an optimal random-effects structure for (G)LMMs, we pursued a data-
454 driven approach using backward model selection. To minimize the risk of Type I error, we
455 started with the maximal random-effects structure justified by the design (Barr et al., 2013).
456 For Experiment 1, where the same contrast coding was used for all dependent variables, the
457 maximal variance-covariance matrix of the random effects is provided in Appendix C. Across
458 experiments, none of these maximal models converged (maximal number of iterations: 10^6).
459 For LMMs, the maximal random-effects structure was backwards-reduced using the *step*
460 function of the R package *lmerTest* (version 3.1-2; Kuznetsova, Brockhoff, & Christensen,
461 2017). If the final fitted model returned by the algorithm had convergence issues, we
462 proceeded to fit zero-correlation parameter (zcp) models in which the random slopes are
463 retained but the correlation parameters are set to zero (Matuschek, Kliegl, Vasissth, Baayen,
464 & Bates, 2017; Seedorff, Oleson, & McMurray, 2019). The full random-effects structure of
465 the zcpLMM required 16 (Experiment 1) and 12 (Experiment 2) variance components to be
466 estimated. This random-effects structure was evaluated and backwards-reduced to arrive at
467 the model that was justified by the data.

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

471 For parameter optimization, the bobyqa optimizer was used for LMMs, and a
472 combination of Nelder-Mead and bobyqa for GLMMs. LMMs were estimated using the
473 restricted maximum likelihood criterion. GLMMs were fit by Laplace approximation. For the

474 coded contrasts, coefficient estimates (b) and their standard errors (SE) along with the
475 corresponding t -values (LMM: $t = b/SE$) or z -values (GLMM: $z = b/SE$) are reported. For
476 GLMMs, p -values are additionally provided. For LMMs, a two-tailed criterion ($|t| > 1.96$)
477 was used to determine significance at the alpha level of .05 (Baayen, Davidson, & Bates,
478 2008).

479 In the (G)LMM analyses, data from individual trials (subject–item combinations)
480 were considered. For the data depicted in Figures 4, 5, 8, and 9, means were calculated for
481 each subject, and these were then averaged across subjects. Result figures display the data on
482 their original scale. When using the T.E.A. to prepare the stimulus material, for one of the
483 photographs the different versions were not saved into the correct folders on the lab computer
484 due to human error. For three additional scenes, participants had difficulty finding the low-
485 salience target. These four scenes were therefore excluded from analysis.

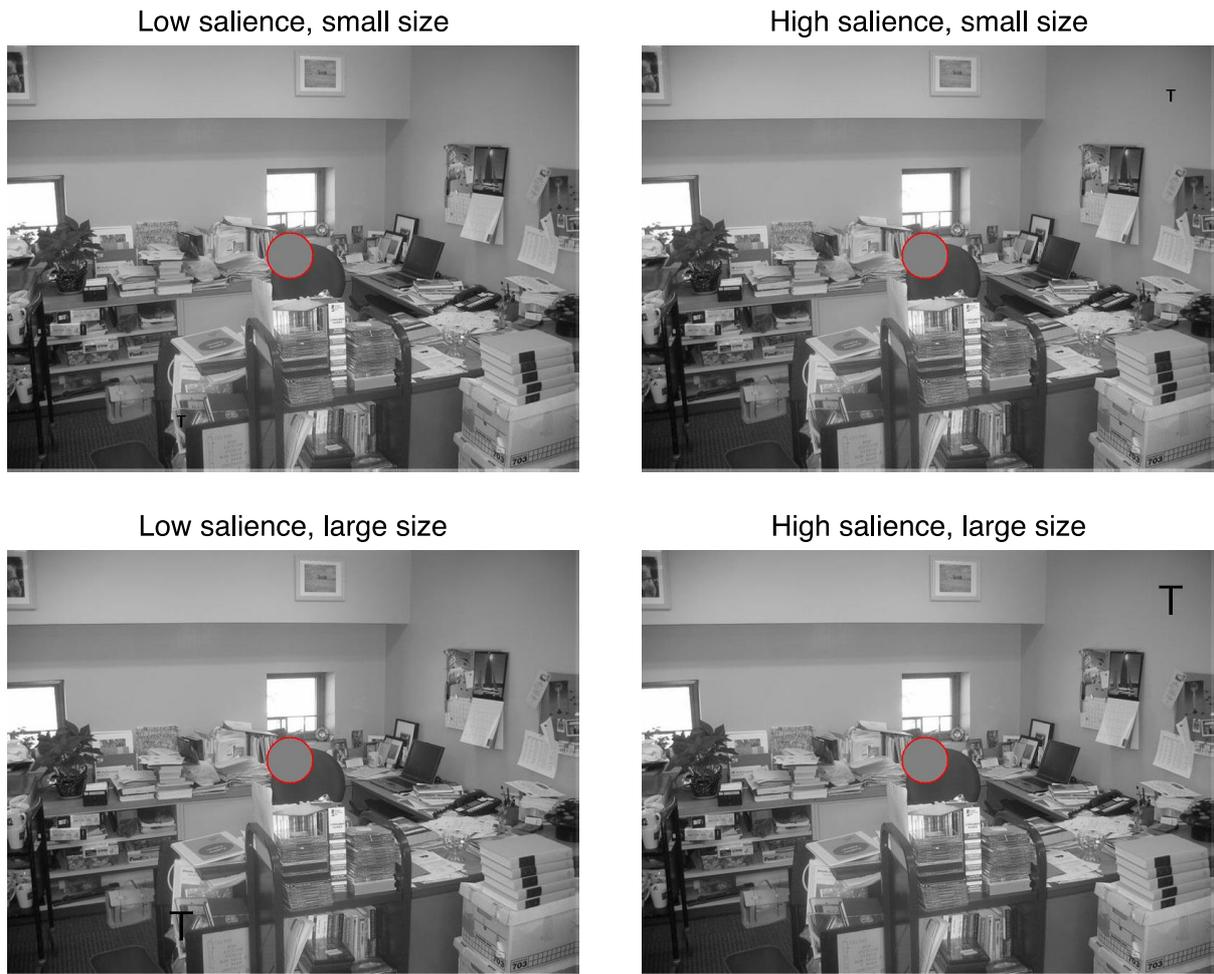
486 **3 Experiment 1**

487 **3.1 Design**

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

496

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.



497

498 Figure 3. Four foveal-scotoma conditions for one of the scenes used in Experiment 1. Left
 499 column: low-salience targets, right column: high-salience targets; top row: small targets,
 500 bottom row: large targets. The gray disk in the center of the image is the foveal mask that
 501 moved concomitantly with the participant’s gaze. In the figure, the foveal scotoma is
 502 highlighted with a red circle. In the experiment, each observer searched each scene in one of
 503 the size \times salience conditions only, either with or without a simulated foveal scotoma.

504

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

512 The visual field manipulation was blocked so that participants completed two blocks
 513 of trials in the experiment: in one block observers’ foveal vision was available, in the other

514 block it was obstructed by a simulated foveal scotoma. Each block started with four practice
515 trials, one for each target salience \times size condition. The order of blocks was counterbalanced
516 across subjects. Within a block, scenes were presented randomly.

517 **3.2 Results**

518 In a first step, we analyzed different measures of search accuracy as indicators of
519 search efficiency. For correct trials, we then analyzed search time and its subcomponents.
520 Finally, we examined saccade amplitude and fixation duration across the viewing period.

521 **3.2.1 Search accuracy**

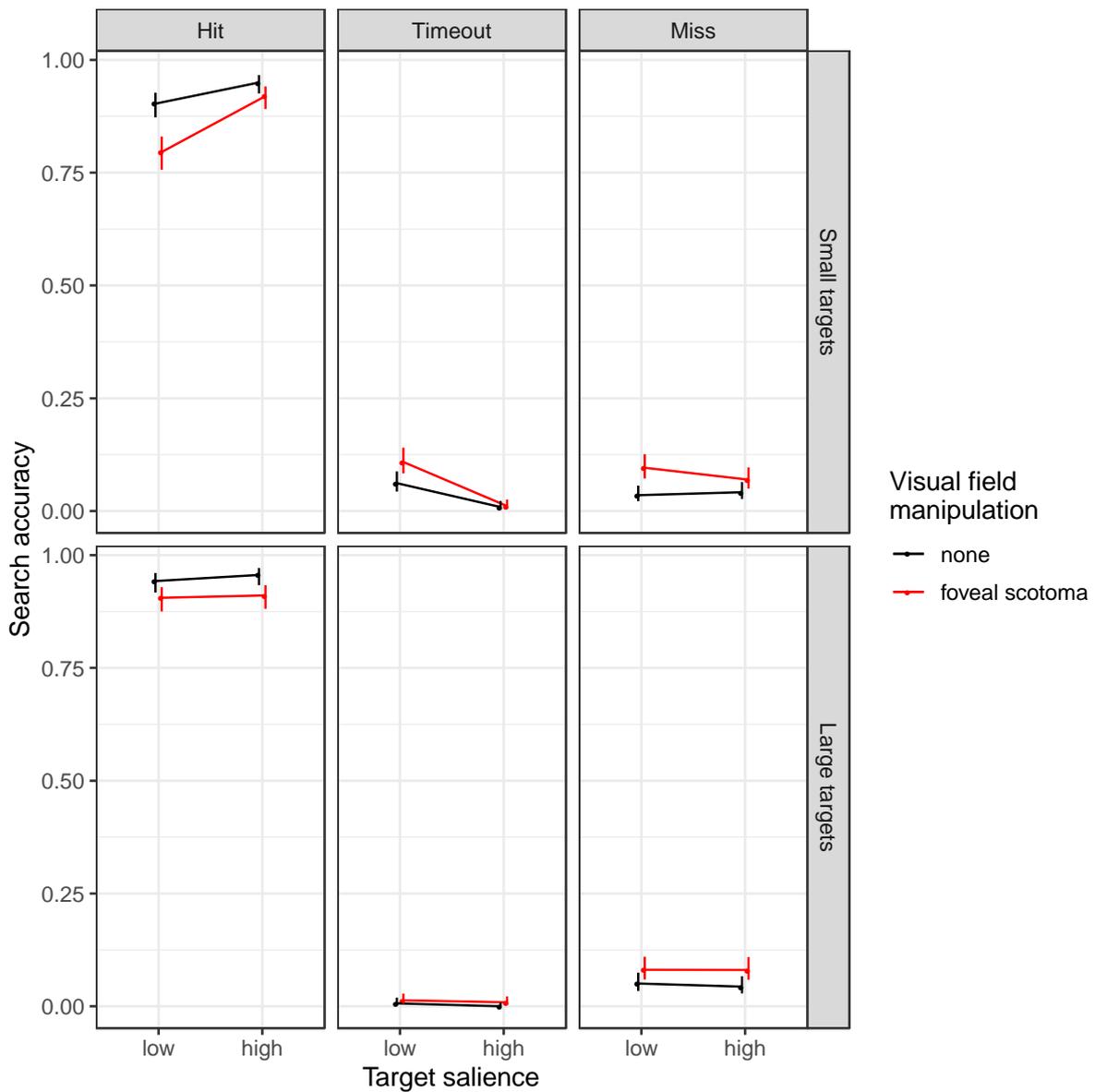
522 The first set of analyses examined the likelihood of finding the target letter in the
523 scene. Performance for each experimental condition was divided into probabilities of “hit,”
524 “miss,” and “timeout” cases (Clayden et al., 2020; Nuthmann, 2014). If the participant had
525 not responded within 15 s, the trial was coded as a “timeout.” A response was scored as a
526 “hit” if the participant indicated to have located the target by button press and their gaze was
527 within the rectangular area of interest (AOI) comprising the target; otherwise, the response
528 was scored as a “miss.” The AOI was $2.9^\circ \times 2.9^\circ$ in size (Clayden et al., 2020). It was the
529 same for both target sizes and included a buffer, following recommendations by Holmqvist
530 and Andersson (2017).

531 There was a significant effect of scotoma on the probability of “hitting” the target
532 such that participants were less likely to correctly locate and accept the target when foveal
533 vision was not available, $b = -0.70$, $SE = 0.13$, $z = -5.49$, $p < .001$ (Figure 4, left column).
534 Moreover, search accuracy was significantly higher for large as compared to small targets, b
535 $= 0.41$, $SE = 0.13$, $z = 3.22$, $p = 0.001$, and it was higher for high-salience compared to low-
536 salience targets, $b = 0.56$, $SE = 0.13$, $z = 4.43$, $p < .001$. Only one of the interactions was
537 significant (Table 1). Specifically, there was a significant size \times salience interaction, $b = -$
538 0.73 , $SE = 0.25$, $z = -2.87$, $p = 0.004$, indicating that the salience effect was smaller for large
539 as compared to small targets. As a matter of fact, the data displayed in Figure 4 suggest that
540 the effect of one variable was absent for the easier condition of the other variable. To test this
541 explicitly, we specified a post-hoc GLMM using dummy coded variables with the following
542 reference levels: large targets, high-salience targets, foveal scotoma. The simple effect for
543 target size, representing the size effect for high-salience targets, was not significant, $b = 0.10$,
544 $SE = 0.24$, $z = 0.41$, $p = 0.685$. The simple effect for target salience, representing the salience
545 effect for large targets, was also not significant ($b = -0.07$, $SE = 0.23$, $z = -0.31$, $p = 0.754$).

546 However, the size \times salience interaction was significant, $b = -1.05$, $SE = 0.31$, $z = -3.37$, $p <$
 547 $.001$.

548 When searching with a scotoma, the probability of missing the target was increased, b
 549 $= 0.72$, $SE = 0.14$, $z = 5.00$, $p < .001$. Timeout probability was low, with no timeouts for large
 550 high-salience targets; no statistical analysis was performed.

551



552

553 Figure 4. Measures of search accuracy for Experiment 1. Top row: small targets, bottom row:
 554 large targets. Each column presents means obtained for a designated dependent variable (see
 555 text for definitions). In each panel, data are shown for low- and high-salience targets during
 556 visual search with a simulated foveal scotoma (red) or without one (black). Data points are

557 binomial proportions, error bars are 95% binomial proportion confidence intervals (Wilson,
558 1927).

559 -----
560 Table 1 about here
561 -----

562 **3.2.2 Search time and its subcomponents**

563 Search time is the time taken from scene onset to participants' button press
564 terminating the search. Participants' gaze data were used to split search time into three
565 subcomponents: search initiation time, scanning time, and verification time (e.g., Clayden et
566 al., 2020; Malcolm & Henderson, 2009; Nuthmann, 2014; Nuthmann & Malcolm, 2016).
567 Search initiation time is the interval between scene onset and the initiation of the first saccade
568 (i.e., initial saccade latency). Scanning time is the time from the first eye movement until the
569 participant's gaze enters the target's area of interest. Verification time is the time from first
570 entering the target interest area until the participant confirms their decision via button press.
571 While the scanning time measure reflects the process of localizing the target in space,
572 verification time reflects the time needed to decide that the fixated object is the target
573 (Malcolm & Henderson, 2009). Longer scanning times indicate weaker target guidance. Long
574 verification times tend to include instances in which observers fixated the target but then
575 continued searching before returning to it (Castelhana, Pollatsek, & Cave, 2008; Clayden et
576 al., 2020; Rutishauser & Koch, 2007; Zhaoping & Frith, 2011; Zhaoping & Guyader, 2007).
577 Moreover, in the absence of foveal or central vision the eyes may move off the target to
578 unmask it and then process it in parafoveal or peripheral vision (Clayden et al., 2020;
579 Nuthmann, 2014). In both cases, there will be off-target fixations between the first and final
580 fixation on the target, the number of which appears to depend on the difficulty of the search
581 (Clayden et al., 2020; Rutishauser & Koch, 2007).

582 We manipulated both the target's size and its salience to explore how the effects
583 combine. Specifically, if high salience helps more for small targets, we should observe an
584 interaction between target size and target salience. In previous letter-in-scene search
585 experiments, in which target size was varied, we found that the verification process was
586 slowed down when foveal vision was not available, whereas the actual search process,
587 indexed by scanning time, remained unaffected (Clayden et al., 2020). Moreover, we tested
588 whether the importance of foveal vision to target verification depended on the size of the
589 target, but the data remained ambiguous (Clayden et al., 2020). With the present experiment,

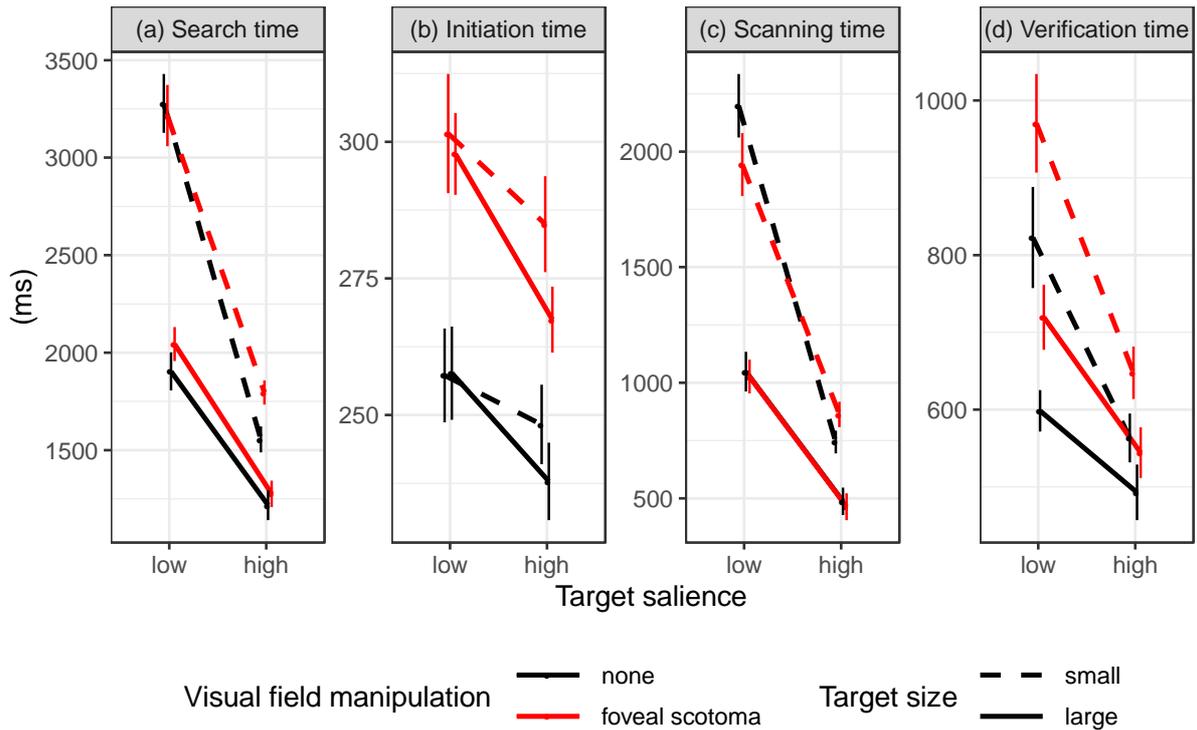
590 we wanted to test whether the availability of foveal vision during target verification was more
591 important if the target's salience was reduced, along with its size. If that were the case, the
592 foveal scotoma should be more detrimental for low-salience than for high-salience targets,
593 and it should be most detrimental for targets that are small and low in salience.

594 The analysis of search times showed a significant effect of target size with faster
595 searches for large as compared to small targets, $b = -927.79$, $SE = 99.45$, $t = -9.33$. The effect
596 of target salience was also significant, with shorter search times for high-salience as
597 compared to low-salience targets, $b = -1230.15$, $SE = 121.26$, $t = -10.14$. There was also a
598 significant interaction between target size and salience such that the salience effect was
599 smaller for large targets, $b = 958.04$, $SE = 166.65$, $t = 5.75$. Analyses of the three sub-
600 processes of search showed the same pattern of results (Table 1). The only exception was a
601 non-significant target size \times salience interaction for search initiation time, $b = -9.33$, $SE =$
602 6.88 , $t = -1.35$.

603 The presence of a foveal scotoma had a significant effect on search initiation and
604 verification, with both sub-processes of search being slowed down (Table 1). Importantly,
605 scanning time was not prolonged when searching with a foveal scotoma, $b = -16.72$, $SE =$
606 46.4 , $t = -0.36$. Button-press search times are the sum of search initiation, scanning, and
607 verification times. For the untransformed data, the search-time difference between the foveal
608 scotoma and the control condition was not significant, $b = 149.09$, $SE = 81.54$, $t = 1.83$. For
609 the transformed data, however, the effect of scotoma was significant, $b = 0.003$, $SE = 0.001$, t
610 $= 3.84$; it was qualified by a significant scotoma \times salience interaction such that the
611 detrimental effect of a foveal scotoma was larger for high-salience targets, $b = 0.002$, $SE =$
612 0.001 , $t = 3.16$.

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

617



618

619 Figure 5. Search time and its three epochs for Experiment 1. Each panel displays the means
 620 for a designated dependent variable (see panel title); note the different y-axis scales for the
 621 different measures. Targets differed in visual salience (x-axis) and size (small: dashed line,
 622 large: solid line). Observers searched the scene either with a simulated foveal scotoma (red
 623 line) or without one (black line). Search times are the sum of search initiation, scanning, and
 624 verification times. Error bars are within-subjects standard errors, using the Cousineau-Morey
 625 method (Cousineau, 2005; Morey, 2008).

626

627 3.2.3 Saccade amplitudes and fixation durations

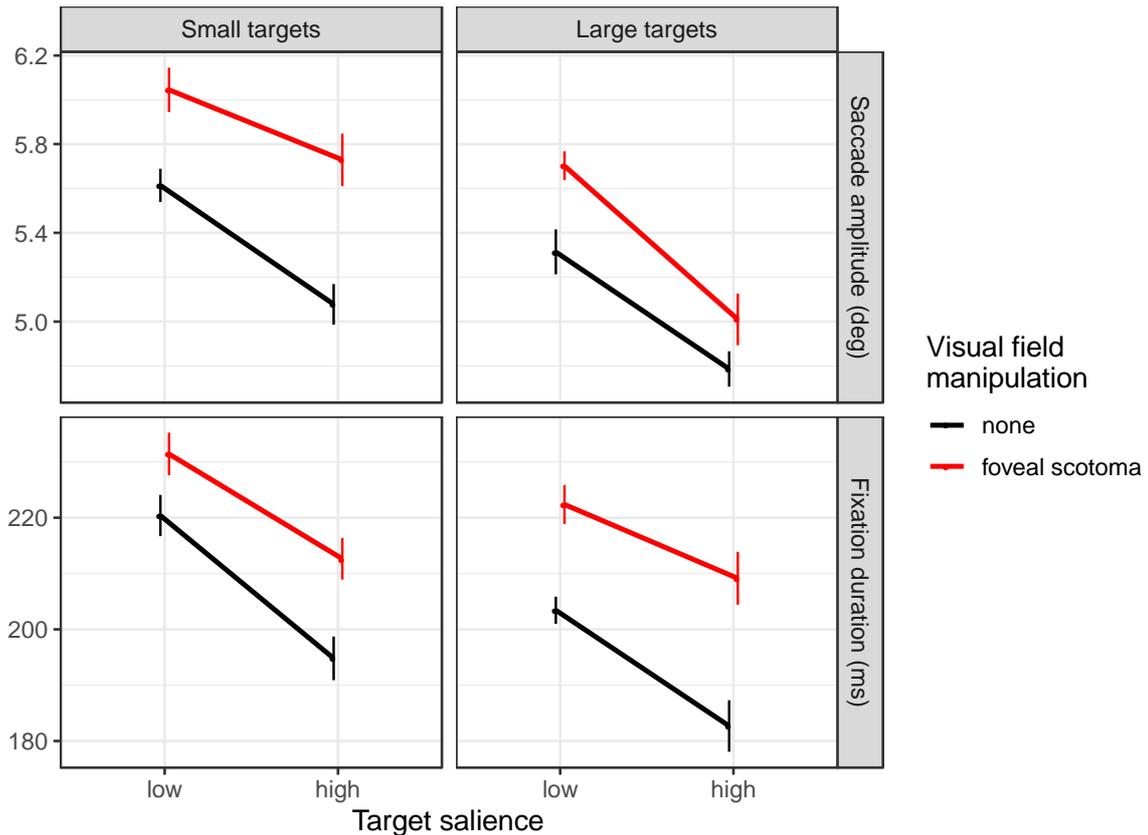
628 Saccade amplitudes and fixation durations were analyzed to characterize eye-
 629 movement behavior during visual search (Figure 6). During scene search with a simulated
 630 foveal scotoma, we expected to observe larger saccade amplitudes and longer fixation
 631 durations (Clayden et al., 2020;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

634 For saccade amplitudes we observed a significant effect of scotoma, with longer
 635 saccades when searching with a foveal scotoma than without ($b = 0.40$, $SE = 0.08$, $t = 4.89$,
 636 Figure 6, top row). There was also a significant effect of target size with shorter saccade
 637 amplitudes for large as compared to small targets, $b = -0.43$, $SE = 0.06$, $t = -6.96$. In addition,

638 there was a significant effect of target salience with shorter saccade amplitudes for high-
639 salience as compared to low-salience targets, $b = -0.57$, $SE = 0.10$, $t = -5.99$. The interaction
640 between target size and scotoma was significant, $b = -0.26$, $SE = 0.12$, $t = -2.11$, indicating
641 that the size effect was larger (i.e., more negative) with a foveal scotoma than without. For
642 the transformed data, however, this interaction was not significant, $b = -0.07$, $SE = 0.04$, $t = -$
643 1.76 . Thus, the interaction was transformed away, making it non-interpretable (Loftus, 1978;
644 Wagenmakers, Kryptos, Criss, & Iverson, 2012). None of the other interactions were
645 significant (Table 1).

646 The analysis of fixation durations revealed a similar pattern of results. There was a
647 significant effect of scotoma, with longer fixation durations when searching with a foveal
648 scotoma than without ($b = 19.85$, $SE = 3.88$, $t = 5.12$, Figure 6, bottom row). There was also a
649 significant effect of target size with shorter fixation durations for large as compared to small
650 targets, $b = -9.49$, $SE = 2.55$, $t = -3.72$. In addition, there was a significant effect of target
651 salience with shorter fixation durations for high-salience as compared to low-salience targets,
652 $b = -20.01$, $SE = 2.86$, $t = -6.99$. Furthermore, there was a significant size \times salience
653 interaction, $b = 7.68$, $SE = 3.52$, $t = 2.18$, which was absent for the transformed data, $b =$
654 0.08 , $SE = 0.05$, $t = 1.71$. Moreover, there was a significant scotoma \times salience interaction, b
655 $= 7.09$, $SE = 3.52$, $t = 2.02$, indicating that the salience effect was smaller with a foveal
656 scotoma than without. None of the other interactions were significant (Table 1).

657



658

659 Figure 6. Mean saccade amplitudes (top row) and fixation durations (bottom row) for small
 660 targets (left column) as opposed to large targets (right column) in Experiment 1. In each
 661 panel, data are presented for low- and high-salience targets during visual search with or
 662 without a simulated foveal scotoma. Error bars are within-subjects standard errors.

663

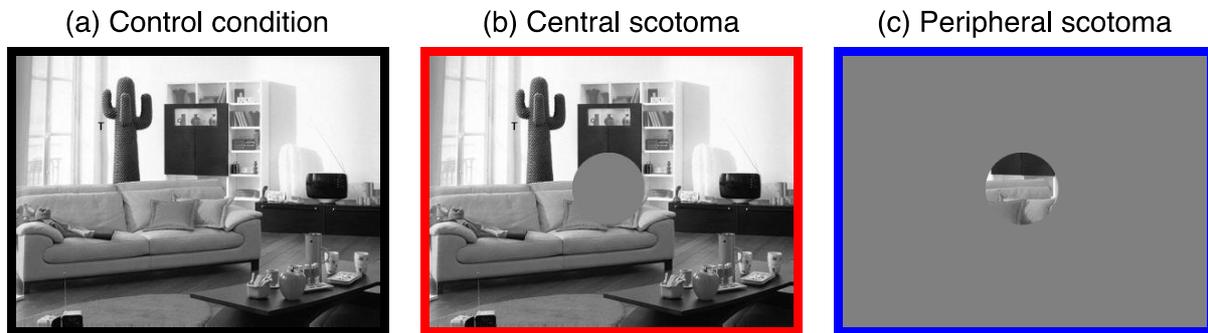
664 4 Experiment 2

665 4.1 Design

666 In Experiment 2, we dropped the manipulation of target size and instead used the
 667 small targets from Experiment 1 throughout. As in Experiment 1, we manipulated the visual
 668 salience of the target letter (low vs. high). This was crossed with another visual field
 669 manipulation: observers searched for the target with a central or peripheral scotoma, for
 670 which the normal-vision control condition provided a baseline (Figure 7). Compared with the
 671 foveal scotoma in Experiment 1 (radius: 1°), the central scotoma in Experiment 2 had a larger
 672 radius (2.5°). The central scotoma was contrasted with the inverse manipulation of a
 673 peripheral scotoma with the same radius. In the visual-cognition literature, central vision is
 674 defined as extending to about 5° from fixation, with peripheral vision being everything
 675 beyond 5° (Loschky, Szaffarczyk, Beugnet, Young, & Boucart, 2019). Technically, our

676 central scotoma did not completely cover central vision, and our peripheral scotoma obscured
677 more than peripheral vision.

678



679

680 Figure 7. Scotoma conditions used in Experiment 2. Observers searched the scene either with
681 full vision (control condition), or with a central or peripheral scotoma (radius: 2.5°). Note that
682 the colored borders match the colors used to distinguish the scotoma-type conditions in
683 Figures 7 to 10. Search targets varied in visual salience; the example scene used for this
684 figure includes the high-salience target.

685

686 To facilitate comparisons across experiments, we used the same scenes with the same
687 locations for low- and high-salience targets as in Experiment 1. A given participant saw each
688 of the 120 scene items once, with 20 scenes in each of the six experimental conditions. The
689 visual field manipulation was blocked so that participants completed three blocks of trials in
690 the experiment. Each block started with four practice trials, two for each target salience
691 condition. The order of blocks was counterbalanced across subjects. Within a block, scenes
692 were presented randomly.

693 4.2 Results

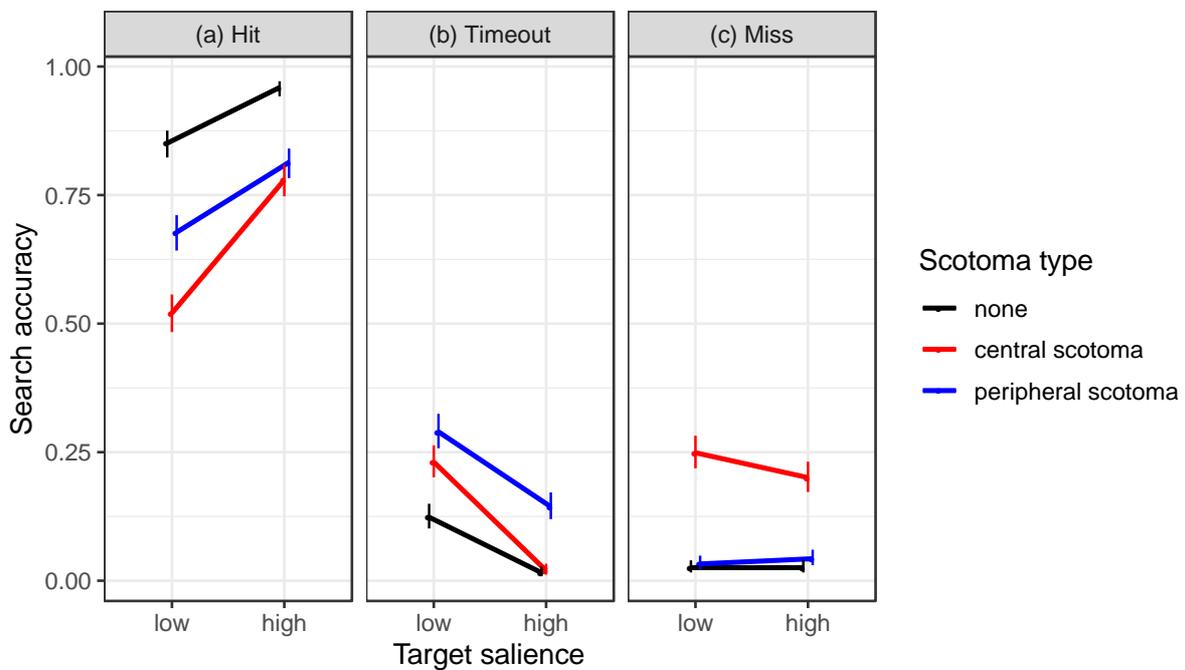
694 4.2.1 Search accuracy

695 The type of the simulated scotoma affected the probability of “hitting” the target, with
696 highest probabilities in the no-scotoma control condition and lowest probabilities for the
697 central scotoma (Figure 8a). The effect of scotoma type on search accuracy was tested using
698 backward difference coding (Table 2). The GLMM results substantiated that search accuracy
699 was significantly reduced for the peripheral scotoma condition (P) compared to the no-
700 scotoma control condition (No), P-No: $b = -1.44$, $SE = 0.14$, $z = -10.19$, $p < .001$. For the
701 central scotoma (C), search accuracy was lower than for the peripheral scotoma, C-P: $b = -$
702 0.77 , $SE = 0.13$, $z = -6.01$, $p < .001$. As in Experiment 1, there was a significant main effect

703 of target salience on search accuracy, with better performance for high-salience than for low-
 704 salience targets, $b = 1.21$, $SE = 0.10$, $z = 11.94$, $p < .001$. The salience effect was significantly
 705 reduced for the peripheral scotoma compared to the no-scotoma control condition, salience \times
 706 P-No interaction: $b = -0.62$, $SE = 0.27$, $z = -2.31$, $p = 0.021$. The salience effect was
 707 significantly increased for the central scotoma compared to the peripheral scotoma, salience \times
 708 C-P interaction: $b = 0.74$, $SE = 0.19$, $z = 3.82$, $p < .001$.

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

714



715

716 Figure 8. Measures of search accuracy for Experiment 2. Each panel presents means obtained
 717 for a designated dependent variable, which is specified in the panel title. Data are shown for
 718 low- and high-salience targets and for different scotoma types (red: central scotoma, blue:
 719 peripheral scotoma, black: no-scotoma control condition). Data points are binomial
 720 proportions, error bars are 95% binomial proportion confidence intervals (Wilson, 1927).

721

722

Table 2 about here

723

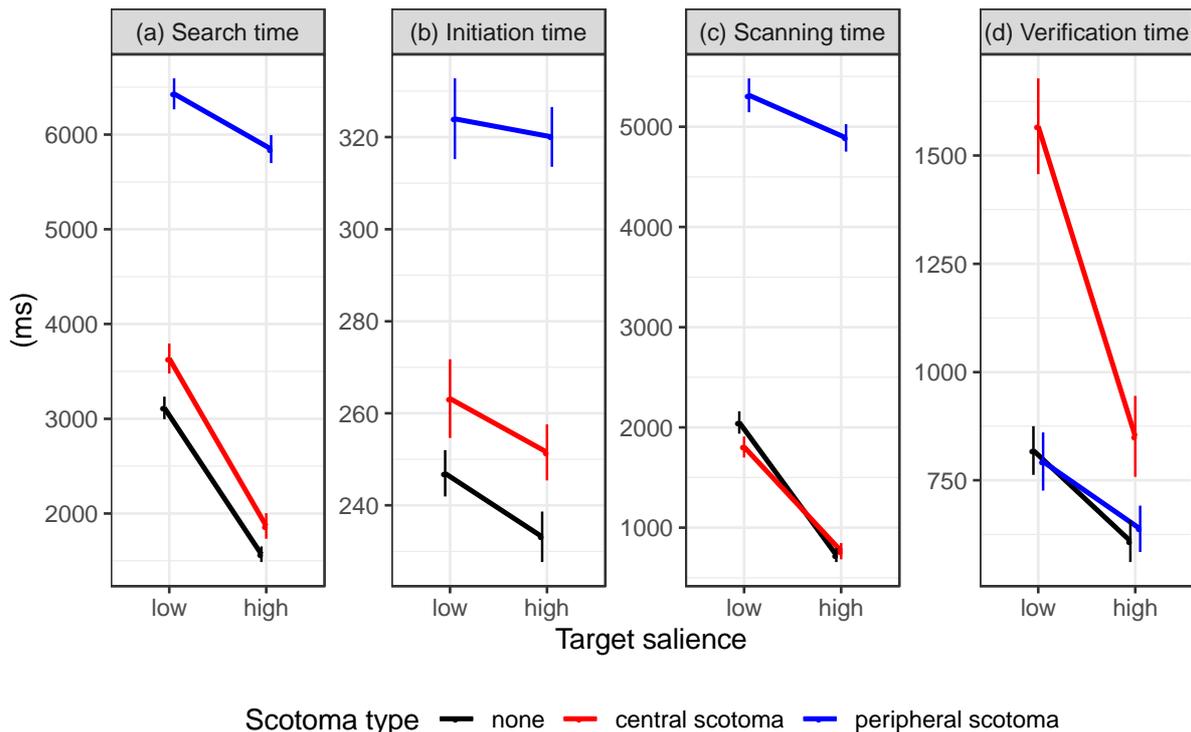
724 4.2.2 Search time and its subcomponents

725 Trials with correct responses were analyzed further. The type of the simulated
726 scotoma affected button-press search times, which were shortest in the no-scotoma control
727 condition and longest when searching with a peripheral scotoma (Figure 9a). The effect of
728 scotoma type on search times was tested using backward difference coding. Search times
729 were significantly longer during search with a central scotoma than during search without a
730 scotoma, C-No: $b = 756.36$, $SE = 117.71$, $t = 6.43$. Search times were further increased for
731 the peripheral scotoma compared to the central scotoma, P-C: $b = 2995.29$, $SE = 214.97$, $t =$
732 13.93 . Moreover, there was a significant main effect of target salience with shorter search
733 times for high-salience compared to low-salience targets, $b = -1523.04$, $SE = 156.45$, $t = -$
734 9.74 . The salience effect was significantly increased for the central scotoma compared to the
735 no-scotoma control condition, salience \times C-No interaction: $b = -707.37$, $SE = 227.02$, $t = -$
736 3.12 . The salience effect was significantly reduced for the peripheral scotoma compared to
737 the central scotoma, salience \times P-C interaction: $b = 1761.53$, $SE = 354.56$, $t = 4.97$.

738 Based on participants' gaze data, button-pressed search times were decomposed into
739 search initiation, scanning, and verification times (Figure 9b, c, d). To evaluate the effect of
740 scotoma type, we used simple coding with the no-scotoma control condition as the reference
741 level. For search with a peripheral scotoma, search initiation time was significantly increased,
742 $b = 81.99$, $SE = 7.43$, $t = 11.03$. Search initiation times were also increased for the central
743 scotoma; this effect was significant for the untransformed data, $b = 17.09$, $SE = 8.18$, $t = 2.09$,
744 but not for the transformed data, $b = 2.16$, $SE = 1.16$, $t = 1.86$. Moreover, there was a
745 significant main effect of target salience with shorter search initiation times for high-salience
746 compared to low-salience targets, $b = -8.62$, $SE = 3.77$, $t = -2.29$. The two interactions
747 involving salience were not significant (Table 2).

748 Scanning time was significantly prolonged when searching with a peripheral scotoma,
749 $b = 3725.53$, $SE = 187.46$, $t = 19.87$. For the central scotoma, there was a numerical increase
750 in scanning time which was not significant, $b = 85.21$, $SE = 108.27$, $t = 0.79$; for the
751 transformed data, however, it was significant, $b = 0.14$, $SE = 0.06$, $t = 2.38$. Scanning times
752 were shorter for high-salience compared to low-salience targets, $b = -1043.68$, $SE = 124.55$, $t =$
753 -8.38 . The effect of target salience was significantly reduced for the peripheral scotoma, $b =$
754 978.91 , $SE = 304.12$, $t = 3.22$, but not for the central scotoma, $b = 6.18$, $SE = 209.5$, $t =$
755 0.03 .

756 Verification time was significantly prolonged when searching with a central scotoma,
 757 $b = 564.89$, $SE = 98.63$, $t = 5.73$, but not when searching with a peripheral scotoma, $b = -$
 758 36.76 , $SE = 64.28$, $t = -0.57$. Verification times were shorter for high-salience compared to
 759 low-salience targets, $b = -442.75$, $SE = 63.09$, $t = -7.02$. This effect was significantly
 760 increased for the central scotoma, $b = -660.97$, $SE = 172.18$, $t = -3.84$, but not for the
 761 peripheral scotoma, $b = 54.2$, $SE = 84.33$, $t = 0.64$.
 762



763
 764 Figure 9. Search time and its three epochs for Experiment 2. Each panel displays the means
 765 for a designated dependent variable (see panel title); note the different y-axis scales for the
 766 different measures. Results are presented for low- and high-salience targets and for different
 767 scotoma types (red: central scotoma, blue: peripheral scotoma, black: no-scotoma control
 768 condition). Error bars are within-subjects standard errors, using the Cousineau-Morey method
 769 (Cousineau, 2005; Morey, 2008).
 770

771 4.2.3 Saccade amplitudes and fixation durations

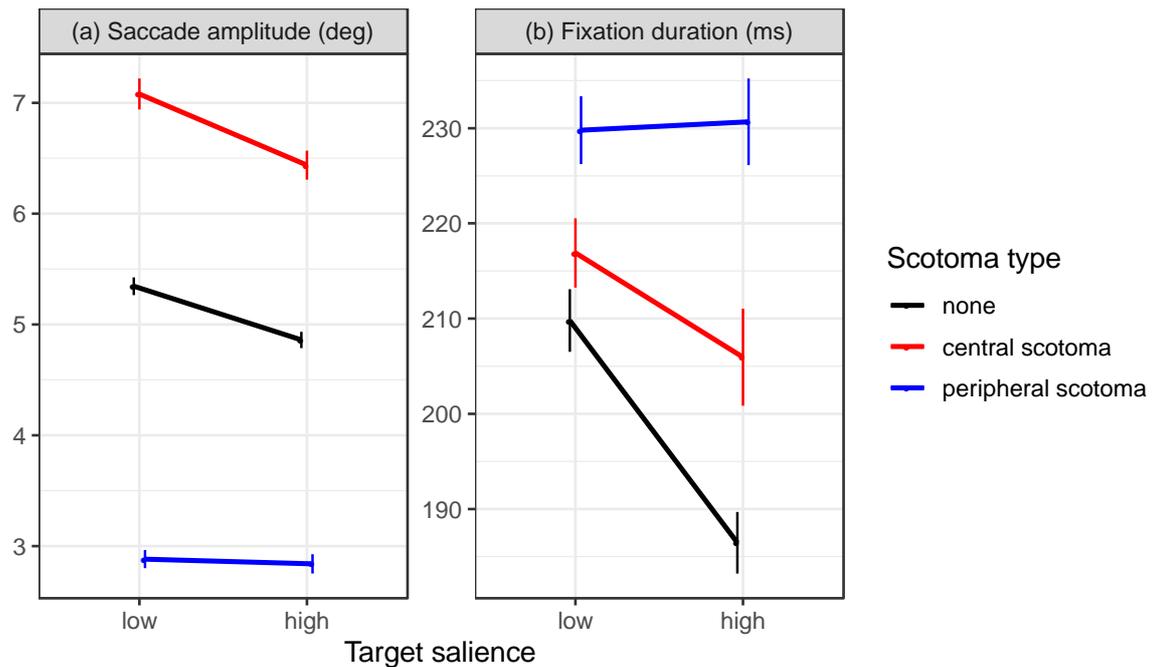
772 Moving-window studies that implemented something akin to our peripheral scotoma
 773 have consistently reported shorter saccade amplitudes and longer fixation durations than in a
 774 normal vision control condition (e.g., Loschky & McConkie, 2002; Nuthmann, 2014). By

775 contrast, masking or degrading central vision tends to increase both saccade amplitudes and
776 fixation durations (Miellet et al., 2010; Nuthmann, 2014).

777 The present data replicate the “windowing effect” on saccade amplitudes. Compared
778 to the no-scotoma control condition, saccade amplitudes were significantly longer when
779 searching with a central scotoma, $b = 1.6$, $SE = 0.15$, $t = 10.63$, and significantly shorter when
780 searching with a peripheral scotoma, $b = -2.2$, $SE = 0.1$, $t = -21.7$. Moreover, as in
781 Experiment 1 there was a significant main effect of target salience with shorter saccade
782 amplitudes for high-salience compared to low-salience targets, $b = -0.42$, $SE = 0.06$, $t = -6.5$.
783 There was also a significant salience \times peripheral scotoma interaction, $b = 0.43$, $SE = 0.08$, t
784 $= 5.27$, indicating that the effect of target salience was reduced for the peripheral scotoma.
785 The interaction between salience and central scotoma was not significant (Table 2).

786 The type of the simulated scotoma also affected fixation durations, which were
787 shortest in the no-scotoma control condition and longest when searching with a peripheral
788 scotoma (Figure 10b). The effect of scotoma type on fixation durations was tested using
789 backward difference coding (Table 2). The LMM results substantiated that fixation durations
790 were significantly longer during search with a central scotoma than during search without a
791 scotoma, C-No: $b = 17.09$, $SE = 4.5$, $t = 3.8$. For the peripheral scotoma, fixation durations
792 were significantly increased compared to the central scotoma, P-C: $b = 18.79$, $SE = 6.33$, $t =$
793 2.97 . As in Experiment 1, there was also a significant main effect of target salience with
794 shorter fixation durations for high-salience compared to low-salience targets, $b = -12.24$, SE
795 $= 2.12$, $t = -5.78$. The salience effect was significantly reduced for the central scotoma
796 compared to the no-scotoma control condition, salience \times C-No interaction: $b = 11.58$, $SE =$
797 4.07 , $t = 2.85$. The salience effect was further reduced for the peripheral scotoma compared to
798 the central scotoma, salience \times P-C interaction: $b = 10.29$, $SE = 4.23$, $t = 2.43$.

799



800

801 Figure 10. Mean saccade amplitudes (a) and fixation durations (b) in Experiment 2 as a
 802 function of target salience and scotoma type, red: central scotoma, blue: peripheral scotoma,
 803 black: no-scotoma control condition. Error bars are within-subjects standard errors.

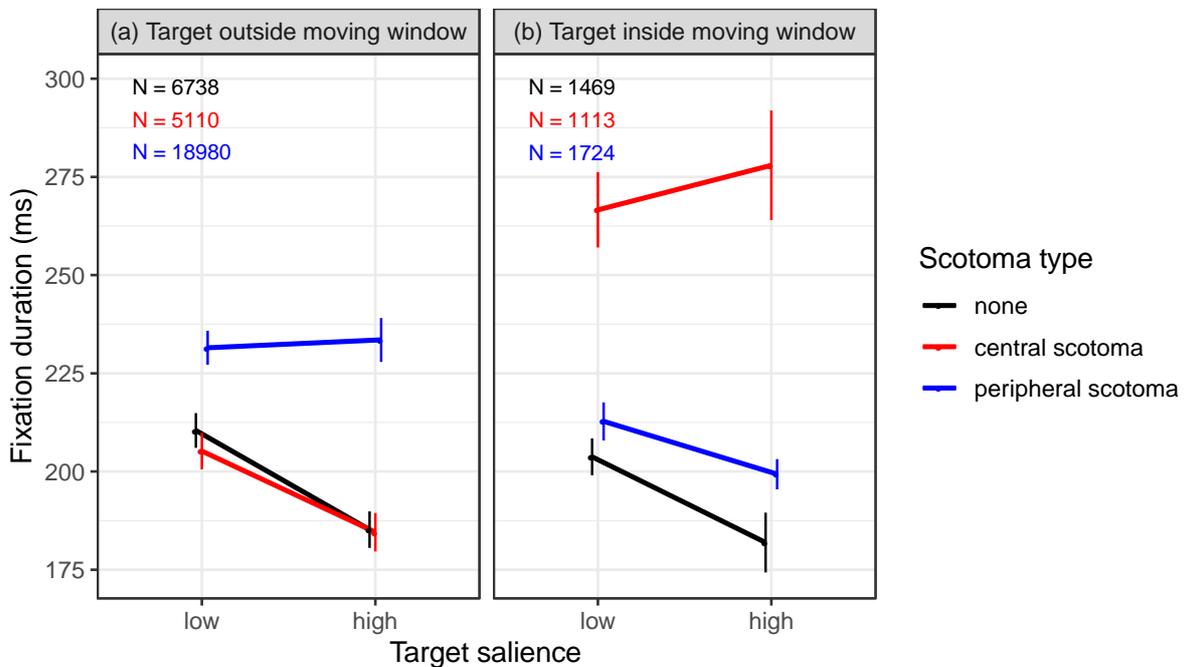
804

805 4.2.4 Control analyses

806 With a peripheral scotoma, the target was not visible to the observer during their
 807 initial fixation at the center of the scene. During most subsequent valid fixations, the target
 808 remained invisible as it was outside the window in which scene content was available. Thus,
 809 search initiation times, saccade amplitudes, and fixation durations should be unaffected by
 810 target salience in this condition. To test this explicitly, we specified additional LMMs using
 811 dummy coding and the peripheral scotoma as reference level. In such a model, the simple
 812 effect for target salience represents the salience effect for the peripheral scotoma. No
 813 significant salience effects were found (search initiation times: $b = -0.52$, $SE = 6.32$, $t = -$
 814 0.08 ; saccade amplitudes: $b = -0.03$, $SE = 0.04$, $t = -0.87$; fixation durations: $b = -0.02$, $SE =$
 815 0.01 , $t = -1.64$).

816 Results from existing studies suggest that visual information within both foveal,
 817 parafoveal, and peripheral vision can influence fixation duration (Einhäuser, Atzert, &
 818 Nuthmann, 2020, for review). Therefore, an additional analysis explored whether effects of
 819 target salience on fixation duration arise from both central and peripheral processing. For
 820 each individual fixation, we determined whether the target was inside or outside the circular

821 window that was used to create the two scotomas. As an approximation, the midpoint of the
 822 target was used for this evaluation. For the central scotoma, the target was visible if it was
 823 outside the window (see Figure 7b), and invisible if it was inside the window. Conversely, for
 824 the peripheral scotoma the target was visible if it was inside the window, and invisible if it
 825 was outside the window (see Figure 7c). We expected target salience to only modulate
 826 fixation durations if the target was visible. The data are consistent with this prediction. For
 827 the central scotoma, the salience effect was present when the target was outside the window
 828 (Figure 11a), whereas it was absent when the target was inside the window (Figure 11b). For
 829 the peripheral scotoma, a salience effect emerged if the target was inside the window (Figure
 830 11b), whereas it was absent when the target was outside the window (Figure 11a). For the no-
 831 scotoma control condition, where the target was always present, the salience effect was
 832 present for both types of fixations. Interestingly, the data also suggest that fixation durations
 833 during search with the central scotoma were not elevated when the target was visible in the
 834 periphery (Figure 11a). Given the post-hoc nature of this explorative analysis, no formal
 835 statistical analyses were conducted. The number of cases in which the target was outside the
 836 window during the fixation amounted to 88% (see Figure 11 for a breakdown). This is why
 837 the analysis of all valid fixations yielded no salience effect for the peripheral scotoma and a
 838 reduced salience effect for the central scotoma (Figure 10b).
 839



840
 841 Figure 11. Mean fixation durations in Experiment 2 as a function of target salience, scotoma
 842 type, and whether the target was outside (a) or inside (b) the scotoma window that moved

843 with the participants' eyes. Error bars are within-subjects standard errors. N = number of
844 observations for a given scotoma-type condition.

845

846 **5 General Discussion**

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

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

869 Investigating the causal influence of features on gaze guidance during scene search
870 requires one to use an experimental approach in which objects or regions in natural scenes are
871 manipulated (Foulsham & Underwood, 2007). In the studies reviewed in the Introduction, the
872 approach has been to select targets based on the output from versions of a popular saliency
873 map model. When manipulating properties of real-world objects in naturalistic scenes, it is
874 impossible to exert perfect experimental control over relevant dimensions. Therefore, the
875 possibility exists that—in some existing scene sets—visual salience is confounded with other

876 variables like object size, eccentricity and semantic congruency. To address these issues, we
877 used the T.E.A. (Clayden et al., 2020) to parametrically manipulate target salience and size in
878 a letter-in-scene search task. In this task, the location of the target is not predicted by the
879 meaning of the scene or by the identity of objects in the scene. Our task still approximates
880 natural behavior because there are real-world searches for which there is minimal guidance
881 by scene context (e.g., search for a fly). Moreover, scene processing and object identification
882 are not totally suppressed when searching for a “T” overlaid onto the scene (T. H. W.
883 Cornelissen & Võ, 2017). One caveat regarding generalizing from letter search to object
884 search in scenes is that the letter targets tend to violate the physical rules of the scene
885 environment in which they appear, such as gravity and surface reflectance. Additionally,
886 although we used images of naturalistic scenes to improve the ecological validity of the
887 search task, these scenes are still two-dimensional static representations of the environment,
888 and so generalization to the natural world should be made with caution.

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

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

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

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

925 The results for the foveal, central, and peripheral scotomas tell us how important the
926 different regions of the visual field are for visual search and its sub-processes. During search
927 with any type of scotoma, observers were significantly less likely to find the target than with
928 normal vision. However, when the target was found despite the presence of a simulated
929 foveal scotoma (Experiment 1), search times were not much elevated (Figure 5a, Table 1). In
930 contrast, the presence of a central or peripheral scotoma (Experiment 2) led to clear search
931 time costs (Figure 9a, Table 2). As expected, the peripheral scotoma was much more
932 detrimental than the central scotoma, confirming that eye movements are guided by
933 peripheral vision. Analyzing sub-processes of search allowed for testing the assumption of a
934 central-peripheral dichotomy according to which peripheral vision is mainly for selecting or
935 looking, while central vision is mainly for seeing or recognizing (Zhaoping, 2019). In
936 Experiment 1, we found that verification times, but not scanning times were significantly
937 prolonged when searching with a foveal scotoma (see also Clayden et al., 2020). In
938 Experiment 2, we found that scanning times were prolonged for the peripheral but not for the
939 central scotoma, whereas verification times were prolonged for the central scotoma but not
940 for the peripheral scotoma (cf. Nuthmann, 2014). Collectively, the data highlight the
941 importance of peripheral vision for target localization, and the importance of foveal and

942 central vision for target verification. This pattern of results is consistent with the central-
943 peripheral dichotomy (Zhaoping, 2019).

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

949 By comparison, the peripheral scotoma weakened the effect of target salience on
950 scanning time (Figure 9c) and also total search time (Figure 9a). This finding is different
951 from results by Foulsham and Underwood (2011). In their Experiment 2, the authors used a
952 gaze-contingent 6° square window to selectively remove image features from the periphery.
953 They tested three image features that are important for saccade target selection under the
954 saliency map hypothesis: (1) color, (2) high-spatial frequency information, and (3) contrast
955 (i.e., the contrast of the image was globally lowered). If saliency in peripheral vision was
956 guiding eye movements towards the target, then peripheral filtering should eliminate or at
957 least diminish the effect of target salience. Contrary to these predictions, the authors found a
958 significant salience effect on the latency to first fixate the target in all three filtering
959 conditions. Unfortunately, their Experiment 2 lacked a full-vision control condition.
960 Moreover, the filtering manipulations left some of the saliency map representation intact. By
961 comparison, the peripheral scotoma in our Experiment 2 blocked out peripheral vision
962 completely. Compared with the no-scotoma control condition, we observed a greatly
963 diminished effect of target salience during target localization, as indexed by scanning time
964 (Figure 9c). Our results therefore suggest that salience in peripheral vision was guiding eye
965 movements towards the target.

966 The central scotoma increased the effect of target salience on verification time (Figure
967 9d, Table 2). Compared with the no-scotoma control condition, mean target verification times
968 were not only elevated, but they also showed a significantly increased effect of target
969 salience. Thus, our results not only suggest that central vision benefits target verification (cf.
970 Nuthmann, 2014), but also that this sub-process of search is influenced by target salience.
971 The increase in the salience effect for verification times was large enough to produce an
972 increased salience effect on total search time as well (Figure 9a, Table 2).

973 The data from both experiments replicate the well-known “windowing effect” on
974 saccade amplitudes, which reflects a tendency to fixate more locations in the non-degraded
975 scene area than the degraded area (Loschky & McConkie, 2002; Mielliet et al., 2010;

976 Nuthmann, 2014; Reingold & Loschky, 2002). Moreover, the present results replicate the
977 finding that fixation durations are elevated in the presence of an artificial scotoma (Clayden
978 et al., 2020; Miellet et al., 2010; Nuthmann, 2014). In our experiments, we experimentally
979 manipulated properties of the search target, and our analysis of saccade amplitudes and
980 fixation durations considered the entire search period. On a given fixation, the target was
981 situated in either foveal, central, or peripheral vision, where it could be obscured by a
982 simulated scotoma or not. In Experiment 1, global eye-movement parameters were affected
983 by target properties such that large targets and high-salience targets were associated with
984 shorter saccade amplitudes and shorter fixation durations; for target size, similar results were
985 obtained by Clayden et al. (2020). Interestingly, significant effects of target size and salience
986 were already present for the duration of the very first fixation, measured as search initiation
987 time (Figure 5b, Table 1). Previous research has demonstrated that the “story,” or gist of a
988 scene can be gleaned from it within around 100 ms of the onset of a scene (Oliva, 2005;
989 Potter, 1975). Scene gist is typically perceived without recognizing any individual object.
990 Therefore, we tentatively propose that observers, during the first glance of the scene, may
991 form a hypothesis about the scene’s search difficulty in terms of target size and salience, and
992 globally adjust their fixation durations and saccade amplitudes accordingly.

993 In Experiment 2, we replicated effects of target salience on saccade amplitudes,
994 fixation durations, and search initiation times (Table 2). Interestingly, the results for the
995 scotoma conditions in both experiments provide clues about necessary conditions for these
996 effects to occur. The peripheral scotoma in Experiment 2 prevented observers from analyzing
997 the scene gist and covered the target during most fixations, including the very first. In this
998 condition, no differences for low- and high-salience targets were observed for saccade
999 amplitudes and search initiation times; the same was true for fixation durations, as long as the
1000 target was outside the window in which scene content was visible. For saccade amplitudes,
1001 the effect of target salience was unchanged when searching with a foveal scotoma
1002 (Experiment 1) or with a central scotoma (Experiment 2). For fixation durations, the effect of
1003 target salience was reduced when searching with a foveal scotoma. For the central scotoma,
1004 the salience effect was present when the target was visible (outside the scotoma), and absent
1005 when it was not visible due to being masked by the scotoma. When all valid fixations were
1006 analyzed together, the salience effect was therefore smaller in the central-scotoma condition
1007 than in the no-scotoma control condition. Collectively, the data suggest that the salience
1008 effect on fixation durations arises from both foveal, central, and peripheral processing.

1009 Moreover, peripheral vision needs to be intact to observe effects of target salience on saccade
1010 amplitudes.

1011 **6 Conclusions**

1012 Methodologically, reliably disentangling stimulus-driven and task-driven influences
1013 on human behavior requires researchers to exert experimental control over relevant stimulus
1014 dimensions, which is challenging when working with images of naturalistic scenes. Here, we
1015 placed context-free targets within scenes using the T.E.A. (Clayden et al., 2020), which
1016 allowed us to manipulate their salience and size parametrically. When using these stimuli for
1017 a target acquisition task in two experiments, clear effects of target salience on search
1018 performance and eye-movement parameters were found. Moreover, the results obtained in
1019 different simulated scotoma conditions lend further support to the central-peripheral
1020 dichotomy (Zhaoping, 2019).

References

- 1021
1022 Adeli, H., Vitu, F., & Zelinsky, G. J. (2017). A model of the superior colliculus predicts
1023 fixation locations during scene viewing and visual search. *Journal of Neuroscience*,
1024 37(6), 1453–1467. <https://doi.org/10.1523/JNEUROSCI.0825-16.2016>
- 1025 Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed
1026 random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–
1027 412. <https://doi.org/10.1016/j.jml.2007.12.005>
- 1028 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for
1029 confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*,
1030 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- 1031 Bates, D., Maechler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects
1032 models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
1033 <https://doi.org/10.18637/jss.v067.i01>
- 1034 Bertera, J. H. (1988). The effect of simulated scotomas on visual search in normal subjects.
1035 *Investigative Ophthalmology & Visual Science*, 29(3), 470–475.
- 1036 Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural
1037 images. *Journal of the Optical Society of America A-Optics Image Science and Vision*,
1038 19(6), 1096–1106. <https://doi.org/10.1364/JOSAA.19.001096>
- 1039 Biederman, I. (1972). Perceiving real-world scenes. *Science*, 177(4043), 77–80.
1040 <https://doi.org/10.1126/science.177.4043.77>
- 1041 Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting
1042 and judging objects undergoing relational violations. *Cognitive Psychology*, 14(2), 143–
1043 177. [https://doi.org/10.1016/0010-0285\(82\)90007-X](https://doi.org/10.1016/0010-0285(82)90007-X)
- 1044 Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H.,
1045 & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology
1046 and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135.
1047 <https://doi.org/10.1016/j.tree.2008.10.008>
- 1048 Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions*
1049 *on Pattern Analysis and Machine Intelligence*, 35(1), 185–207.
1050 <https://doi.org/10.1109/tpami.2012.89>
- 1051 Borji, A., Sihite, D. N., & Itti, L. (2013a). Quantitative analysis of human-model agreement
1052 in visual saliency modeling: a comparative study. *IEEE Transactions on Image*
1053 *Processing*, 22(1), 55–69. <https://doi.org/10.1109/TIP.2012.2210727>
- 1054 Borji, A., Sihite, D. N., & Itti, L. (2013b). What stands out in a scene? A study of human

1055 explicit saliency judgment. *Vision Research*, 91, 62–77.
1056 <https://doi.org/http://dx.doi.org/10.1016/j.visres.2013.07.016>

1057 Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal*
1058 *Statistical Society Series B-Statistical Methodology*, 26(2), 211–252.
1059 <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>

1060 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.
1061 <https://doi.org/10.1163/156856897X00357>

1062 Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the “Spotlight” reveals
1063 universal information use for face recognition. *PLOS ONE*, 5(3), e9708.
1064 <https://doi.org/10.1371/journal.pone.0009708>

1065 Castelhana, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an
1066 unspecified target, but only in identification and not in attentional guidance. *Psychonomic*
1067 *Bulletin & Review*, 15(4), 795–801. <https://doi.org/10.3758/PBR.15.4.795>

1068 Clayden, A. C., Fisher, R. B., & Nuthmann, A. (2020). On the relative (un)importance of
1069 foveal vision during letter search in naturalistic scenes. *Vision Research*, 177, 41–55.
1070 <https://doi.org/10.1016/j.visres.2020.07.005>

1071 Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking
1072 with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods,*
1073 *Instruments, & Computers*, 34(4), 613–617. <https://doi.org/10.3758/BF03195489>

1074 Cornelissen, T. H. W., & Vö, M. L.-H. (2017). Stuck on semantics: Processing of irrelevant
1075 object-scene inconsistencies modulates ongoing gaze behavior. *Attention Perception &*
1076 *Psychophysics*, 79(1), 154–168. <https://doi.org/10.3758/s13414-016-1203-7>

1077 Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to
1078 Loftus and Masson’s method. *Tutorials in Quantitative Methods for Psychology*, 1(1),
1079 42–45. <https://doi.org/10.20982/tqmp.01.1.p042>

1080 Demidenko, E. (2013). *Mixed models: Theory and applications with R* (2d ed.). Hoboken,
1081 New Jersey: John Wiley & Sons. <https://doi.org/10.1002/9781118651537>

1082 Duchowski, A. T., & Çöltekin, A. (2007). Foveated gaze-contingent displays for peripheral
1083 LOD management, 3D visualization, and stereo Imaging. *ACM Transactions on*
1084 *Multimedia Computing Communications and Applications*, 3(4):24, 1–18.
1085 <https://doi.org/10.1145/1314303.1314309>

1086 Einhäuser, W., Atzert, C., & Nuthmann, A. (2020). Fixation durations in natural scene
1087 viewing are guided by peripheral scene content. *Journal of Vision*, 20(4):15, 1–15.
1088 <https://doi.org/10.1167/jov.20.4.15>

1089 Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse
1090 the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*,
1091 8(2):2, 1–19. <https://doi.org/10.1167/8.2.2>

1092 Foulsham, T., Alan, R., & Kingstone, A. (2011). Scrambled eyes? Disrupting scene structure
1093 impedes focal processing and increases bottom-up guidance. *Attention Perception &*
1094 *Psychophysics*, 73(7), 2008–2025. <https://doi.org/10.3758/s13414-011-0158-y>

1095 Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the
1096 potency of visual saliency in scene perception? *Perception*, 36(8), 1123–1138.
1097 <https://doi.org/10.1068/p5659>

1098 Foulsham, T., & Underwood, G. (2011). If visual saliency predicts search, then why?
1099 Evidence from normal and gaze-contingent search tasks in natural scenes. *Cognitive*
1100 *Computation*, 3(1), 48–63. <https://doi.org/10.1007/s12559-010-9069-9>

1101 Geringswald, F., Baumgartner, F. J., & Pollmann, S. (2013). A behavioral task for the
1102 validation of a gaze-contingent simulated scotoma. *Behavior Research Methods*, 45(4),
1103 1313–1321. <https://doi.org/10.3758/s13428-013-0321-6>

1104 Glaholt, M. G., Rayner, K., & Reingold, E. M. (2012). The mask-onset delay paradigm and
1105 the availability of central and peripheral visual information during scene viewing.
1106 *Journal of Vision*, 12(1):9, 1–19. <https://doi.org/10.1167/12.1.9>

1107 Henderson, J. M., Brockmole, J. R., Castelano, M. S., & Mack, M. (2007). Visual saliency
1108 does not account for eye movements during visual search in real-world scenes. In R. P.
1109 G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A*
1110 *window on mind and brain* (pp. 537–562). Oxford: Elsevier.
1111 <https://doi.org/http://dx.doi.org/10.1016/B978-008044980-7/50027-6>

1112 Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M.
1113 Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye*
1114 *movements and the visual world* (pp. 1–58). New York: Psychology Press.

1115 Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive
1116 relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, 16(5),
1117 850–856. <https://doi.org/10.3758/PBR.16.5.850>

1118 Holmqvist, K., & Andersson, R. (2017). *Eye tracking: A comprehensive guide to methods,*
1119 *paradigms and measures*. Lund, Sweden: Lund Eye-Tracking Research Institute.

1120 Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in
1121 the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and*
1122 *scene perception* (pp. 29–53). Oxford: Elsevier Science Ltd.

1123 <https://doi.org/10.1016/B978-008043361-5/50003-1>

1124 Itti, L. (2006). Quantitative modelling of perceptual salience at human eye position. *Visual*

1125 *Cognition*, 14(4–8), 959–984. <https://doi.org/10.1080/13506280500195672>

1126 Itti, L., & Koch, C. (1999). Comparison of feature combination strategies for saliency-based

1127 visual attention systems. *Proc. SPIE*, Vol. 3644, pp. 473–482.

1128 <https://doi.org/10.1117/12.348467>

1129 Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of

1130 visual attention. *Vision Research*, 40(10–12), 1489–1506.

1131 [https://doi.org/10.1016/S0042-6989\(99\)00163-7](https://doi.org/10.1016/S0042-6989(99)00163-7)

1132 Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid

1133 scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

1134 20(11), 1254–1259. <https://doi.org/10.1109/34.730558>

1135 Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not)

1136 and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.

1137 <https://doi.org/10.1016/j.jml.2007.11.007>

1138 Kleiner, M., Brainard, D., & Pelli, D. (2007). What’s new in Psychtoolbox-3? *Perception*, 36,

1139 14.

1140 Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying

1141 neural circuitry. *Human Neurobiology*, 4(4), 219–227.

1142 Koehler, K., Guo, F., Zhang, S., & Eckstein, M. P. (2014). What do saliency models predict?

1143 *Journal of Vision*, 14(3):14, 1–27. <https://doi.org/10.1167/14.3.14>

1144 Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in

1145 linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.

1146 <https://doi.org/10.18637/jss.v082.i13>

1147 Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, 6(3), 312–319.

1148 <https://doi.org/10.3758/BF03197461>

1149 Loschky, L. C., & McConkie, G. W. (2002). Investigating spatial vision and dynamic

1150 attentional selection using a gaze-contingent multiresolutional display. *Journal of*

1151 *Experimental Psychology: Applied*, 8(2), 99–117. [https://doi.org/10.1037/1076-](https://doi.org/10.1037/1076-898X.8.2.99)

1152 [898X.8.2.99](https://doi.org/10.1037/1076-898X.8.2.99)

1153 Loschky, L. C., Szaffarczyk, S., Beugnet, C., Young, M. E., & Boucart, M. (2019). The

1154 contributions of central and peripheral vision to scene-gist recognition with a 180

1155 degrees visual field. *Journal of Vision*, 19(5):15, 1–21. <https://doi.org/10.1167/19.5.15>

1156 Loschky, L. C., & Wolverton, G. S. (2007). How late can you update gaze-contingent

1157 multiresolutional displays without detection? *ACM Transactions on Multimedia*
1158 *Computing, Communications and Applications*, 3(4):25, 1–10.
1159 <https://doi.org/10.1145/1314303.1314310>

1160 Malcolm, G. L., Groen, I. I. A., & Baker, C. I. (2016). Making sense of real-world scenes.
1161 *Trends in Cognitive Sciences*, 20(11), 843–856.
1162 <https://doi.org/10.1016/j.tics.2016.09.003>

1163 Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on
1164 visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*,
1165 9(11):8, 1–13. <https://doi.org/10.1167/9.11.8>

1166 Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I
1167 error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–
1168 315. <https://doi.org/10.1016/j.jml.2017.01.001>

1169 McConkie, G. W., & Loschky, L. C. (2002). Perception onset time during fixations in free
1170 viewing. *Behavior Research Methods Instruments & Computers*, 34(4), 481–490.
1171 <https://doi.org/10.3758/BF03195477>

1172 McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation
1173 in reading. *Perception & Psychophysics*, 17(6), 578–586.
1174 <https://doi.org/10.3758/BF03203972>

1175 McIlreavy, L., Fiser, J., & Bex, P. J. (2012). Impact of simulated central scotomas on visual
1176 search in natural scenes. *Optometry and Vision Science*, 89(9), 1385–1394.
1177 <https://doi.org/10.1097/OPX.0b013e318267a914>

1178 Mielle, S., Zhou, X., He, L., Rodger, H., & Caldara, R. (2010). Investigating cultural
1179 diversity for extrafoveal information use in visual scenes. *Journal of Vision*, 10(6):21,
1180 1–18. <https://doi.org/10.1167/10.6.21>

1181 Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau
1182 (2005). *Tutorial in Quantitative Methods for Psychology*, 4, 61–64.
1183 <https://doi.org/10.20982/tqmp.04.2.p061>

1184 Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual*
1185 *Cognition*, 21(7), 803–837. <https://doi.org/10.1080/13506285.2013.832449>

1186 Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in
1187 real-world scenes? Evidence from eye movements. *Journal of Experimental*
1188 *Psychology: Human Perception and Performance*, 40(1), 342–360.
1189 <https://doi.org/10.1037/a0033854>

1190 Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image

1191 features on fixation selection in scenes. *Annals of the New York Academy of Sciences*,
1192 1339(1), 82–96. <https://doi.org/10.1111/nyas.12705>

1193 Nuthmann, A., Einhäuser, W., & Schütz, I. (2017). How well can saliency models predict
1194 fixation selection in scenes beyond central bias? A new approach to model evaluation
1195 using generalized linear mixed models. *Frontiers in Human Neuroscience*, 11, 491.
1196 <https://doi.org/10.3389/fnhum.2017.00491>

1197 Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene
1198 viewing. *Journal of Vision*, 10(8):20, 1–19. <https://doi.org/10.1167/10.8.20>

1199 Nuthmann, A., & Malcolm, G. L. (2016). Eye guidance during real-world scene search: The
1200 role color plays in central and peripheral vision. *Journal of Vision*, 16(2):3, 1–16.
1201 <https://doi.org/10.1167/16.2.3>

1202 Nuthmann, A., Schütz, I., & Einhäuser, W. (2020). Saliency-based object prioritization
1203 during active viewing of naturalistic scenes in young and older adults. *Scientific*
1204 *Reports*, 10, 22057. <https://doi.org/10.1038/s41598-020-78203-7>

1205 Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology*
1206 *of attention* (pp. 251–256). San Diego, CA: Elsevier.

1207 Potter, M. C. (1975). Meaning in visual search. *Science*, 187(4180), 965–966.
1208 <https://doi.org/10.1126/science.1145183>

1209 Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science*, 206(4417), 468–469.
1210 <https://doi.org/10.1126/science.504987>

1211 Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network:*
1212 *Computation in Neural Systems*, 10(4), 341–350. [https://doi.org/10.1088/0954-](https://doi.org/10.1088/0954-898X/10/4/304)
1213 898X/10/4/304

1214 Reingold, E. M., & Loschky, L. C. (2002). Saliency of peripheral targets in gaze-contingent
1215 multiresolutional displays. *Behavior Research Methods Instruments & Computers*,
1216 34(4), 491–499. <https://doi.org/10.3758/BF03195478>

1217 Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Review of*
1218 *Vision Science*, 2(1), 437–457. <https://doi.org/10.1146/annurev-vision-082114-035733>

1219 Rutishauser, U., & Koch, C. (2007). Probabilistic modeling of eye movement data during
1220 conjunction search via feature-based attention. *Journal of Vision*, 7(6):5, 1–20.
1221 <https://doi.org/10.1167/7.6.5>

1222 Saunders, D. R., & Woods, R. L. (2014). Direct measurement of the system latency of gaze-
1223 contingent displays. *Behavior Research Methods*, 46(2), 439–447.
1224 <https://doi.org/10.3758/s13428-013-0375-5>

- 1225 Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori
 1226 contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*,
 1227 104038. <https://doi.org/10.1016/j.jml.2019.104038>
- 1228 Seedorff, M., Oleson, J., & McMurray, B. (2019). Maybe maximal: Good enough mixed
 1229 models optimize power while controlling Type I error. *PsyArXiv*.
 1230 <https://doi.org/10.31234/osf.io/xmhfr>
- 1231 Stoll, J., Thrun, M., Nuthmann, A., & Einhäuser, W. (2015). Overt attention in natural
 1232 scenes: Objects dominate features. *Vision Research*, *107*, 36–48.
 1233 <https://doi.org/10.1016/j.visres.2014.11.006>
- 1234 Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive*
 1235 *Psychology*, *12*(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- 1236 Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency
 1237 influence eye movements when inspecting pictures. *Quarterly Journal of Experimental*
 1238 *Psychology*, *59*(11), 1931–1949. <https://doi.org/10.1080/17470210500416342>
- 1239 Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye
 1240 movements during scene inspection: A test of the saliency map hypothesis. *European*
 1241 *Journal of Cognitive Psychology*, *18*(3), 321–342.
 1242 <https://doi.org/10.1080/09541440500236661>
- 1243 Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary
 1244 for object identification? Evidence from eye movements during the inspection of real-
 1245 world scenes. *Consciousness and Cognition*, *17*(1), 159–170.
 1246 <https://doi.org/10.1016/j.concog.2006.11.008>
- 1247 van Diepen, P. M. J., De Graef, P., & Van Rensbergen, J. (1994). On-line control of moving
 1248 masks and windows on a complex background using the ATVista videographics adapter.
 1249 *Behavior Research Methods Instruments & Computers*, *26*(4), 454–460.
 1250 <https://doi.org/10.3758/BF03204665>
- 1251 Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New
 1252 York: Springer. <https://doi.org/10.1007/978-0-387-21706-2>
- 1253 Wagenmakers, E.-J., Kryptos, A.-M., Criss, A. H., & Iverson, G. (2012). On the
 1254 interpretation of removable interactions: A survey of the field 33 years after Loftus.
 1255 *Memory & Cognition*, *40*(2), 145–160. <https://doi.org/10.3758/s13421-011-0158-0>
- 1256 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2d ed.). New York:
 1257 Springer.
- 1258 Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference.

1259 *Journal of the American Statistical Association*, 22(158), 209–212.
1260 <https://doi.org/10.1080/01621459.1927.10502953>

1261 Wolfe, J. M. (2015). Visual search. In A. Kingstone, J. M. Fawcett, & E. F. Risko (Eds.), *The*
1262 *Handbook of Attention* (pp. 27–56). MITGogNet.

1263 Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological*
1264 *Review*, 115(4), 787–835. <https://doi.org/10.1037/a0013118>

1265 Zhaoping, L. (2019). A new framework for understanding vision from the perspective of the
1266 primary visual cortex. *Current Opinion in Neurobiology*, 58, 1–10.
1267 <https://doi.org/10.1016/j.conb.2019.06.001>

1268 Zhaoping, L., & Frith, U. (2011). A clash of bottom-up and top-down processes in visual
1269 search: The reversed letter effect revisited. *Journal of Experimental Psychology: Human*
1270 *Perception and Performance*, 37(4), 997–1006. <https://doi.org/10.1037/a0023099>

1271 Zhaoping, L., & Guyader, N. (2007). Interference with bottom-up feature detection by
1272 higher-level object recognition. *Current Biology*, 17(1), 26–31.
1273 <https://doi.org/10.1016/j.cub.2006.10.050>
1274

1275

Tables

1276 Table 1

1277 *Linear and generalized linear mixed models (LLM and GLMM respectively) for Experiment 1: Means (b), standard errors (SE), and test*
 1278 *statistics (LLMs: t-values; GLMMs: z-values and p-values) for fixed effects*

| | | Intercept | Target size | Target salience | Foveal scotoma | Size × Salience | Size × Scotoma | Salience × Scotoma | Size × Salience × Scotoma |
|------------------------|-----------|-----------|-------------|-----------------|----------------|-----------------|----------------|--------------------|---------------------------|
| Probability correct | <i>b</i> | 2.71 | 0.41 | 0.56 | -0.7 | -0.73 | 0.08 | 0.1 | -0.67 |
| | <i>SE</i> | 0.15 | 0.13 | 0.13 | 0.13 | 0.25 | 0.25 | 0.25 | 0.51 |
| | <i>z</i> | 18.35 | 3.22 | 4.43 | -5.49 | -2.87 | 0.32 | 0.38 | -1.32 |
| | <i>p</i> | < .001 | 0.001 | < .001 | < .001 | 0.004 | 0.746 | 0.704 | 0.186 |
| Search time | <i>b</i> | 2086.33 | -927.79 | -1230.15 | 149.09 | 958.04 | -62.7 | 75.1 | -251.91 |
| | <i>SE</i> | 112.75 | 99.45 | 121.26 | 81.54 | 166.65 | 101.36 | 128.17 | 202.45 |
| | <i>t</i> | 18.5 | -9.33 | -10.14 | 1.83 | 5.75 | -0.62 | 0.59 | -1.24 |
| Search initiation time | <i>b</i> | 269.41 | -9.02 | -19.83 | 38.14 | -9.33 | -8.04 | -11.61 | 0.35 |
| | <i>SE</i> | 8.35 | 3.44 | 4.1 | 12.1 | 6.88 | 6.88 | 6.87 | 13.76 |
| | <i>t</i> | 32.26 | -2.62 | -4.84 | 3.15 | -1.35 | -1.17 | -1.69 | 0.03 |
| Scanning time | <i>b</i> | 1127.95 | -723.15 | -968.64 | -16.72 | 768.16 | 6.73 | 185.87 | -296.28 |
| | <i>SE</i> | 71.43 | 80.07 | 104.24 | 46.4 | 135.65 | 92.8 | 92.7 | 185.42 |
| | <i>t</i> | 15.79 | -9.03 | -9.29 | -0.36 | 5.66 | 0.07 | 2.01 | -1.6 |
| Verification time | <i>b</i> | 677.29 | -178.51 | -225.95 | 118.25 | 172.78 | -60.1 | -92.32 | 42.14 |
| | <i>SE</i> | 69.39 | 36.97 | 37.34 | 45.73 | 64.39 | 50.45 | 57.28 | 100.78 |
| | <i>t</i> | 9.76 | -4.83 | -6.05 | 2.59 | 2.68 | -1.19 | -1.61 | 0.42 |
| Saccade amplitude | <i>b</i> | 5.3 | -0.43 | -0.57 | 0.4 | -0.22 | -0.26 | 0.01 | -0.4 |
| | <i>SE</i> | 0.11 | 0.06 | 0.1 | 0.08 | 0.12 | 0.12 | 0.12 | 0.24 |
| | <i>t</i> | 49.19 | -6.96 | -5.99 | 4.89 | -1.81 | -2.11 | 0.04 | -1.63 |
| Fixation duration | <i>b</i> | 204.56 | -9.49 | -20.01 | 19.85 | 7.68 | 7.85 | 7.09 | 3.63 |
| | <i>SE</i> | 4.1 | 2.55 | 2.86 | 3.88 | 3.52 | 4.09 | 3.52 | 9.69 |
| | <i>t</i> | 49.85 | -3.72 | -6.99 | 5.12 | 2.18 | 1.92 | 2.02 | 0.37 |

1279 *Note: Non-significant coefficients are set in bold (LLMs: $|t| < 1.96$; GLMMs: $p > .05$). See text for further details.*

1280 Table 2

1281 *Linear and generalized linear mixed models (LLM and GLMM respectively) for Experiment 2: Means (b), standard errors (SE), and test*

1282 *statistics (LLMs: t-values; GLMMs: z-values and p-values) for fixed effects*

| Dependent variable | Contrast coding (scotoma type) | Reference level | Scot 1 (definition) | Scot 2 (definition) | | Intercept | Target salience | Scot 1 | Scot 2 | Saliency × Scot 1 | Saliency × Scot 2 |
|------------------------|--------------------------------|-----------------|---------------------|---------------------|-----------|-----------|-----------------|-------------|--------------|-------------------|-------------------|
| Probability correct | BWD | No – P – C | P - No | C - P | <i>b</i> | 1.76 | 1.21 | -1.44 | -0.77 | -0.62 | 0.74 |
| | | | | | <i>SE</i> | 0.15 | 0.1 | 0.14 | 0.13 | 0.27 | 0.19 |
| | | | | | <i>z</i> | 11.9 | 11.94 | -10.19 | -6.01 | -2.31 | 3.82 |
| | | | | | <i>p</i> | < .001 | < .001 | < .001 | < .001 | 0.021 | < .001 |
| Search time | BWD | No – C – P | C - No | P - C | <i>b</i> | 3932.88 | -1523.04 | 756.36 | 2995.29 | -707.37 | 1761.53 |
| | | | | | <i>SE</i> | 123.55 | 156.45 | 117.71 | 214.97 | 227.02 | 354.56 |
| | | | | | <i>t</i> | 31.83 | -9.74 | 6.43 | 13.93 | -3.12 | 4.97 |
| | | | | | | | | | | | |
| Search initiation time | simple | no scotoma | C - No | P - No | <i>b</i> | 273.21 | -8.62 | 17.09 | 81.99 | 4.06 | 12.85 |
| | | | | | <i>SE</i> | 6.83 | 3.77 | 8.18 | 7.43 | 8.91 | 9.45 |
| | | | | | <i>t</i> | 39.98 | -2.29 | 2.09 | 11.03 | 0.46 | 1.36 |
| | | | | | | | | | | | |
| Scanning time | simple | no scotoma | C - No | P - No | <i>b</i> | 2709.84 | -1043.68 | 85.21 | 3725.53 | 6.18 | 978.91 |
| | | | | | <i>SE</i> | 92.54 | 124.55 | 108.27 | 187.46 | 209.5 | 304.12 |
| | | | | | <i>t</i> | 29.28 | -8.38 | 0.79 | 19.87 | 0.03 | 3.22 |
| | | | | | | | | | | | |
| Verification time | simple | no scotoma | C - No | P - No | <i>b</i> | 910.92 | -442.75 | 564.89 | -36.76 | -660.97 | 54.2 |
| | | | | | <i>SE</i> | 59.84 | 63.09 | 98.63 | 64.28 | 172.18 | 84.33 |
| | | | | | <i>t</i> | 15.22 | -7.02 | 5.73 | -0.57 | -3.84 | 0.64 |
| | | | | | | | | | | | |
| Saccade amplitude | simple | no scotoma | C - No | P - No | <i>b</i> | 4.87 | -0.42 | 1.6 | -2.2 | -0.18 | 0.43 |
| | | | | | <i>SE</i> | 0.1 | 0.06 | 0.15 | 0.1 | 0.16 | 0.08 |
| | | | | | <i>t</i> | 47.11 | -6.5 | 10.63 | -21.7 | -1.16 | 5.27 |
| | | | | | | | | | | | |
| Fixation duration | BWD | No – C – P | C - No | P - C | <i>b</i> | 211.77 | -12.24 | 17.09 | 18.79 | 11.58 | 10.29 |
| | | | | | <i>SE</i> | 3.82 | 2.12 | 4.5 | 6.33 | 4.07 | 4.23 |
| | | | | | <i>t</i> | 55.42 | -5.78 | 3.8 | 2.97 | 2.85 | 2.43 |
| | | | | | | | | | | | |

1283 *Note: Non-significant coefficients are set in bold (LLMs: $|t| < 1.96$; GLMMs: $p > .05$). See text for further details.*

Appendix A

Mathematical Definition of Local Contrast

1284

1285

1286

1287 As a measure of visual salience, the root-mean-square (RMS) contrast was calculated as

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

1288 where L is either 11 (patch width 23) or 23 (patch width 47), $p(r, c)$ is the pixel value at row r

1289 and column c , $\bar{p}(R, C)$ is the mean of the patch calculated as

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

1290 and \bar{p}_{img} is the mean of the image.

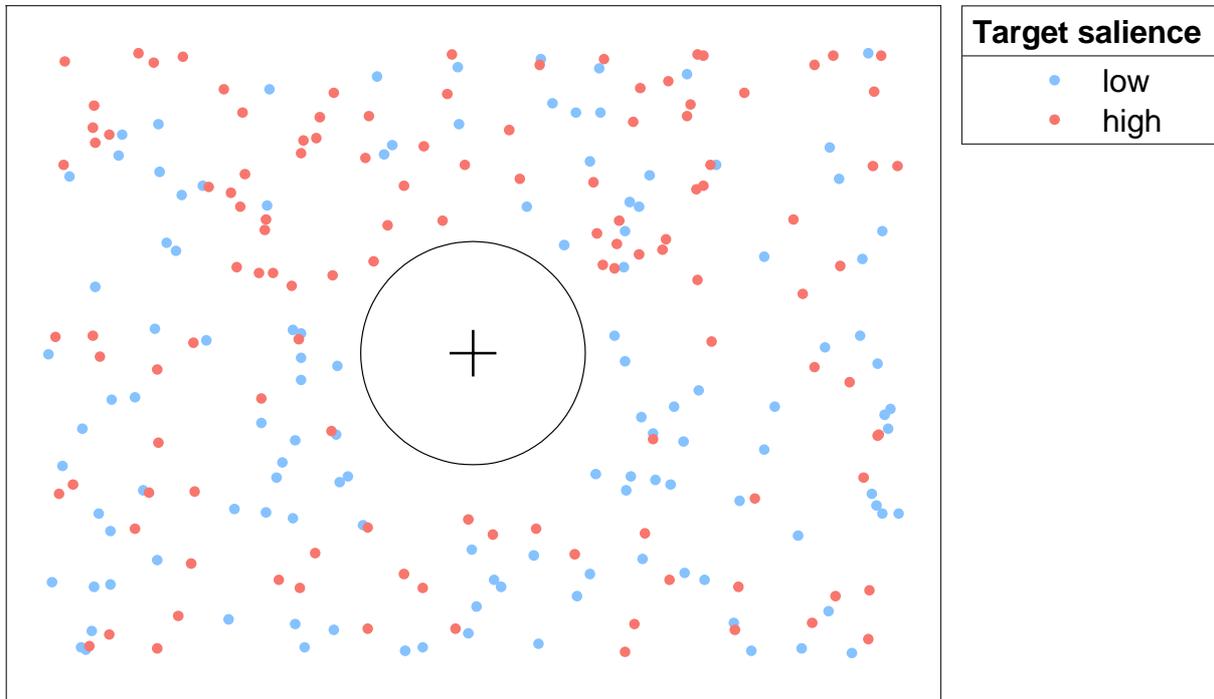
1291

1292

1293
1294
1295

Appendix B

Distribution of Search Targets in the Scenes



1296
1297
1298
1299
1300
1301

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°).

