

1
2 **On the Relative (Un)importance of Foveal Vision during Letter Search in Naturalistic**
3 **Scenes**

4
5 Adam C. Clayden^{1,2}, Robert B. Fisher³ and Antje Nuthmann^{1,4}

6 ¹ Psychology Department, School of Philosophy, Psychology and Language Sciences,
7 University of Edinburgh, UK

8 ² School of Engineering, Arts, Science and Technology, University of Suffolk, UK

9 ³ School of Informatics, University of Edinburgh, UK

10 ⁴ Institute of Psychology, University of Kiel, Germany

11
12
13
14 **Author Note**

15
16 Antje Nuthmann  <http://orcid.org/0000-0003-3338-3434>

17 The T.E.A is available at <https://github.com/AdamClayden93/tea>.

18 The project was supported by the Engineering and Physical Sciences Research
19 Council (UK). Portions of this research were presented at the 18th European Conference on
20 Eye Movements (Vienna, Austria, 2015).

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

24
25

26 **Abstract**

27 The importance of high-acuity foveal vision to visual search can be assessed by denying
28 foveal vision using the gaze-contingent Moving Mask technique. Foveal vision was
29 necessary to attain normal performance when searching for a target letter in alphanumeric
30 displays, *Perception & Psychophysics*, 62 (2000) 576-585. In contrast, foveal vision was not
31 necessary to correctly locate and identify medium-sized target objects in natural scenes,
32 *Journal of Experimental Psychology: Human Perception and Performance*, 40 (2014) 342-
33 360. To explore these task differences, we used grayscale pictures of real-world scenes which
34 included a target letter (Experiment 1: T, Experiment 2: T or L). To reduce between-scene
35 variability with regard to target salience, we developed the Target Embedding Algorithm
36 (T.E.A.) to place the letter in a location for which there was a median change in local contrast
37 when inserting the letter into the scene. The presence or absence of foveal vision was crossed
38 with four target sizes. In both experiments, search performance decreased for smaller targets,
39 and was impaired when searching the scene without foveal vision. For correct trials, the
40 process of target localization remained completely unimpaired by the foveal scotoma, but it
41 took longer to accept the target. We reasoned that the size of the target may affect the
42 importance of foveal vision to the task, but the present data remain ambiguous. In summary,
43 the data highlight the importance of extrafoveal vision for target localization, and the
44 importance of foveal vision for target verification during letter-in-scene search.

45

46 243 words

47

48 *Keywords:* naturalistic scenes; visual search; target size; eye movements; foveal
49 vision

50

51 1. Introduction

52 How important is the availability of high-acuity foveal vision to visual search? This
53 question has been investigated with different search tasks, ranging from letter search in
54 alphanumeric displays (Bertera & Rayner, 2000) to object-in-scene search (Nuthmann, 2014),
55 either highlighting the relative importance (letter search) or unimportance (scene search) of
56 foveal vision. The aim of the present work was to combine design features from both search
57 paradigms to better understand these task differences. In Experiment 1, observers searched
58 for the letter “T” embedded in grayscale pictures of real-world scenes, with or without foveal
59 vision. In Experiment 2, we added a letter recognition component to the search task (“Is it a T
60 or an L?”). In both experiments, we also varied the size of the letter target to investigate the
61 degree to which the importance of foveal vision depends on the size of the search target.

62 Visual acuity is highest at the fovea before declining rapidly as it approaches the
63 periphery (Strasburger, Rentschler, & Jüttner, 2011, for review). Whereas the foveal region is
64 typically defined as the central 2° of vision, the parafoveal region extends from the foveal
65 region out to about 5° from fixation; the fovea and parafovea together are commonly referred
66 to as central vision. The peripheral region is everything beyond the parafoveal region. During
67 each eye fixation, information may be extracted from foveal, parafoveal, and peripheral
68 regions of the visual field.

69 The importance of foveal vision was first studied in sentence reading by means of the
70 gaze-contingent *Moving Mask* technique. To this end, Rayner and Bertera (1979) aligned a
71 visual mask with the reader’s gaze to wipe out the text in view. The size of the mask ranged
72 between 1 and 17 characters (1° = three characters). Simulating reading without a fovea in
73 that manner reduced the reader’s reading speed by increasing the number of fixations,
74 fixation duration, and reducing saccade length. Moreover, reading comprehension suffered.
75 The same authors also investigated the importance of foveal vision in visual search (Bertera

76 & Rayner, 2000). In this study, participants searched for the target letter “y” within a
77 randomly arranged array of alphanumeric characters, with or without a simulated scotoma.
78 Five different scotoma sizes, ranging from 0.3° to 3°, were tested. As the mask size increased,
79 the lower the search accuracy, the longer the search time, and the more fixations were made.
80 Geringswald, Baumgartner, and Pollmann (2012) investigated the impact of a large simulated
81 central scotoma (diameter: 9°) on contextual cueing in visual search. Participants searched for
82 a T-shaped target among L-shaped distractors. Blocking out central vision eliminated the
83 search facilitation which is oftentimes observed for targets appearing in repeated
84 configurations (see also Geringswald & Pollmann, 2015).

85 Interestingly, visual search studies involving naturalistic scenes have found rather
86 different results (McIlreavy, Fiser, & Bex, 2012; Nuthmann, 2014). In the study by
87 Nuthmann (2014), participants searched for a specific object in a colored image of a real-
88 world scene (e.g., a blender in a kitchen scene). Search was cued with a word label and
89 search objects had an average size of 2.5° × 2.5° (medium size). When searching the scene
90 with artificially impaired foveal or central vision¹, search performance was surprisingly
91 unimpaired. *Foveal* vision was not necessary to attain normal search performance. When
92 searching without *central* vision, participants’ gaze data revealed that they were not impaired
93 in locating the search object in the scene, but in verifying that the target was in fact the target.
94 In the study by Nuthmann (2014), the scene image contained contextually relevant search
95 targets (cf. Torralba, Oliva, Castelhana, & Henderson, 2006). McIlreavy et al. (2012)
96 excluded such contextual guidance towards the target by asking observers to look for spatial
97 distortions (Bex, 2010), which were embedded at random places in grayscale images of
98 natural scenes. The results for search times were similar to the ones by Nuthmann (2014).

¹ The size (i.e., radius) of the scotoma was manipulated as the standard deviation of the two-dimensional Gaussian distribution that was used to mix the high-resolution foreground with a low-resolution background image; foveal scotoma: $\sigma_{x,y} = 1.6^\circ$, central scotoma: $\sigma_{x,y} = 4.1^\circ$.

99 Searching with a foveal scotoma ($\sigma_{x,y} = 1^\circ$) had no detrimental effect on performance. Only
100 the largest central scotoma condition ($\sigma_{x,y} = 4^\circ$) led to a significant increase in mean search
101 time.

102 During overt search of any kind, the information extracted during eye fixations
103 subserves both a peripheral selection task as well as a central discrimination task (Hooge &
104 Erkelens, 1999; Shen, Reingold, Pomplun, & Williams, 2003). The peripheral selection task
105 determines the target location for the next saccade, whereas the central discrimination task
106 involves an accept/reject decision about whether the fixated object is the target. Since foveal
107 analysis allows for encoding fine perceptual detail, making foveal vision unavailable should
108 be disruptive to the central discrimination task. However, such reasoning ignores the fact that
109 the processing of the fixated object or region can begin prior to the start of fixation via
110 extrafoveal processing (Reichle & Reingold, 2013; Reingold & Glaholt, 2014). Thus, simple
111 search and scene search may differ in the way extrafoveal processing enables the extraction
112 of information that is required to reject distractors and to accept the target. Moreover, the
113 relation between foveal analysis and peripheral selection may be task dependent (cf. Shen et
114 al., 2003).

115 Target size is a feature that may be relevant in this regard. Both McIlreavy et al.
116 (2012) and Nuthmann (2014) discuss that target size could be an important mediating factor
117 for their findings on the (un)importance of foveal vision. Before elaborating on this
118 argument, we briefly review research on size and eccentricity effects in (normal) visual
119 search. A common paradigm is to use fairly small simple displays which observers search
120 covertly in the absence of eye movements. Using this approach, Duncan and Humphreys
121 (1989) investigated the effect of stimulus size and showed that search is more difficult for
122 small letters than for large letters. A related finding is the eccentricity effect: search
123 performance deteriorates as the target is presented at farther peripheral locations (Carrasco,

124 Evert, Chang, & Katz, 1995; Geisler & Chou, 1995). This reduction in search efficiency may
125 be due to the poorer spatial resolution in the periphery. Consistent with this view, enlarging
126 the stimuli according to the cortical magnification factor (Rovamo & Virsu, 1979) eliminated
127 the eccentricity effect (Carrasco & Frieder, 1997; Carrasco, McLean, Katz, & Frieder, 1998;
128 but see Wolfe, O'Neill, & Bennett, 1998, Experiment 4a). The eccentricity effect is also
129 observed in the presence of eye movements (Scialfa & Joffe, 1998; Zelinsky, 2008).

130 In the context of visual search in real-world scenes, the effect of target size has
131 received little systematic investigation. Wolfe, Alvarez, Rosenholtz, Kuzmova, and Sherman
132 (2011, Experiment 1) had observers search for annotated objects in photographs of real-world
133 scenes. The objects showed a natural variability in size and eccentricity and search times
134 were found to increase for both smaller as well as more eccentric targets. Miellet, Zhou, He,
135 Rodger, and Caldara (2010) asked both Eastern and Western observers to search for animals
136 in zoo photographs. In the experiment, target size and the size of a gaze-contingent moving
137 mask were parametrically manipulated (size/ diameter: 2°, 5°, or 8°). Search performance was
138 better for larger targets. As the simulated scotoma got larger, performance increasingly
139 suffered (cf. McIlreavy et al., 2012; Nuthmann, 2014). Importantly, there was an interaction
140 between mask size and target size such that the deleterious effect of mask size was more
141 pronounced for smaller targets. In the 2°-*Blindspot* condition, making foveal vision
142 unavailable, search performance was reduced for 2° targets but not for 8° targets. Although
143 suggestive, any findings involving target size in this study need to be treated cautiously
144 because target salience (Itti & Koch, 2000) was not controlled for. Other potential confounds
145 are target eccentricity (i.e., distance from scene center) and contextual guidance.

146 The goal of the present research was to further investigate the importance of foveal
147 vision to visual search. Stimuli were grayscale pictures of real-world scenes in which a target
148 letter was inserted (Experiment 1: T, Experiment 2: T or L). Four letter sizes, ranging from

149 0.25° to 1.5° in width, were crossed with the presence vs. absence of foveal vision. To control
150 for visual salience, the letter was algorithmically placed for each scene in a location for
151 which there was a medium change in local contrast when inserting the letter. Letter targets
152 were used for a number of reasons. The small to large animal targets in Miellet et al. (2010)
153 were all part of different scenes. Our approach allowed us to place letter targets of variable
154 size at the same location within a given scene. In addition, using context-free letter targets
155 rather than contextually relevant search targets prevents observers from using their
156 knowledge about the likely positions of targets to guide their eye movements (cf. McIlreavy
157 et al., 2012). Our task still approximates natural behavior because there are real-world
158 searches for which there is minimal guidance by scene context (e.g., search for a fly). Perhaps
159 more importantly, scene processing and object identification are not totally suppressed when
160 searching for a “T” overlaid onto the scene (T. H. W. Cornelissen & Vö, 2017). In
161 Experiment 1, on each trial participants were asked to look for the letter “T”. In Experiment
162 2, we added a recognition component to the task. The target was either a “T” or an “L”,
163 and—once they found the letter—participants had to indicate which one it was. We chose
164 these two letters because they share exactly the same features (strokes) and differ only in
165 their spatial arrangement (Duncan & Humphreys, 1989). Because we used participants’ eye-
166 movement data to verify that targets had indeed been found, there were no target-absent trials
167 (Nuthmann, 2013, 2014; Nuthmann & Malcolm, 2016).²

168 If foveal vision is necessary to achieve normal search performance during letter-in-
169 scene search, then we should observe a reduction in performance—lower search accuracy and
170 longer search time—when searching the scene with a simulated foveal scotoma, compared
171 with a normal-vision control condition. Moreover, we expected to find effects of target size,

² In a typical laboratory search experiment, the observer’s task is to establish whether the target is present or absent amongst other distractor items (Wolfe, 2014). In the present experiments, observers are asked to acquire the target with their eyes so the task has been referred to as target acquisition rather than search (Zelinsky, 2008).

172 with better performance for larger targets. Critically, the experimental design allowed us to
173 investigate whether the importance of foveal vision depended on the size of the search target
174 (cf. Mielliet et al., 2010). Why would size matter? Here, our hypotheses concern two separate
175 sub-processes of search: scanning for the target and accepting the target. The scanning
176 process involves the localization of the target in space, the duration of which (scanning time)
177 is indexed as the time between the first saccade and the first fixation on the target (Malcolm
178 & Henderson, 2009). Similarly, verification time is the elapsed time between the beginning
179 of the first fixation on the target and search termination.

180 The possibility exists that the actual search process, indexed by the scanning time, is
181 slowed down when foveal vision is absent. The reject decision during scanning epoch
182 fixations may be impaired if the extraction of information in extrafoveal vision (on the
183 previous, but also on the current fixation) cannot compensate for the lack of foveal analysis.
184 Moreover, the difficulty of central discrimination may affect the efficiency of peripheral
185 selection, if the two tasks share resources (see Shen et al., 2003, for discussion).
186 Alternatively, blocking out foveal vision may only affect the verification process, as
187 explained next.

188 Upon fixation with a foveal scotoma, all of the target—or some part of it—will be
189 covered by the scotoma. The extent of this masking depends on both the size of the target and
190 the initial fixation position on the search target (Nuthmann, 2014). If the available
191 information is not sufficient to make the accept decision, the eyes may move off the target to
192 unmask the letter and to process it in parafoveal or peripheral vision (cf. Nuthmann, 2014).
193 Such behavior would increase verification times. We hypothesized that any detrimental effect
194 of the foveal scotoma may only occur for smaller targets, or may be more pronounced for
195 those. Moreover, in Experiment 2 we changed the task to involve not only target detection
196 but also target identification. At least for small letters, letter identification may require the

197 extraction of fine detail via foveal analysis. Therefore, we reasoned that any adverse effect of
198 the foveal scotoma, and its interaction with target size, may be stronger in Experiment 2 than
199 in Experiment 1.

200 **2. Methods**

201 *2.1. Participants*

202 Thirty-two participants (12 males) between the ages of 18 and 27 (mean age 20 years)
203 participated in Experiment 1. Thirty-two different participants (8 males) between the ages of
204 18 and 27 (mean age 22 years) participated in Experiment 2. All participants had normal or
205 corrected-to-normal vision by self-report. They gave their written consent prior to the
206 experiment and either received study credit or were paid at a rate of £7 per hour for their
207 participation. Ethics approval was obtained from the Psychology Research Ethics Committee
208 of the University of Edinburgh.

209 *2.2. Apparatus*

210 Working with gaze-contingent displays requires minimizing the latency of the system.
211 This was achieved by using (a) an eye tracker with high temporal resolution, (b) modern
212 graphics hardware, and (c) a monitor with a high refresh rate. Stimuli were presented on a 21-
213 inch CRT monitor with a refresh rate of 140 Hz at a viewing distance of 90 cm, taking up a
214 $24.8^\circ \times 18.6^\circ$ (width \times height) field of view. A chin and forehead rest was used to keep the
215 participants' head position stable. During stimulus presentation, the eye movements of the
216 participants were recorded binocularly with an SR Research EyeLink 1000 Desktop mount
217 system with high accuracy (0.15° best, $0.25\text{-}0.5^\circ$ typical) and high precision (0.01° RMS).
218 The EyeLink 1000 was equipped with the 2000 Hz camera upgrade, allowing for binocular
219 recordings at a sampling rate of 1000 Hz per eye. The experiments were programmed in
220 MATLAB 2013a (The MathWorks, Natick, MA) using the OpenGL-based Psychophysics
221 Toolbox 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007) which incorporates the EyeLink

222 Toolbox extensions (F. W. Cornelissen, Peters, & Palmer, 2002). A game controller was used
223 to record participants' behavioral responses.

224 2.3. *Stimuli*

225 In Experiment 1, stimuli consisted of 120 grayscale images of naturalistic scenes (800
226 × 600 pixels), which came from a variety of categories; 104 of these photographs were
227 previously used as colored images in Nuthmann (2014). Example scenes are shown in
228 Figures 1 and 3. Eight additional images were used as practice scenes. Image processing
229 techniques (See Section Target Embedding Algorithm below) were used to insert the letter T
230 in four sizes at the same location within a given scene, such that the chosen location was of
231 median salience, as explained below. Note that in the experiment, each participant viewed a
232 given scene only once, in one of the four target size conditions (and either with or without
233 foveal vision).

234 In Experiment 2, 128 (+ 8 practice) grayscale images of real-world scenes were used,
235 120 of which were from experiment 1 with 8 new images. The new images were chosen
236 because the experimental design required an equal number of T- and L-scenes in each target-
237 size condition. The search target was either a letter T or L that was again algorithmically
238 placed into the scene at a median salience location.

239 2.4. *Design*

240 Both experiments used a 2 × 4 within-subjects design with 2-level factor foveal vision
241 (present vs. absent) and 4-level factor target size. The factor foveal vision refers to the
242 implementation of a foveal scotoma. In the scotoma condition, foveal vision was blocked by
243 a gaze-contingent moving mask (foveal vision absent, or scotoma on). This was contrasted
244 with a normal-vision control condition (foveal vision present, or scotoma off).

245 In both experiments, the presence or absence of foveal vision was crossed with four
246 target sizes. In Experiment 1, they were equally spaced as follows: S - Small (letter width

247 0.25°), M - Medium (0.66°), L - Large (1.08°), and XL - Extra Large (1.5°). The XL target
248 size was chosen such that the foveal scotoma, which had a radius of 1°, completely obscured
249 the target when observers fixated the center of the letter. In Experiment 2, we removed the
250 XL targets; instead, we added targets of intermediate size (0.41°) halfway between the small
251 and medium targets. These adjustments were informed by the results obtained in Experiment
252 1: search efficiency was much worse for small targets compared with medium-sized targets,
253 while performance differences between large and extra-large targets were much less
254 pronounced.

255 In Experiment 1, the 120 T-scenes were assigned to eight lists of 15 scenes each. The
256 scene lists were rotated over participants, such that a given participant was exposed to a list
257 for only one of the eight experimental conditions created by the 2×4 design. There were
258 eight groups of four participants, and each group of participants was exposed to unique
259 combinations of list and experimental condition. To summarize, participants viewed each of
260 the 120 scene items once, with 15 scenes in each of the eight experimental conditions. Across
261 the 32 participants, each scene item appeared in each condition four times.

262 For Experiment 2, each of the 128 original scene images was submitted to the Target
263 Embedding Algorithm to produce four T-scenes and four L-scenes, one for each target size.
264 In the experiment, half of the original scenes were used as T-scenes, the other half as L-
265 scenes. Since the algorithm placed the Ts and Ls of four different sizes in the same location,
266 there were a few cases where the horizontal bar of the T or the vertical bar of the L blended
267 with a dark scene background. Therefore, the decision about which scenes to use in either
268 category was guided by visual inspection. We then created eight scene lists, each comprising
269 eight T-scenes and eight L-scenes. Apart from that, the same counterbalancing procedure as
270 in Experiment 1 was used to control for item effects.

271 The foveal vision manipulation was blocked so that participants completed two blocks
272 of trials in the experiment: in one block observers' foveal vision was available, in the other
273 block it was obstructed by a gaze-contingent scotoma. Each block started with four practice
274 trials, one for each target size condition. The order of blocks was counterbalanced across
275 subjects. Within a block, scenes with targets of different sizes and types (Experiment 2 only)
276 were presented randomly.

277 *2.4.1. Target Embedding Algorithm – T.E.A.*

278 It is important to manipulate target size within scenes rather than between scenes.
279 Regarding target placement, different degrees of randomness are conceivable. If the target
280 was placed randomly on a given trial, targets of different sizes would be located at varying
281 eccentricities in a given scene. Moreover, the degree to which the target stands out from its
282 neighboring regions (i.e., its visual salience) would differ widely between scenes and
283 between target sizes per scene. Therefore, it is important to place targets of different sizes at
284 the same location within a given scene. In principle, this common location can be picked
285 randomly (McIlreavy et al., 2012). When using letter targets, random placement would
286 inevitably lead to considerable differences in target salience between scenes. To reduce this
287 variability, we developed a target embedding algorithm (T.E.A.) that took target salience into
288 account.

289 While there are many methods of constructing salience maps for images of real-world
290 scenes (Borji, Sihite, & Itti, 2013), it is widely held that simple stimulus features such as
291 color, orientation and intensity (luminance contrast) contribute to the computation of visual
292 salience (Itti & Koch, 2000). Using the output of a computational model of visual salience as
293 input for our algorithm would be prohibitively computationally expensive. As a practical
294 alternative, we used a version of root-mean-square (RMS) contrast: when stepping through
295 the scene, the standard deviation of luminance values of all pixels in the evaluated region was

296 divided by the mean luminance of the image. Calculating luminance contrast this way is
297 consistent with measures of detectability in natural scenes (Bex & Makous, 2002), and with
298 filter properties of early vision (Moulden, Kingdom, & Gatley, 1990). Moreover, it has been
299 used in experimental studies on fixation selection in scenes (e.g., Nuthmann & Einhäuser,
300 2015; Reinagel & Zador, 1999).

301 The target was placed at an image position that caused a median RMS contrast
302 change. To compute this, a rectangular region that was slightly larger than the target moved
303 pixel-by-pixel through the image. The RMS contrast M_o was calculated at each position.
304 Afterwards, the target was inserted and the RMS contrast M_w was computed at each position.
305 By computing $\Delta C = M_w - M_o$ at each pixel, we obtained an image map comprising the
306 contrast difference values within the image. After calculating the contrast difference map for
307 each target size, the four resultant maps were summed together to obtain a final summed
308 difference map. This summing acted as a way for the algorithm to compute a single location
309 for all target sizes, as the values of each individual difference map varied slightly. The
310 distribution of values from the summed map was computed. From the distribution different
311 contrast levels could be selected to control the desired amount of contrast change arising
312 from placing the letter in the scene. We used the median contrast difference as a compromise
313 between harder (smaller contrast difference) and easier (larger contrast difference) target
314 positions.

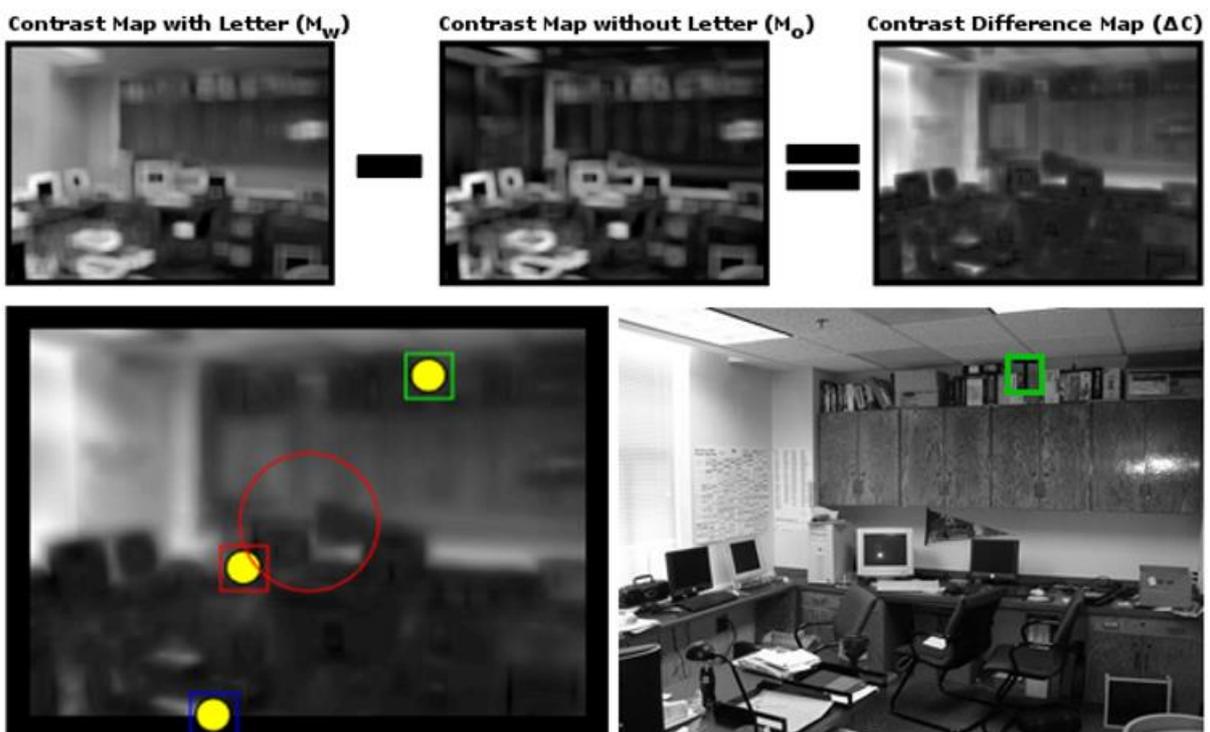
315 This final map was then probed by our algorithm to locate all pixel (i.e., potential
316 target) positions with the median change in contrast. Some positions were eliminated by the
317 following two criteria. First, locations within 3° of visual angle from the center were excluded
318 from evaluation due to the central region being the initial location of both the participant's
319 gaze and the gaze-contingent scotoma. Participants were not aware of this constraint. Second,
320 locations at the boundaries of the image were also excluded to avoid truncation of the letter.

321 From all remaining possible median contrast target positions, one was selected at random as
 322 the location of the target for that stimulus.

323 For Experiment 2, the algorithm was extended to handle multiple target letters. In this
 324 case, a new ‘TL’ contrast difference map was generated by computing:

$$\Delta C(r, c) = \sum_{\mathcal{L}_s} |\Delta C_{\mathcal{L}}^{[s]}(r, c) - t_{\mathcal{L}_s}|$$

325 where $\Delta C_{\mathcal{L}}^{[s]}$ is the difference map for a given font size [s] and letter $\mathcal{L} \in \{T, L\}$, with [r,c]
 326 denoting the map’s rows and columns. Each of its values were then subtracted by the median
 327 contrast of a given map, denoted by $t_{\mathcal{L}_s}$. This process was repeated for both letters and all
 328 four scales before adding the resultant image maps together. By subtracting $t_{\mathcal{L}_s}$, the lowest
 329 value in this new map (with a minimum of zero) is the pixel closest to the target value $t_{\mathcal{L}_s}$,
 330 and the coordinates of this pixel defined the target position for that image. As before, central
 331 and boundary pixel positions were eliminated from consideration. Figure 1 provides an
 332 illustration by depicting the contrast difference map and the algorithmic probing.
 333



335 *Figure 1.* Illustration of the T.E.A. depicting the initial creation of the contrast difference
336 map. The T.E.A. creates 3 contrast maps (from left to right): contrast with the letter placed at
337 each pixel position, contrast without the letter, and the difference between them. Using the
338 contrast difference map, the algorithm then probes the scene, excluding locations near the
339 boundary of the screen (example: blue boxed dot) and inside the central circle (example: red
340 boxed dot). If multiple positions are found (example: yellow dots), one satisfying the above
341 two constraints is chosen at random (example: green boxed dot) for target insertion.

342

343 2.4.2. *Creation of gaze-contingent scotoma*

344 The foveal scotoma was created using texture-mapping and OpenGL (Open Graphics
345 Library). This technique provides various blending operations that enable simple image
346 combinations to take place via an image's alpha channel (see Duchowski & Çöltekin, 2007,
347 for details on the general technique). The scotoma was a symmetric circular mask with a
348 radius of 1°. The scotoma size was chosen to completely obscure foveal vision. The foveal
349 mask moved concomitantly with the participant's gaze. To this end, the average horizontal
350 and vertical position of the two eyes (Nuthmann, 2013, for discussion) was continuously
351 evaluated online. Updating the display contingent on the viewer's gaze required 1 ms to
352 receive a sample from the eye tracker, less than 1 ms to draw the image textures and up to 7
353 ms to refresh the screen. Thus, the display was updated depending on observers' gaze
354 position in close to real time. A detailed account of the gaze-contingent implementation is
355 provided in Nuthmann (2013, 2014).

356 There are some subtle differences between the implementation of the foveal scotoma
357 in a previous study from our lab (Nuthmann, 2014) and here. Nuthmann (2014) used full-
358 color images, and foveal vision was degraded by applying a very strong low-pass filter to the
359 currently fixated scene region (the foveal scotoma was only one of six conditions with

360 degraded vision). Moreover, a Gaussian mask was used, and the size of the scotoma was
361 defined as the standard deviation of the two-dimensional Gaussian distribution (1.6° for the
362 foveal scotoma, or small *Blindspot*). In the present experiments, using grayscale images, we
363 used a circular mask drawn in gray. To avoid a sharp-boundary mask and to reduce
364 perceptibility of slight mask position jitter, the perimeter of the circular mask was slightly
365 faded through low-pass filtering, while the interior remained untouched. When investigating
366 the importance of foveal vision (i.e., a relatively small region of the visual field), it seems
367 more appropriate to define the size of the moving mask as the radius of a circle rather than
368 the standard deviation of a Gaussian.

369 2.5. Procedure

370 At the beginning of the experiment, a 9-point calibration procedure was performed,
371 followed by a 9-point calibration accuracy test (validation). At the beginning of each trial a
372 fixation cross was presented at the center of the screen for 600 ms, and acted as a fixation
373 check. The fixation check was deemed successful if gaze position, averaged across both eyes,
374 continuously stayed within an area of 40×40 pixels ($1.24^\circ \times 1.24^\circ$) for 200 ms. If this
375 condition was not met, the fixation check timed out after 500 ms. In this case, the fixation
376 check procedure was either repeated or replaced by another calibration procedure. If the
377 fixation check was successful, the scene image appeared on the screen. Once subjects had
378 found the target letter, they were instructed to fixate their gaze on it and press a button on the
379 controller to end the trial (cf. Glaholt, Rayner, & Reingold, 2012; Nuthmann, 2014). In
380 experiment 1, participants could press any button to indicate that they had found the T. Upon
381 identifying the target in Experiment 2, observers pressed one of two triggers on the controller
382 corresponding to either “T” or “L”. Trials timed-out 15 s after stimulus presentation if no
383 response was made. There was an inter-trial interval of 1 s before the next fixation cross was
384 presented.

385 2.6. Data analysis

386 The SR Research Data Viewer software with default settings was used to convert the
387 raw data obtained by the eye tracker into a fixation sequence matrix. The behavioral and eye-
388 movement data were further processed and analyzed using the R system for statistical
389 computing (R Development Core Team). Figures were created using MATLAB (Figures 1
390 and 3) or the *ggplot2* package (Wickham, 2016) supplied in R (remaining figures).

391 The T.E.A. was programmed in MATLAB. When using the T.E.A. to prepare the
392 stimulus material for Experiment 1, due to an input error the target was not inserted into an
393 adequate scene location for eight of the scenes. Moreover, the algorithm did not catch that
394 one scene had a different aspect ratio. As a result, nine scenes were excluded when analyzing
395 the data from Experiment 1.

396 Analyses of fixation durations and saccade lengths excluded fixations that were
397 interrupted with blinks. Analysis of fixation durations disregarded fixations that were the first
398 or last fixation in a trial. Fixation durations that are very short or very long are typically
399 discarded, based on the assumption that they are not determined by on-line cognitive
400 processes (Inhoff & Radach, 1998). This precaution was not followed in the present study
401 because the presence of a foveal scotoma may affect eye movements (e.g., fixations were
402 predicted to be longer than normal).

403 Distributions of continuous response variables were positively skewed. In this case,
404 variables are oftentimes transformed to produce model residuals that are more normally
405 distributed. To find a suitable transformation, we estimated the optimal λ -coefficient for the
406 Box-Cox power transformation (Box & Cox, 1964) using the *boxcox* function of the R
407 package *MASS* (Venables & Ripley, 2002) with $y(\lambda) = (y^\lambda - 1)/\lambda$ if $\lambda \neq 0$ and $\log(y)$ if $\lambda = 0$.
408 For all continuous dependent variables, the optimal λ was different from 1, making
409 transformations appropriate. Whenever λ was close to 0, a log transformation was chosen.

410 Non-linear transformations distort the ratio scale properties of the measured variables
411 (Stevens, 1946). As a result, the significance of main effects can change, although this rarely
412 happens (Kliegl, Masson, & Richter, 2010). Perhaps more importantly, some interactions can
413 be transformed away, making them non-interpretable (Loftus, 1978; Wagenmakers,
414 Krypotos, Criss, & Iverson, 2012). Here, we analyzed both untransformed and transformed
415 data. As a default, we report the results for the raw untransformed data and additionally
416 supply the results for the transformed data when they differ from the analysis of the
417 untransformed data.

418 2.7. Statistical analysis using mixed models

419 Continuous response variables were analyzed using linear mixed-effects models
420 (LMM), and search accuracy was analyzed using binomial generalized linear mixed-effects
421 models (GLMM) with a logit link function. The analyses were conducted with the *lme4*
422 package (version 1.1.-23; Bates, Maechler, Bolker, & Walker, 2015) supplied in R, using the
423 bobyqa optimizer for LMMs, and a combination of Nelder-Mead and bobyqa for GLMMs.
424 Separate (G)LMMs were estimated for each dependent variable.

425 A mixed-effects model contains both fixed-effects and random-effects terms. Fixed-
426 effects parameters were estimated via contrast coding for which we used the nomenclature
427 and example code provided by the UCLA Statistical Consulting Group (2011). For the factor
428 scotoma, simple coding was used (-0.5/ +0.5, reference: no scotoma). To test effects of target
429 size, Helmert coding was used to compare each level of the factor target size to the mean of
430 the subsequent levels. The first contrast compared the mean of a given DV for S-targets with
431 the mean for all larger targets (Experiment 1: M-, L-, and XL-targets). For Experiment 1, the
432 second target-size contrast compared the mean for M-targets with the mean across L- and
433 XL-targets, and the third contrast compared the mean for L-targets with the mean for XL-
434 targets. Three additional interaction terms allowed for testing whether the scotoma effect was

435 significantly different for different target-size contrasts. Given that the fixed effects were
436 centered around zero, the intercept of the models reflected the grand mean of the DV.

437 The mixed models included subjects and scene items as crossed random factors. In
438 experimental research, it is common to treat subjects as the sole random factor in the analysis
439 (Judd, Westfall, & Kenny, 2012). However, in research on real-world scene perception and
440 search, the variance introduced by stimulus sampling cannot be ignored (e.g., Nuthmann &
441 Einhäuser, 2015; Nuthmann & Malcolm, 2016). We used counterbalancing to assign scene
442 items to experimental conditions and refrained from placing the search target randomly in the
443 scene. While algorithmic target placement reduces between-scene variability, it does not
444 eliminate it completely. Therefore, scene items were included as random factor.

445 The overall mean for each subject and scene item were estimated as random
446 intercepts. In principle, the variance-covariance matrix of the random effects not only
447 includes random intercepts but also random slopes as well as correlations between intercepts
448 and slopes. Random slopes estimate the degree to which each main effect and/or interaction
449 varies across subjects and/or scene items.

450 To select an optimal random-effects structure for (G)LMMs, we pursued a data-
451 driven approach using backward model selection. To minimize the risk of Type I error, we
452 started with the maximal random-effects structure justified by the design (Barr, Levy,
453 Scheepers, & Tily, 2013). However, the maximal random-effects structure would require
454 estimating 72 parameters (by subject: random intercept, 7 random slopes, 28 correlation
455 terms; by item: same as by subject). Across experiments, none of these maximal models
456 converged (maximal number of iterations: 10^6). To reduce model complexity without taking
457 the risk of inflating the Type I error, we proceeded to fit zero-correlation parameter (zcp)
458 models in which the random slopes are retained but the correlation parameters are set to zero
459 (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017; Seedorff, Oleson, & McMurray,

460 2019). The full random-effects structure of the zcpLMM required 16 variance components to
461 be estimated. This random-effects structure was backwards-reduced using the *step* function
462 of the R package *lmerTest* (version 3.1-2; Kuznetsova, Brockhoff, & Christensen, 2017) to
463 arrive at a model that was justified by the data. For GLMMs we report random intercept
464 models, because random slope models did not converge. Due to the way GLMMs are
465 estimated, model non-convergence tends to be a much larger issue than with LMMs
466 (Seedorff et al., 2019).

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

473 For the (G)LMM, data were not averaged, and modelled at the level of individual
474 observations instead. For the data depicted in Figures 4 and 6, means were calculated for each
475 subject, and these were then averaged across subjects.

476 **3. Results and Discussion**

477 The results of the two letter-in-scene search experiments are presented in three main
478 sections. First, different measures of search accuracy were analyzed as indicators of search
479 efficiency. Second, the time to find the target was analyzed. Behavioral search times were
480 then decomposed based on participants' gaze data to illuminate disruptions in specific sub-
481 processes of search (e.g., Malcolm & Henderson, 2009; Nuthmann, 2014). Third, we
482 examined saccade amplitude and fixation duration across the viewing period as general eye-
483 movement measures.

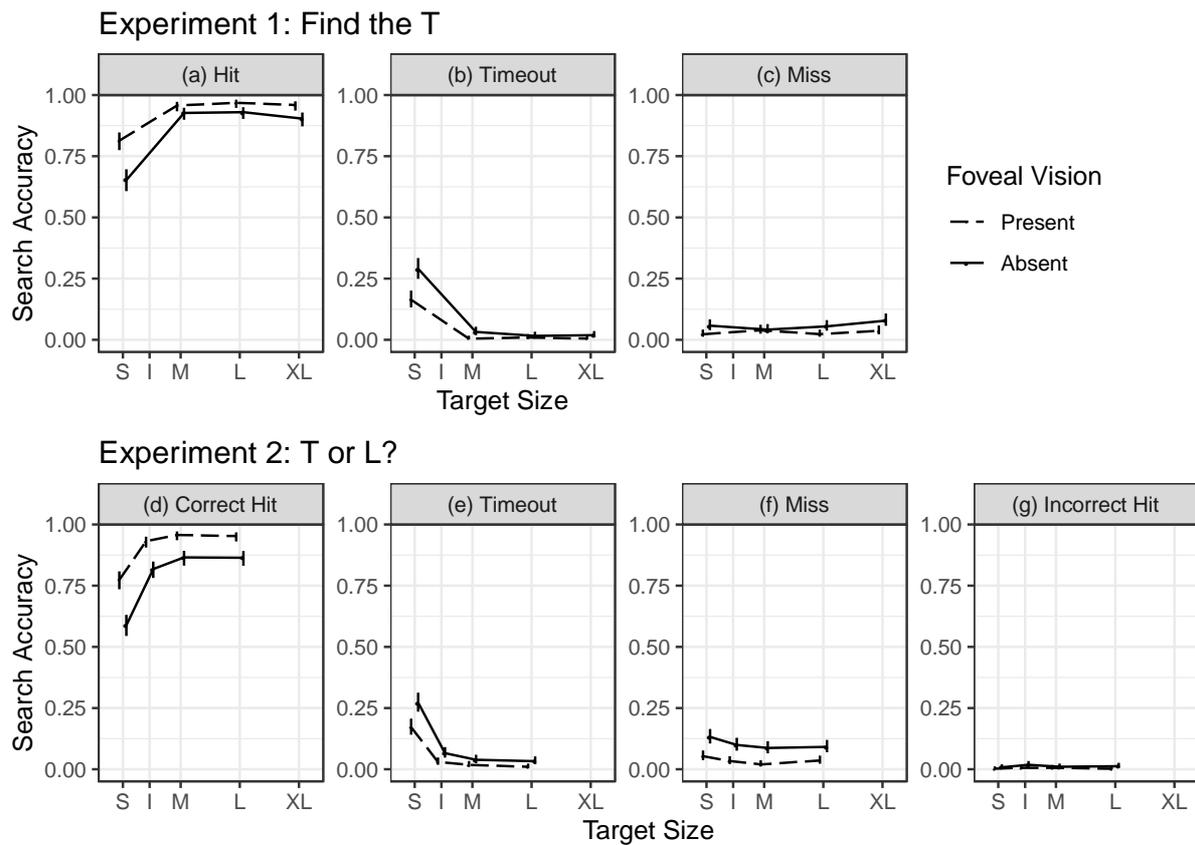
484 *3.1. Search Accuracy*

485 The first set of analyses examined the likelihood of finding the target letter in the
486 scene. Performance for each experimental condition was divided into probabilities of “hit”,
487 “miss”, and “timeout” cases (Nuthmann, 2014). Since we used a target acquisition task, a
488 target was present on all trials. A response was scored as a “hit” if the participant indicated to
489 have located the target by button press and his or her gaze was within the rectangular area of
490 interest (AOI) comprising the target. In signal detection experiments, including yes-no search
491 tasks, trials in which a non-target stimulus is identified as a target are labelled as “false
492 alarms” (Palmer, Verghese, & Pavel, 2000; Tanner & Swets, 1954). In our experiments,
493 incorrect responses included true false alarms where participants were fixating a non-target
494 location and their eyes were not in the vicinity of the target when the button-press response
495 was made. Incorrect responses also included cases where participants fixated near the target
496 but their fixation did not fall within the AOI. Given the difficulty in distinguishing between
497 these two cases, all trials with incorrect responses were labelled as “misses.” The third
498 category comprised trials in which the participant had not responded within 15 s. Trials with
499 no responses were coded as “timeouts.”

500 The size of an AOI that can be given to target stimuli is limited by (a) the spatial
501 (in)accuracy and (im)precision of the eye tracker, and (b) the inaccuracy of the visuo-
502 oculomotor system when targeting relatively small objects (Pajak & Nuthmann, 2013). For
503 high-end eye-trackers like the EyeLink 1000, the minimum AOI size is about 1 to 1.5°, and
504 the recommendation has been made to add a buffer of that size around any target object
505 (Holmqvist & Andersson, 2017). Here, we chose to use the same AOI for all target sizes; this
506 AOI was somewhat larger than the XL target letter with an additional 0.5° of padding to
507 either side (2.9° × 2.9° in total).

508 The search accuracy results for both experiments are depicted in Figure 2. The
509 GLMM results are summarized in Tables 1 and 2 for Experiments 1 and 2, respectively.

510



511

512 *Figure 2.* Measures of search accuracy for Experiment 1 (top row) and Experiment 2 (bottom
 513 row). Each column presents a designated dependent variable, which is specified in the panel
 514 title (see text for definitions). Target sizes on the x -axis are described by letters (S: Small, I:
 515 Intermediate - Experiment 2 only, M: Medium, L: Large, XL: Extra Large - Experiment 1
 516 only; see text for actual sizes in degrees of visual angle). The x -axis is scaled to show all
 517 target sizes across both experiments; the spacing on the x -axis preserves the relative distances
 518 between target sizes. Data points are binomial proportions, error bars are 95% binomial
 519 proportion confidence intervals (Wilson, 1927).

520

521 3.1.1. Experiment 1

522 There was a significant effect of scotoma on the probability of “hitting” the target
 523 such that participants were less likely to correctly locate and accept the target when foveal

524 vision was not available, $b = -0.82$, $SE = 0.14$, $z = -5.88$, $p < .05$ (Figure 2a). Moreover, mean
 525 search accuracy was significantly lower for S-targets compared to the mean of M- through
 526 XL-targets, $b = -1.85$, $SE = 0.12$, $z = -15.26$, $p < .05$; the other target-size contrasts were not
 527 significant (Table 1). Scotoma and target size did not interact (Table 1). The drop in
 528 performance for small targets was due to an increase in timed out trials (Figure 2b). Timeout
 529 probability was low for all other target sizes, with or without a foveal scotoma. The
 530 probability of missing the target was low, with and without a scotoma (Figure 2c).

531 -----
 532 Insert Table 1 about here
 533 -----

534 3.1.2. Experiment 2

535 Experiment 2 included an additional letter recognition component (is the target a “T”
 536 or an “L”?). Therefore, we distinguished between hit trials with correct and incorrect
 537 recognition responses. The probability of incorrect hits was very low in all experimental
 538 conditions (Figure 2g). For correct hit trials, there was a significant effect of scotoma such
 539 that participants were less likely to locate and correctly identify the target without foveal
 540 vision, $b = -1.15$, $SE = 0.11$, $z = -10.36$, $p < .05$ (Figure 2d). Accuracy was lower for smaller
 541 targets; specifically, the contrasts testing S-targets and I-targets against respective larger
 542 targets were significant (S-targets vs. mean for I-, M-, and L-targets: $b = -1.62$, $SE = 0.10$, $z =$
 543 -15.96 , $p < .05$; I-targets vs. mean for M- and L-targets: $b = -0.41$, $SE = 0.14$, $z = -2.98$, $p <$
 544 $.05$). Scotoma and target size did not interact (Table 2). The drop in performance for search
 545 without foveal vision also shows in increased probabilities of missing the target (Figure 2f)
 546 and not responding within 15 s (Figure 2e).

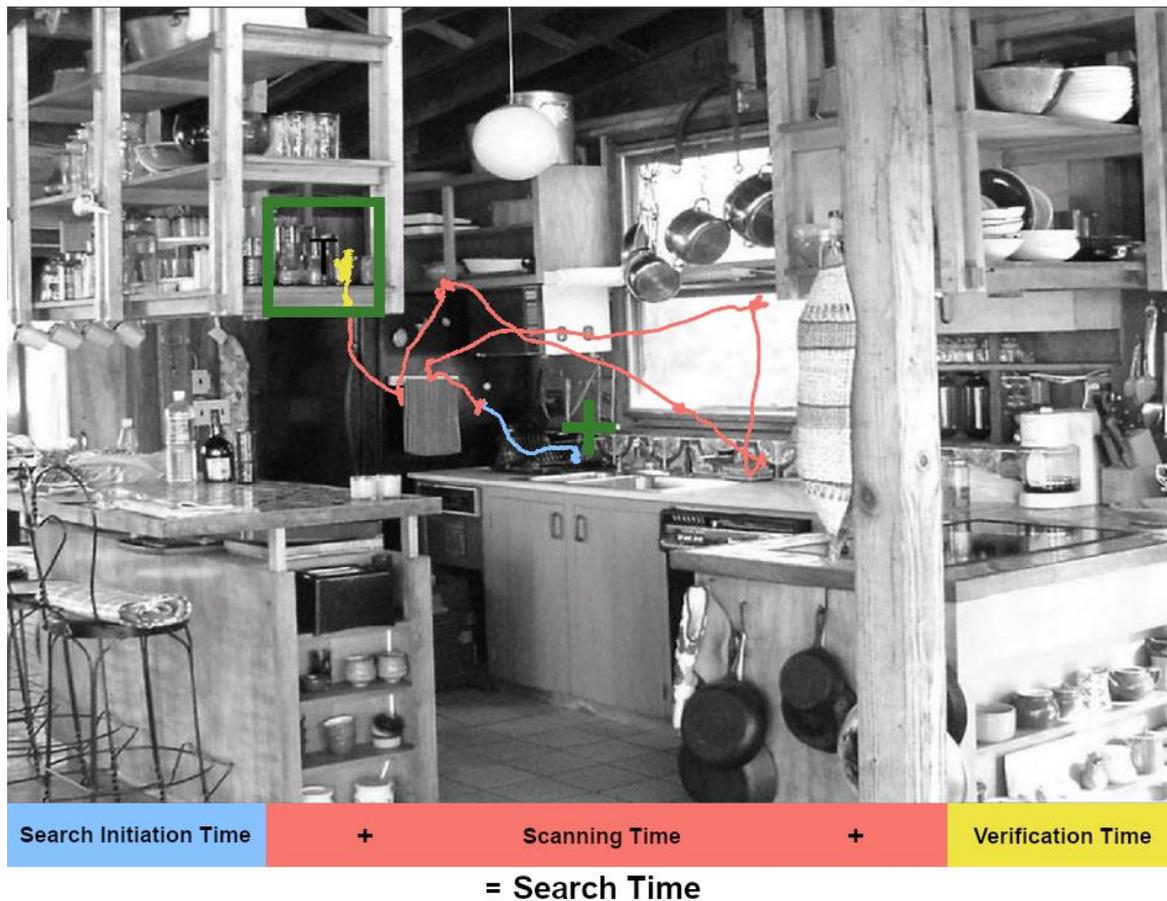
547 -----
 548 Insert Table 2 about here

549

550 *3.2. Search time and its subcomponents*

551 Search behavior was analyzed further for correct trials (“hits”) only. Search time is
552 the overall time taken from scene onset to a user response terminating the search. We then
553 used participants’ gaze data to divide search time into three behaviorally defined epochs:
554 search initiation time, scanning time, and verification time (e.g., Malcolm & Henderson,
555 2009; Nuthmann, 2014; Nuthmann & Malcolm, 2016; Spotorno, Malcolm, & Tatler, 2015).
556 This was done to test how the availability of foveal vision as well as the size of the target
557 would affect different sub-processes of search. Search initiation time is the interval between
558 scene onset and the initiation of the first saccade (i.e., initial saccade latency, or time to
559 move). This epoch measures the time needed to choose a target location for the first saccade.
560 Scanning time (or time to target) is the time from the first eye movement until the
561 participant’s gaze enters the target’s area of interest (minus the first saccade). The scanning
562 time measure reflects the process of localizing the target in space (Malcolm & Henderson,
563 2009), with longer times indicating weaker target guidance. The sum of search initiation time
564 and scanning time represents the latency to first fixate the target (Castelhano, Pollatsek, &
565 Cave, 2008). Our main objective in removing search initiation time from the target latency
566 was to obtain a “clean” measure of scanning time. Finally, the verification process is indexed
567 by the time taken from first entering the target interest area until the participant confirms their
568 decision via button press. This component of search may also include time spent
569 subsequently exploring other scene regions to be sure that they do not contain the target
570 (Castelhano et al., 2008). The segmentation of search time by oculomotor behavior is
571 visualized in Figure 3.

572



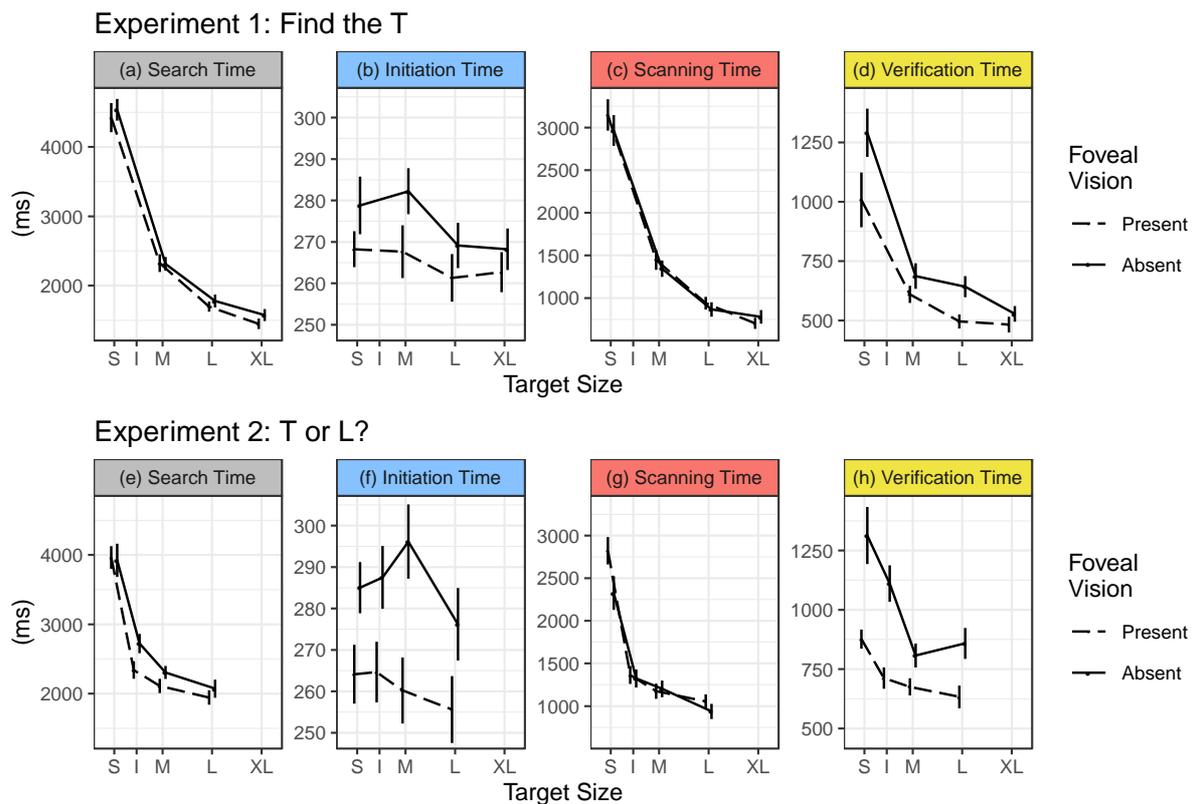
573

574 *Figure 3.* Gaze-based decomposition of search time. For an example search trial, the scene
 575 image is presented together with the raw gaze data from one observer (curvy lines are
 576 saccades, clustered data points are fixations). Visualizing the division of search time, blue
 577 represents search initiation (i.e., initial saccade latency); red, scanning time; and yellow,
 578 verification time. When summed, they yield the total search time. The blue segment includes
 579 saccade execution to visualize the change in gaze position during the first eye movement. The
 580 green box is the interest area around the target letter “T”.

581

582 As outlined in the Introduction, our hypotheses concerned the scanning and
 583 verification time epochs, but not search initiation. Our main objective was to explore the
 584 degree to which scanning times and/or verification times are lengthened when foveal vision is
 585 unavailable. Moreover, we wanted to test whether target size affects the importance of foveal

586 vision to the task; in particular, we hypothesized that verification times may reveal an
 587 interaction between target size and scotoma. A final question was whether any effects on sub-
 588 processes of search—each operating on a different timescale—were large enough to drive
 589 corresponding effects on overall search times. The results are depicted in Figure 4; the LMM
 590 results are summarized in Tables 1 and 2.
 591



592

593 *Figure 4.* Search time and its three epochs for Experiment 1 (top row) and Experiment 2
 594 (bottom row). Each column presents means obtained for a designated dependent variable (see
 595 panel title). For a given dependent variable, the y-axis has been normalized across plots for
 596 ease of comparison between the two experiments; but note the different y-axis scales for the
 597 different measures. For the three sub-processes of search (initiation, scanning, verification),
 598 subplot titles use the color scheme from Figure 3. Solid bold lines represent the scotoma
 599 condition in which foveal vision was absent; dashed lines represent the control condition in
 600 which foveal vision was present. Target sizes on the x-axis are described by letters (S: Small,

601 I: Intermediate - Experiment 2 only, M: Medium, L: Large, XL: Extra Large - Experiment 1
602 only). The x -axis is scaled to show all target sizes across both experiments; the spacing on the
603 x -axis preserves the relative distances between target sizes. Error bars are within-subjects
604 standard errors, using the Cousineau-Morey method (Cousineau, 2005; Morey, 2008).

605

606 3.2.1. Experiment 1

607 The search-time difference between the foveal scotoma and control condition was
608 significant, $b = 170.38$, $SE = 80.14$, $t = 2.13$; for the transformed data, this difference was not
609 significant, $b = 0.0013$, $SE = 0.0008$, $t = 1.62$. Moreover, search times became progressively
610 faster for larger targets, with all three target-size contrasts yielding statistically significant
611 differences (Table 1). Scotoma and target size did not interact (Table 1).

612 For search initiation time, there were no significant effects (Table 1). Importantly,
613 scanning time was not prolonged when searching with a foveal scotoma, $b = -21.93$, $SE =$
614 76.84 , $t = -0.29$. However, scanning times became progressively faster for larger targets, with
615 all three target-size contrasts yielding statistically significant differences (Table 1). Scotoma
616 and target size did not interact (Table 1).

617 Interestingly, verification time was significantly prolonged when searching with a
618 foveal scotoma, $b = 167.81$, $SE = 43.87$, $t = 3.82$. For larger targets, target verification was
619 completed faster. Specifically, the contrasts testing S-targets and M-targets against respective
620 larger targets were significant (Table 1). Moreover, the effect of scotoma was significantly
621 stronger for S-targets compared to the mean effect of scotoma for M- through XL-targets, $b =$
622 285.18 , $SE = 141.38$, $t = 2.02$. For the transformed data, however, this interaction was not
623 significant, $b = 0.092$, $SE = 0.084$, $t = 1.09$.

624 3.2.2. Experiment 2

625 Search times were significantly longer with a foveal scotoma than without, $b =$
626 299.94 , $SE = 87.76$, $t = 3.42$. Moreover, search times were faster for larger targets, with all
627 three target-size contrasts yielding statistically significant differences (Table 2). The effect of
628 scotoma was significantly stronger for I-targets compared to the mean effect for M- and L-
629 targets, $b = 300.99$, $SE = 143.18$, $t = 2.1$. There were no other significant interaction effects
630 for search time (Table 2).

631 In contrast to Experiment 1, the effect of scotoma on search initiation time was
632 statistically significant, $b = 25.6$, $SE = 13$, $t = 1.967$. Moreover, the target-size contrast
633 comparing M-targets with L-targets was significant, $b = 11.46$, $SE = 4.96$, $t = 2.31$. As in
634 Experiment 1, scotoma and target size did not interact (Table 2).

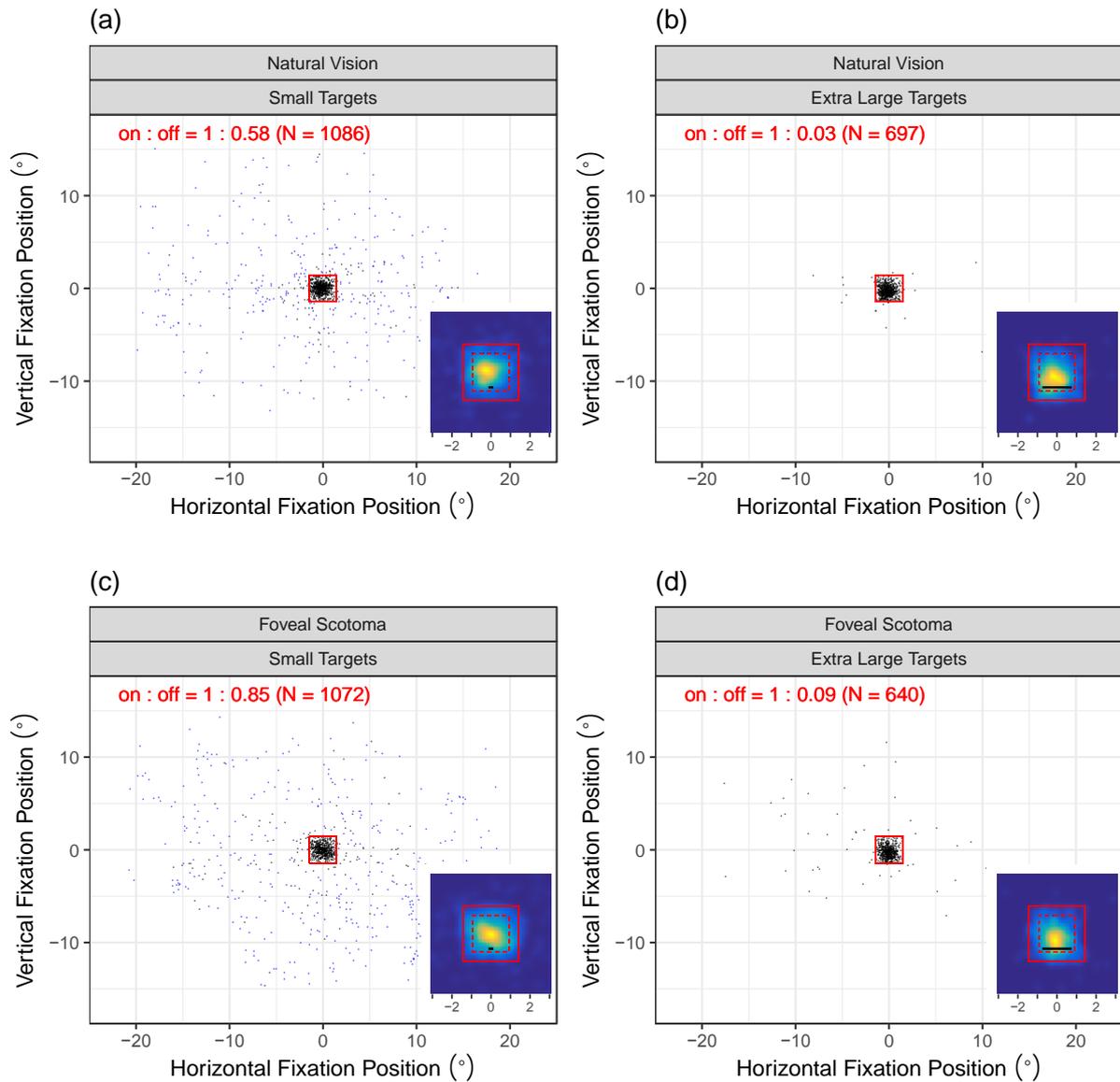
635 For scanning time, there was no significant effect of scotoma, $b = -57.5$, $SE = 55.55$, t
636 $= -1.03$. All three target-size contrasts were significant, the larger the target the faster the
637 search (Table 2). Scotoma and target size did not interact (Table 2).

638 As in Experiment 1, verification time was significantly prolonged when searching
639 with a foveal scotoma, $b = 331.23$, $SE = 60.4$, $t = 5.48$. Verification times were shorter for
640 larger targets; specifically, the contrasts testing S-targets and I-targets against respective
641 larger targets were significant (S-targets vs. mean for I-, M-, and L-targets: $b = 401.5$, $SE =$
642 81.37 , $t = 4.93$; I-targets vs. mean for M- and L-targets: $b = 196.24$, $SE = 52.33$, $t = 3.75$).
643 What about the theoretically salient interaction between scotoma and target size? The first
644 interaction term tested whether the effect of scotoma was significantly different for S-targets
645 compared to the mean effect of scotoma for I- through L-targets; for the untransformed data,
646 the interaction was not significant, $b = 214.12$, $SE = 141.31$, $t = 1.52$, but for the transformed
647 data it was, $b = 0.022$, $SE = 0.011$, $t = 2.09$. The second interaction term compared the effect
648 of scotoma for I-targets to the mean effect of scotoma for M- and L-targets. For the
649 untransformed data, the effect of scotoma was significantly stronger for I-targets compared to

650 the mean effect of scotoma for M- and L-targets, $b = 267.63$, $SE = 96.25$, $t = 2.78$; for the
651 transformed data, however, this interaction was not significant, $b = 0.018$, $SE = 0.010$, $t =$
652 1.77. The third interaction, comparing the effect of scotoma for M-targets to the effect of
653 scotoma for L-targets, was not significant (Table 2).

654 3.2.3. *Where are the eyes during the verification epoch?*

655 Two more questions arise regarding the last component of search. Why are
656 verification times longer for smaller targets? And what are the eyes doing when foveal
657 analysis of the search target is not possible during fixation? In the scotoma conditions of our
658 previous study (Nuthmann, 2014), observers had no problem selecting the target in
659 parafoveal vision and fixating their gaze on it. Within-object fixation positions showed a
660 central *Preferred Viewing Location* (PVL) such that most initial fixations were placed in
661 proximity to object center (Pajak & Nuthmann, 2013). Moreover, prolonged verification
662 times in the central-scotoma condition were due to an increased number of off-target
663 fixations to unmask the object and to further analyze it in peripheral vision. Here, we used
664 one common AOI for all target sizes; thus, the margin around the actual target was larger for
665 smaller targets. Therefore, differences in oculomotor behavior for the different target-size
666 conditions are not well captured by a binary distinction between on-target and off-target
667 fixations. Instead, we explored fixation positions during the verification epoch through two-
668 dimensional scatter and density plots. Since the AOI was used for data scoring, we still refer
669 to fixations within the AOI as on-target fixations and fixations outside the AOI as off-target
670 fixations. We summarize important aspects of a complex data pattern by comparing extreme
671 target sizes, that is S-targets and XL-targets from Experiment 1 (Figure 5).



672

673 *Figure 5.* Analysis of fixation positions during the verification epoch. The four panels show
 674 data for small (left) and extra-large (right) targets in the natural-vision (top) and foveal-
 675 scotoma (bottom) conditions from Experiment 1. The red square with solid lines represents
 676 the area of interest (AOI) used for distinguishing between on-target fixations (within the
 677 AOI) and off-target fixations (outside the AOI). Fixation positions are expressed relative to
 678 the center of the AOI. The scatter plots show all fixations made during the verification epoch.
 679 Fixations belonging to sequences with five or more successive off-target fixations are
 680 depicted in blue rather than black. Also presented is the ratio of on-target to off-target

681 fixations, along with the absolute number of fixations. The inset plots zoom into the AOI
682 region ($6^\circ \times 6^\circ$) and display fixation positions as two-dimensional density plots. The
683 frequency information is displayed as variations in color, with colors ranging from blue (few
684 fixations), through the parula colormap to yellow (many fixations). The red square with
685 dashed lines is the AOI encompassing the extra-large letter. The bold black line depicts the
686 actual width of the target letter.

687

688 To unmask the target, the best strategy would be to move the eyes outside the target
689 AOI. However, the scatter plots for XL-targets show very few off-target fixations, both with
690 a foveal scotoma (Figure 5d) and without (Figure 5b). To overcome overplotting for on-target
691 fixations, the inset plots zoom into the region where the AOI was situated and display density
692 heatmaps of fixations. The fixation positions within the AOI comprise initial fixations,
693 immediate refixations, and later revisits. The data for XL-targets show a central “hot spot”,
694 replicating the finding of a PVL, whether foveal vision was available or not. Collectively, the
695 data suggest that extrafoveal information from the last scanning fixation was oftentimes
696 sufficient to identify extra-large targets when foveal vision was not available.

697 For the smaller target sizes, a different pattern of results emerged. There were still
698 many more on-target than off-target fixations, but off-target fixations were much more
699 frequent than for XL-targets. For S-targets (Figure 5a and c), off-target fixations were widely
700 spread around the target AOI. There were also more off-target fixations with a foveal
701 scotoma than without, as reflected by the on : off ratios. In the foveal-scotoma condition,
702 some fixations close to the AOI may have been placed there purposely to unmask the letter
703 and to process it in extrafoveal vision. More generally, off-target fixations are thought to be
704 double-checking fixations to ensure that other scene regions did not contain the target (cf.
705 Castelhana et al., 2008). It is also possible that observers did not actually attend to the target

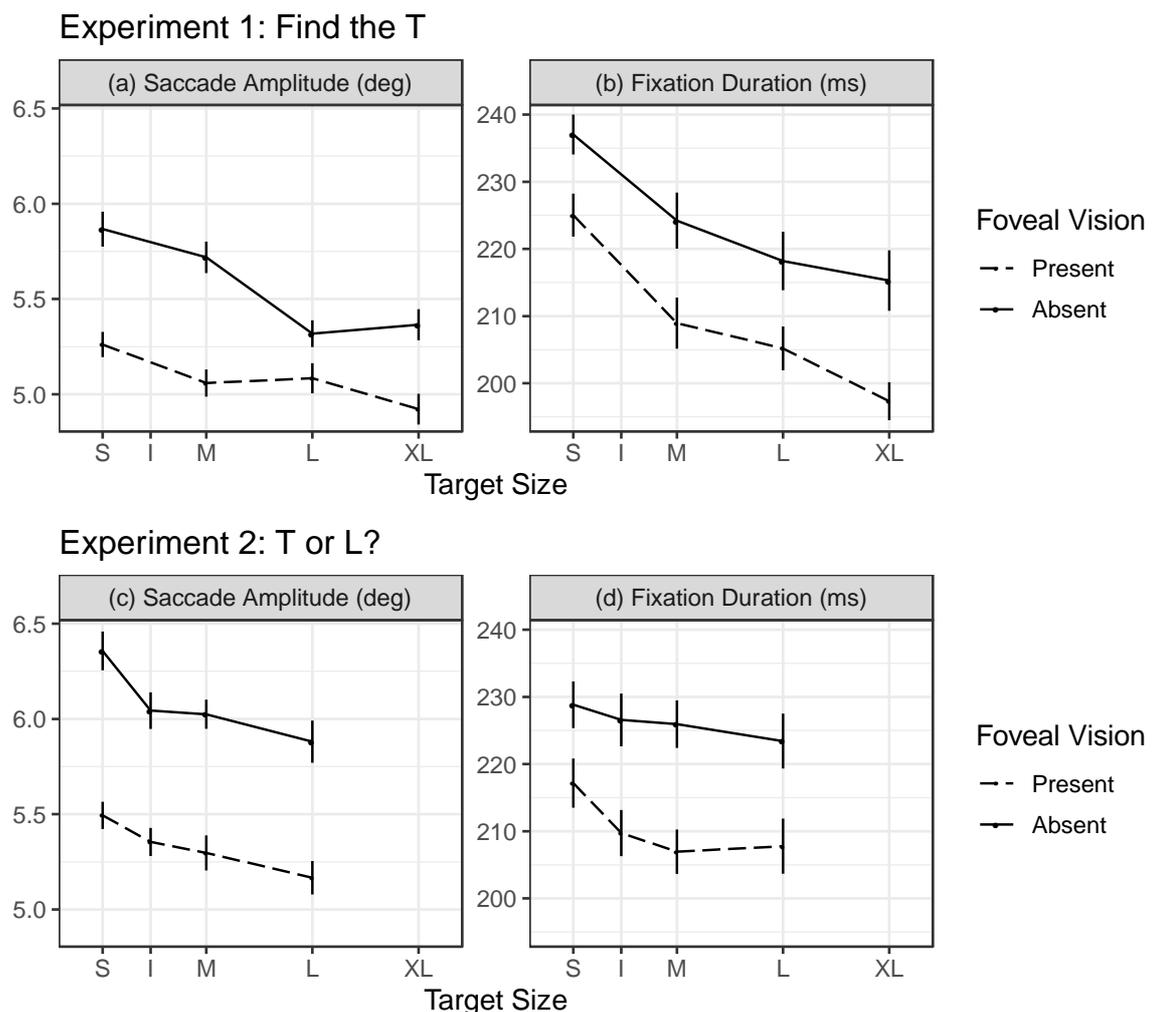
706 when they first encountered it and therefore kept exploring other scene regions. We cannot
707 reliably distinguish between these alternatives. In any case, fixations far away from the AOI
708 tended to come from trials in which longer sequences of successive off-target fixations were
709 made before the eyes returned to the target. To highlight this, in the scatter plots all fixations
710 that come from sequences with five or more successive off-target fixations are presented in
711 blue rather than black (the number 5 was arbitrarily chosen). Fixation positions within the
712 AOI showed a central PVL both in the presence and absence of foveal vision (inset plots in
713 Figure 5a and c).

714 *3.3. Saccade amplitudes and fixation durations*

715 Saccade amplitudes and fixation durations were analyzed to characterize eye-
716 movement behavior during visual search (Figure 6). In the presence of a simulated scotoma,
717 we should observe somewhat larger saccade amplitudes and longer fixation durations
718 (Bertera & Rayner, 2000; F. W. Cornelissen, Bruin, & Kooijman, 2005; Miellet et al., 2010;
719 Nuthmann, 2014). We had no a priori hypotheses regarding the relationship between target
720 size and saccade amplitudes and/or fixation durations.

721 For both experiments, results for mean saccade amplitudes showed a significant effect
722 of scotoma, with larger saccades when searching with a foveal scotoma than without
723 (Experiment 1: $b = 0.49$, $SE = 0.07$, $t = 6.74$, Figure 6a; Experiment 2: $b = 0.74$, $SE = 0.1$, $t =$
724 7.14 , Figure 6c). In both experiments, an increase in target size was associated with shorter
725 saccade amplitudes (Experiment 1: Table 1, Experiment 2: Table 2). For Experiment 1, the
726 two contrasts testing S-targets and M-targets against respective larger targets were
727 significant. For Experiment 2, the effect of target size on saccade amplitudes was driven by
728 S-targets only. For M-Targets in Experiment 1, the effect of scotoma was significantly
729 stronger than the mean effect of scotoma for any larger targets (Table 1). In Experiment 2,
730 scotoma and target size did not interact (Table 2).

731 Fixation durations also showed a significant effect of scotoma, with longer fixation
 732 durations when searching with a foveal scotoma than without (Experiment 1: $b = 16.57$, $SE =$
 733 3.83 , $t = 4.33$, Figure 6b; Experiment 2: $b = 18.12$, $SE = 4.73$, $t = 3.83$, Figure 6d). Moreover,
 734 fixation durations tended to be shorter for larger targets (Tables 1 and 2). For Experiment 1,
 735 the contrasts testing S-targets and M-targets against respective larger targets were significant
 736 (S-targets vs. mean for M-, L-, and XL-targets: $b = 20.48$, $SE = 2.59$, $t = 7.91$; M-targets vs.
 737 mean for L- and XL-targets: $b = 6.25$, $SE = 2.9$, $t = 2.15$). For Experiment 2, mean fixation
 738 duration was significantly increased for S-Targets compared to the mean for I- through L-
 739 targets, $b = 9.54$, $SE = 2.46$, $t = 3.88$. Scotoma and target size did not interact (Tables 1 and
 740 2).
 741



743 *Figure 6.* Mean saccade amplitudes and fixation durations for both experiments. Solid bold
744 lines represent the scotoma condition in which foveal vision was absent; dashed lines
745 represent the normal-vision control condition. Target sizes on the x -axis are described by
746 letters (S: Small, I: Intermediate - Experiment 2 only, M: Medium, L: Large, XL: Extra Large
747 - Experiment 1 only). The x -axis is scaled to show all target sizes across both experiments;
748 the spacing on the x -axis preserves the relative distances between target sizes. Error bars are
749 within-subjects standard errors, using the Cousineau-Morey method (Cousineau, 2005;
750 Morey, 2008).

751

752 **4. General Discussion**

753 Two experiments were conducted to test the degree to which foveal vision was
754 necessary to find context-free target letters in naturalistic scenes. A gaze-contingent moving
755 mask (Rayner & Bertera, 1979) was used to simulate the absence of foveal vision. In
756 Experiment 1, observers searched for the letter “T” which could occur at four different sizes.
757 In Experiment 2, the target was either a “T” or an “L”, and participants had to indicate which
758 letter it was. If foveal vision was necessary to achieve normal search performance, the time
759 taken to find the target should be significantly longer without foveal vision than with.
760 Moreover, we reasoned that the importance of foveal vision may depend on the size of the
761 search target such that foveal vision loss may be more detrimental for smaller targets.

762 While searching for the target without foveal vision, observers were significantly less
763 likely to find the target than with normal vision. Our main analyses considered all correct
764 trials (“hits”), for which we analyzed search times along with three sub-processes of search
765 (cf. Nuthmann, 2014). With a foveal scotoma, search initiation times were significantly
766 prolonged in Experiment 2, but not in Experiment 1. Thus, when foveal vision is not
767 available it may take a little longer to launch the very first saccade, but this is not always the

768 case. In both experiments, without foveal vision participants were not impaired in locating
769 the search target in the scene (indexed by scanning time), but the process of accepting the
770 target and responding was delayed (indexed by verification time).

771 Button-press search times are the sum of search initiation, scanning, and verification
772 times. Average verification times are typically shorter than scanning times, and initiation
773 times are shorter still. The question then arises whether small effects on faster sub-processes
774 of search are large enough to affect total search time. For Experiment 2, search times were
775 significantly prolonged when searching with a foveal scotoma. For Experiment 1, the effect
776 of scotoma was significant for the untransformed data (Table 1) but not for the transformed
777 data. Moreover, when analyzing the search-time data from a given experiment with (less
778 appropriate) two-way repeated measures analyses of variance (*F*₁ test with subject as random
779 effect), no significant effect of scotoma was detected. In summary, the effect of a foveal
780 scotoma on search times was fairly small and not very stable (Experiment 1).

781 The experiments also tested whether target size was a mediating factor for previous
782 findings on the (un)importance of foveal vision during scene search (McIlreavy et al., 2012;
783 Nuthmann, 2014). Not surprisingly, the data from both experiments were indicative of better
784 search performance for larger targets, in keeping with previous research (Miellet et al., 2010).
785 Searching the scenes for small letters proved to be a difficult task, with timed out trials and
786 fairly long search times on successful trials. Button-press search times for medium-sized
787 letters were similar to the ones for contextually relevant objects in our previous study
788 (Nuthmann, 2014). The critical question was whether the size of the search target would
789 affect the importance of foveal vision to the task (cf. Miellet et al., 2010). Specifically, we
790 hypothesized that any detrimental effect of the foveal scotoma on the target verification
791 process may only occur for smaller targets, or may be more pronounced for smaller than for
792 larger targets. Significant interactions between scotoma and target size would lend support to

793 this hypothesis. For both experiments, we found that the presence of significant interaction
794 terms was scale dependent. In Experiment 1, the scotoma \times target size 1 interaction was only
795 significant for the untransformed data. In Experiment 2, the scotoma \times target size 1
796 interaction was only significant for the transformed data, whereas the scotoma \times target size 2
797 interaction was only significant for the untransformed data. Given the discrepant results for
798 untransformed and transformed data, we do not place much interpretative weight on the
799 interaction effects (Brysbaert & Stevens, 2018; Loftus, 1978). It is clear that any effects are
800 small, suggesting the value of a replication study to support these conclusions.

801 Recent results regarding the unimportance of foveal vision when searching for spatial
802 distortions (McIlreavy et al., 2012) or real-world objects (Nuthmann, 2014) in naturalistic
803 scenes were surprising, given the importance of foveal vision in both reading (Rayner &
804 Bertera, 1979) and visual search within alphanumeric displays (Bertera & Rayner, 2000). To
805 better understand these task differences, we combined design features from letter search and
806 scene search tasks by embedding letters into images of real-world scenes. In the following,
807 we discuss the present results in the context of existing literature.

808 Foveal vision appeared to be more important in the present letter-in-scene search
809 tasks than during object-in-scene search (Nuthmann, 2014). Neither search accuracy, nor
810 search time or any of its components were affected by a simulated foveal scotoma in
811 Nuthmann (2014). In contrast, search accuracy was significantly lower, and target
812 verification time significantly prolonged in the present experiments, in which the target was a
813 context-free letter rather than a contextually relevant object. We note that the objects used in
814 Nuthmann (2014) were, on average, larger in size than the largest letters used here. In the
815 present experiments, the simulated scotoma completely masked the target when observers
816 directed their gaze to the geometrical center of the letter target, regardless of its size. Thus,
817 the foveal scotoma could occlude the entire letter. The fact that target verification was still

818 possible demonstrates that it could be done on the basis of extrafoveal information alone. For
819 one, there was extrafoveal information about the target from the last scanning fixation.
820 Moreover, during the subsequent verification epoch—and for all but the XL-targets—
821 observers had an increased tendency to make additional off-target fixations, which increased
822 verification time.

823 Foveal vision appeared to be less important in the present letter-in-scene search tasks
824 compared to letter search in alphanumeric displays for which quite dramatic search-time
825 costs were observed (Bertera & Rayner, 2000). In the experiment by Bertera and Rayner
826 (2000), each array consisted of 26 letters (with 4 letters repeated) and 9 digits. Even though
827 alphanumeric characters are overlearned stimuli, searching such displays for a designated
828 target letter is bound to be relatively inefficient because the distractor items consisted of a
829 large and heterogeneous set of other letters, as well as numbers. In this case, the extraction of
830 fine detail via foveal analysis was found to be beneficial to the task (Bertera & Rayner,
831 2000). Compared to such unguided letter search, the availability of foveal vision may be less
832 important for search displays in which the distractor letters are relatively similar to each other
833 and relatively different from the target letter. For covert search³, it has been shown that
834 search efficiency increases as distractor-distractor similarity increases and target-distractor
835 similarity decreases (Duncan & Humphreys, 1989). During overt search, distractors which
836 are similar to the target receive more fixations than dissimilar distractors (Reingold &
837 Glaholt, 2014, for review). The fact that such saccadic selectivity exists indicates that
838 extrafoveal processing and top-down factors influence the decision about where to look next.
839 Moreover, extrafoveal processing during the scanning epoch may also facilitate later target

³ In this research, displays are smaller than the observer's visual span such that eye movements are not essential (Findlay & Gilchrist, 2003). At the same time, unless the target was located in foveal vision, search success implies that the target has been discriminated outside foveal vision.

840 verification. Thus, it is an open question for future research to determine whether a stronger
841 reliance on guidance mechanisms may render foveal vision less important.

842 When search takes place in real-world scenes, basic feature guidance by object
843 properties is complemented by different types of scene guidance, in particular syntactic,
844 semantic, and episodic guidance (Henderson & Ferreira, 2004, for review). Studying visual
845 search in scenes poses some methodological challenges. It is unclear what to count as an
846 “object” in a real-world scene (Neider & Zelinsky, 2008). Thus, there is no clear separation
847 between targets and distractors. Distractor features tend to be heterogeneous (Wolfe et al.,
848 2011) and the degree to which visual similarity relationships between objects in scenes affect
849 guidance of gaze to search targets is hard to assess (Alexander & Zelinsky, 2012). Here, we
850 used simple targets that were precisely specified (but varied in size) and attempted to control
851 for their local salience. Our naturalistic scenes contained exactly one target letter to be
852 analyzed against the scene background (Experiment 1: T, Experiment 2: T or L). The scenes
853 in which the letter targets were embedded showed natural variation in (a) the number of
854 elements that shared some similarity with the target (e.g., a chair leg), (b) overall target-
855 background similarity (De Vries, Hooge, Wertheim, & Verstraten, 2013, for review), and (c)
856 scene clutter (Rosenholtz, Li, & Nakano, 2007).

857 According to contemporary search theories like the target acquisition model
858 (Zelinsky, 2008), observers compare their target representation to the search scene to obtain a
859 map of evidence for the target at each image location. This map is then used to guide eye
860 movements to target-like patterns in the scene (peripheral selection task). Upon fixation,
861 incoming visual information is analyzed to decide whether this pattern is a target or a
862 distractor (central discrimination task). The cycle of selection (guidance) and discrimination
863 repeats until the target is found (Reingold & Glaholt, 2014; Zelinsky, Peng, Berg, & Samaras,
864 2013). By simulating a foveal scotoma, we selectively masked information that would

865 otherwise be used for the central discrimination task. As a result, individual fixation durations
866 were increased, a common finding. The foveal scotoma did not slow down the process of
867 target localization, as measured by scanning time. This particular result highlights the
868 importance of extrafoveal vision for target localization. It also implies that the reject decision
869 during scanning epoch fixations was not substantially impaired, the logical conclusion being
870 that the resolution of extrafoveal vision was sufficient to make that decision. However,
871 making foveal vision unavailable increased the difficulty of the verification task. The accept
872 decision during verification epoch fixations is thought to require a more complete analysis of
873 the target candidate than the reject decision during scanning fixations (Malcolm &
874 Henderson, 2009). In agreement with this view, in the scotoma condition observers spent
875 more time making the accept decision, during on-target and off-target fixations.

876 As outlined above, there were various reasons for using letter targets.
877 Methodologically, this design choice ensured that the effects of interest could not be
878 mediated by other variables such as contextual constraints, target salience, or eccentricity.
879 Importantly, when searching for a context-free letter target the scene is more than just a
880 patterned background. Processing of scene and object relationships appears to be obligatory,
881 in a sense that it is hard to suppress (T. H. W. Cornelissen & Vö, 2017). To extend the
882 present findings, it would be useful to systematically explore the role played by various forms
883 of scene guidance, using manipulations like scene inversion (Foulsham & Underwood, 2011),
884 scene scrambling (Foulsham, Alan, & Kingstone, 2011), or pseudo-scene viewing (Luke &
885 Henderson, 2016).

886 The present results replicate the finding that fixation durations and saccade
887 amplitudes are both elevated in the presence of an artificial scotoma (Bertera & Rayner,
888 2000; F. W. Cornelissen et al., 2005; Mielliet et al., 2010; Nuthmann, 2014; but see McIlreavy
889 et al., 2012). The saccade amplitude adjustment reflects a tendency to fixate more locations in

890 the non-degraded scene area than the degraded area (Nuthmann, 2014). Both global eye-
891 movement parameters were also affected by target size; a reduction in target size was
892 associated with both larger saccade amplitudes (see also Mielliet et al., 2010) as well as longer
893 fixation durations. These findings were unexpected, because participants had no way of
894 knowing which target size would be displayed next, due to the randomized presentation of
895 scenes. Over the course of scene viewing, there is a tendency for fixation durations to
896 increase and saccade amplitudes to decrease (Pannasch, Helmert, Roth, Herbold, & Walter,
897 2008; Unema, Pannasch, Joos, & Velichkovsky, 2005). In our experiments, search time
898 equates to viewing time, such that the longer search times for small targets could potentially
899 explain the longer fixation durations (but not the larger saccade amplitudes). However, time
900 course analyses (not reported here) provided no evidence for this. Without further research,
901 any account of why or how observers adjust their fixation durations and saccade amplitudes
902 in response to different target sizes (in otherwise identical scenes) remains speculative. A
903 first step toward explaining this counterintuitive finding is to directly compare randomized
904 and blocked presentations of different target sizes (cf. Rothkegel, Schütt, Trukenbrod,
905 Wichmann, & Engbert, 2019).

906 Theories of visual search have largely been built on search for targets in arbitrary 2D
907 arrays of items which observers searched without moving their eyes (Wolfe & Horowitz,
908 2017, for review). However, most real-world search takes place in structured scenes which
909 observers explore through eye movements. The adoption of more ecologically valid stimuli
910 has led to a new brand of image-based search theory (Eckstein, 2011, for review). Most of
911 these models ignore that visual acuity declines systematically from the central fovea into the
912 periphery (Nuthmann, 2014, for discussion). Moreover, visual search models usually aim at
913 explaining the nature of peripheral selection (guidance) rather than central discrimination
914 (Zelinsky et al., 2013). We analyzed both components and found that extrafoveal processing

915 is not only important for selection but also for discrimination (cf. Reingold & Glaholt, 2014).

916 The present results, together with our previous findings, inform future model building by

917 specifying how (un)important the different regions of the visual field are for different sub-

918 processes of search.

919

920 **References**

- 921 Alexander, R. G., & Zelinsky, G. J. (2012). Effects of part-based similarity on visual search:
922 The Frankenbear experiment. *Vision Research*, *54*, 20–30.
923 <https://doi.org/10.1016/j.visres.2011.12.004>
- 924 Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed
925 random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–
926 412. <https://doi.org/10.1016/j.jml.2007.12.005>
- 927 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for
928 confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*,
929 *68*(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- 930 Bates, D., Maechler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects
931 models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.
932 <https://doi.org/10.18637/jss.v067.i01>
- 933 Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of the effective stimulus in
934 visual search. *Perception & Psychophysics*, *62*(3), 576–585.
935 <https://doi.org/10.3758/BF03212109>
- 936 Bex, P. J. (2010). (In) Sensitivity to spatial distortion in natural scenes. *Journal of Vision*,
937 *10*(2):23, 1–15. <https://doi.org/10.1167/10.2.23>
- 938 Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural
939 images. *Journal of the Optical Society of America A-Optics Image Science and Vision*,
940 *19*(6), 1096–1106. <https://doi.org/10.1364/JOSAA.19.001096>
- 941 Borji, A., Sihite, D. N., & Itti, L. (2013). Quantitative analysis of human-model agreement in
942 visual saliency modeling: a comparative study. *IEEE Transactions on Image*
943 *Processing*, *22*(1), 55–69. <https://doi.org/10.1109/TIP.2012.2210727>
- 944 Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal*

- 945 *Statistical Society Series B-Statistical Methodology*, 26(2), 211–252.
946 <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- 947 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.
948 <https://doi.org/10.1163/156856897X00357>
- 949 Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models:
950 a tutorial. *Journal of Cognition*, 1(1), 9. <https://doi.org/10.5334/joc.10>
- 951 Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target
952 eccentricity affects performance on conjunction searches. *Perception & Psychophysics*,
953 57(8), 1241–1261. <https://doi.org/10.3758/BF03208380>
- 954 Carrasco, M., & Frieder, K. S. (1997). Cortical magnification neutralizes the eccentricity
955 effect in visual search. *Vision Research*, 37(1), 63–82. [https://doi.org/10.1016/S0042-](https://doi.org/10.1016/S0042-6989(96)00102-2)
956 6989(96)00102-2
- 957 Carrasco, M., McLean, T. L., Katz, S. M., & Frieder, K. S. (1998). Feature asymmetries in
958 visual search: Effects of display duration, target eccentricity, orientation and spatial
959 frequency. *Vision Research*, 38(3), 347–374. [https://doi.org/10.1016/S0042-](https://doi.org/10.1016/S0042-6989(97)00152-1)
960 6989(97)00152-1
- 961 Castelhana, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an
962 unspecified target, but only in identification and not in attentional guidance. *Psychonomic*
963 *Bulletin & Review*, 15(4), 795–801. <https://doi.org/10.3758/PBR.15.4.795>
- 964 Cornelissen, F. W., Bruin, K. J., & Kooijman, A. C. (2005). The influence of artificial
965 scotomas on eye movements during visual search. *Optometry and Vision Science*, 82(1),
966 27–35.
- 967 Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The EyeLink Toolbox: Eye tracking
968 with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods*,
969 *Instruments, & Computers*, 34(4), 613–617. <https://doi.org/10.3758/BF03195489>

- 970 Cornelissen, T. H. W., & Võ, M. L.-H. (2017). Stuck on semantics: Processing of irrelevant
971 object-scene inconsistencies modulates ongoing gaze behavior. *Attention Perception &*
972 *Psychophysics*, 79(1), 154–168. <https://doi.org/10.3758/s13414-016-1203-7>
- 973 Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to
974 Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, 1(1),
975 42–45. <https://doi.org/10.20982/tqmp.01.1.p042>
- 976 De Vries, J. P., Hooge, I. T. C., Wertheim, A. H., & Verstraten, F. A. J. (2013). Background,
977 an important factor in visual search. *Vision Research*, 86, 128–138.
978 <https://doi.org/10.1016/j.visres.2013.04.010>
- 979 Duchowski, A. T., & Çöltekin, A. (2007). Foveated gaze-contingent displays for peripheral
980 LOD management, 3D visualization, and stereo Imaging. *ACM Transactions on*
981 *Multimedia Computing Communications and Applications*, 3(4):24, 1–18.
982 <https://doi.org/10.1145/1314303.1314309>
- 983 Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity.
984 *Psychological Review*, 96(3), 433–458. <https://doi.org/10.1037//0033-295X.96.3.433>
- 985 Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, 11(5):14, 1–36.
986 <https://doi.org/10.1167/11.5.14>
- 987 Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*.
988 Oxford: University Press.
- 989 Foulsham, T., Alan, R., & Kingstone, A. (2011). Scrambled eyes? Disrupting scene structure
990 impedes focal processing and increases bottom-up guidance. *Attention Perception &*
991 *Psychophysics*, 73(7), 2008–2025. <https://doi.org/10.3758/s13414-011-0158-y>
- 992 Foulsham, T., & Underwood, G. (2011). If visual saliency predicts search, then why?
993 Evidence from normal and gaze-contingent search tasks in natural scenes. *Cognitive*
994 *Computation*, 3(1), 48–63. <https://doi.org/10.1007/s12559-010-9069-9>

- 995 Geisler, W. S., & Chou, K. L. (1995). Separation of low-level and high-level factors in
996 complex tasks: Visual search. *Psychological Review*, *102*(2), 356–378.
997 <https://doi.org/10.1037/0033-295X.102.2.356>
- 998 Geringswald, F., Baumgartner, F., & Pollmann, S. (2012). Simulated loss of foveal vision
999 eliminates visual search advantage in repeated displays. *Frontiers in Human*
1000 *Neuroscience*, *6*. <https://doi.org/10.3389/fnhum.2012.00134>
- 1001 Geringswald, F., & Pollmann, S. (2015). Central and peripheral vision loss differentially
1002 affects contextual cueing in visual search. *Journal of Experimental Psychology:*
1003 *Learning Memory and Cognition*, *41*(5), 1485–1496.
1004 <https://doi.org/10.1037/xlm0000117>
- 1005 Glaholt, M. G., Rayner, K., & Reingold, E. M. (2012). The mask-onset delay paradigm and
1006 the availability of central and peripheral visual information during scene viewing.
1007 *Journal of Vision*, *12*(1):9, 1–19. <https://doi.org/10.1167/12.1.9>
- 1008 Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M.
1009 Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye*
1010 *movements and the visual world* (pp. 1–58). New York: Psychology Press.
- 1011 Holmqvist, K., & Andersson, R. (2017). Eye tracking: A comprehensive guide to methods,
1012 paradigms and measures. Lund, Sweden: Lund Eye-Tracking Research Institute.
- 1013 Hooge, I. T. C., & Erkelens, C. J. (1999). Peripheral vision and oculomotor control during
1014 visual search. *Vision Research*, *39*(8), 1567–1575. [https://doi.org/10.1016/S0042-](https://doi.org/10.1016/S0042-6989(98)00213-2)
1015 [6989\(98\)00213-2](https://doi.org/10.1016/S0042-6989(98)00213-2)
- 1016 Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in
1017 the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and*
1018 *scene perception* (pp. 29–53). Oxford: Elsevier Science Ltd.
1019 <https://doi.org/10.1016/B978-008043361-5/50003-1>

- 1020 Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of
1021 visual attention. *Vision Research*, *40*(10–12), 1489–1506.
1022 [https://doi.org/10.1016/S0042-6989\(99\)00163-7](https://doi.org/10.1016/S0042-6989(99)00163-7)
- 1023 Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in
1024 social psychology: A new and comprehensive solution to a pervasive but largely ignored
1025 problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69.
1026 <https://doi.org/10.1037/a0028347>
- 1027 Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*,
1028 14.
- 1029 Kliegl, R., Masson, M. E. J., & Richter, E. M. (2010). A linear mixed model analysis of
1030 masked repetition priming. *Visual Cognition*, *18*(5), 655–681.
1031 <https://doi.org/10.1080/13506280902986058>
- 1032 Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in
1033 linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.
1034 <https://doi.org/10.18637/jss.v082.i13>
- 1035 Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, *6*(3), 312–319.
1036 <https://doi.org/10.3758/BF03197461>
- 1037 Luke, S. G., & Henderson, J. M. (2016). The influence of content meaningfulness on eye
1038 movements across tasks: Evidence from scene viewing and reading. *Frontiers in*
1039 *Psychology*, *7*. <https://doi.org/10.3389/fpsyg.2015.00257>
- 1040 Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on
1041 visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*,
1042 *9*(11):8, 1–13. <https://doi.org/10.1167/9.11.8>
- 1043 Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I
1044 error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–

- 1045 315. <https://doi.org/10.1016/j.jml.2017.01.001>
- 1046 McIlreavy, L., Fiser, J., & Bex, P. J. (2012). Impact of simulated central scotomas on visual
1047 search in natural scenes. *Optometry and Vision Science*, *89*(9), 1385–1394.
1048 <https://doi.org/10.1097/OPX.0b013e318267a914>
- 1049 Miellel, S., Zhou, X., He, L., Rodger, H., & Caldara, R. (2010). Investigating cultural
1050 diversity for extrafoveal information use in visual scenes. *Journal of Vision*, *10*(6):21,
1051 1–18. <https://doi.org/10.1167/10.6.21>
- 1052 Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau
1053 (2005). *Tutorial in Quantitative Methods for Psychology*, *4*, 61–64.
- 1054 Moulden, B., Kingdom, F., & Gatley, L. F. (1990). The standard deviation of luminance as a
1055 metric for contrast in random-dot images. *Perception*, *19*(1), 79–101.
1056 <https://doi.org/10.1068/p190079>
- 1057 Neider, M. B., & Zelinsky, G. J. (2008). Exploring set size effects in scenes: Identifying the
1058 objects of search. *Visual Cognition*, *16*(1), 1–10.
1059 <https://doi.org/10.1080/13506280701381691>
- 1060 Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual*
1061 *Cognition*, *21*(7), 803–837. <https://doi.org/10.1080/13506285.2013.832449>
- 1062 Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in
1063 real-world scenes? Evidence from eye movements. *Journal of Experimental*
1064 *Psychology: Human Perception and Performance*, *40*(1), 342–360.
1065 <https://doi.org/10.1037/a0033854>
- 1066 Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image
1067 features on fixation selection in scenes. *Annals of the New York Academy of Sciences*,
1068 *1339*(1), 82–96. <https://doi.org/10.1111/nyas.12705>
- 1069 Nuthmann, A., & Malcolm, G. L. (2016). Eye guidance during real-world scene search: The

- 1070 role color plays in central and peripheral vision. *Journal of Vision*, 16(2):3, 1–16.
1071 <https://doi.org/10.1167/16.2.3>
- 1072 Pajak, M., & Nuthmann, A. (2013). Object-based saccadic selection during scene perception:
1073 Evidence from viewing position effects. *Journal of Vision*, 13(5):2, 1–21.
1074 <https://doi.org/10.1167/13.5.2>
- 1075 Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision*
1076 *Research*, 40(10–12), 1227–1268. [https://doi.org/10.1016/S0042-6989\(99\)00244-8](https://doi.org/10.1016/S0042-6989(99)00244-8)
- 1077 Pannasch, S., Helmert, J. R., Roth, K., Herbold, A.-K., & Walter, H. (2008). Visual fixation
1078 durations and saccade amplitudes: Shifting relationship in a variety of conditions.
1079 *Journal of Eye Movement Research*, 2(2):4, 1–19. <https://doi.org/10.16910/jemr.2.2.4>
- 1080 Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science*, 206(4417), 468–469.
1081 <https://doi.org/10.1126/science.504987>
- 1082 Reichle, E. D., & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind
1083 link. *Frontiers in Human Neuroscience*, 7, 361.
1084 <https://doi.org/10.3389/fnhum.2013.00361>
- 1085 Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network:*
1086 *Computation in Neural Systems*, 10(4), 341–350. [https://doi.org/10.1088/0954-](https://doi.org/10.1088/0954-898X/10/4/304)
1087 [898X/10/4/304](https://doi.org/10.1088/0954-898X/10/4/304)
- 1088 Reingold, E. M., & Glaholt, M. G. (2014). Cognitive control of fixation duration in visual
1089 search: The role of extrafoveal processing. *Visual Cognition*, 22(3–4), 610–634.
1090 <https://doi.org/10.1080/13506285.2014.881443>
- 1091 Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*,
1092 7(2):17, 1–22. <https://doi.org/10.1167/7.2.17>
- 1093 Rothkegel, L. O. M., Schütt, H. H., Trukenbrod, H. A., Wichmann, F. A., & Engbert, R.
1094 (2019). Searchers adjust their eye-movement dynamics to target characteristics in

- 1095 natural scenes. *Scientific Reports*, 9, 1635. <https://doi.org/10.1038/s41598-018-37548-w>
- 1096 Rovamo, J., & Virsu, V. (1979). An estimation and application of the human cortical
1097 magnification factor. *Experimental Brain Research*, 37(3), 495–510.
1098 <https://doi.org/10.1007/BF00236819>
- 1099 Scialfa, C. T., & Joffe, K. M. (1998). Response times and eye movements in feature and
1100 conjunction search as a function of target eccentricity. *Perception & Psychophysics*,
1101 60(6), 1067–1082. <https://doi.org/10.3758/BF03211940>
- 1102 Seedorff, M., Oleson, J., & McMurray, B. (2019). Maybe maximal: Good enough mixed
1103 models optimize power while controlling Type I error. *PsyArXiv*.
1104 <https://doi.org/10.31234/osf.io/xmhfr>
- 1105 Shen, J., Reingold, E. M., Pomplun, M., & Williams, D. E. (2003). Saccadic selectivity
1106 during visual search: The influence of central processing difficulty. In J. Hyönä, R.
1107 Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye*
1108 *movement research* (pp. 65–88). Amsterdam: Elsevier. [https://doi.org/10.1016/B978-](https://doi.org/10.1016/B978-044451020-4/50005-0)
1109 [044451020-4/50005-0](https://doi.org/10.1016/B978-044451020-4/50005-0)
- 1110 Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2015). Disentangling the effects of spatial
1111 inconsistency of targets and distractors when searching in realistic scenes. *Journal of*
1112 *Vision*, 15(2):12, 1–21. <https://doi.org/10.1167/15.2.12>
- 1113 Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680.
1114 <https://doi.org/10.1126/science.103.2684.677>
- 1115 Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern
1116 recognition: A review. *Journal of Vision*, 11(5):13, 1–82.
1117 <https://doi.org/10.1167/11.5.13>
- 1118 Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection.
1119 *Psychological Review*, 61(6), 401–409. <https://doi.org/10.1037/h0058700>

- 1120 Torralba, A., Oliva, A., Castelhana, M. S., & Henderson, J. M. (2006). Contextual guidance
1121 of eye movements and attention in real-world scenes: The role of global features in
1122 object search. *Psychological Review*, *113*(4), 766–786. [https://doi.org/10.1037/0033-](https://doi.org/10.1037/0033-295X.113.4.766)
1123 [295X.113.4.766](https://doi.org/10.1037/0033-295X.113.4.766)
- 1124 Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of
1125 information processing during scene perception: The relationship between saccade
1126 amplitude and fixation duration. *Visual Cognition*, *12*(3), 473–494.
1127 <https://doi.org/10.1080/13506280444000409>
- 1128 Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New
1129 York: Springer. <https://doi.org/10.1007/978-0-387-21706-2>
- 1130 Wagenmakers, E.-J., Kryptos, A.-M., Criss, A. H., & Iverson, G. (2012). On the
1131 interpretation of removable interactions: A survey of the field 33 years after Loftus.
1132 *Memory & Cognition*, *40*(2), 145–160. <https://doi.org/10.3758/s13421-011-0158-0>
- 1133 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2d ed.). New York:
1134 Springer.
- 1135 Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference.
1136 *Journal of the American Statistical Association*, *22*(158), 209–212.
1137 <https://doi.org/10.1080/01621459.1927.10502953>
- 1138 Wolfe, J. M. (2014). Approaches to visual search: Feature integration theory and guided
1139 search. In A. C. Nobre & S. Kastner (Eds.), *Oxford Handbook of Attention* (pp. 11–55).
1140 New York: Oxford University Press.
- 1141 Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011).
1142 Visual search for arbitrary objects in real scenes. *Attention Perception & Psychophysics*,
1143 *73*(6), 1650–1671. <https://doi.org/10.3758/s13414-011-0153-3>
- 1144 Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search.

- 1145 *Nature Human Behaviour*, 1, article number 0058. <https://doi.org/10.1038/s41562-017->
1146 0058
- 1147 Wolfe, J. M., O'Neill, P., & Bennett, S. C. (1998). Why are there eccentricity effects in visual
1148 search? Visual and attentional hypotheses. *Perception & Psychophysics*, 60(1), 140–
1149 156. <https://doi.org/10.3758/BF03211924>
- 1150 Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological*
1151 *Review*, 115(4), 787–835. <https://doi.org/10.1037/a0013118>
- 1152 Zelinsky, G. J., Peng, Y., Berg, A. C., & Samaras, D. (2013). Modeling guidance and
1153 recognition in categorical search: Bridging human and computer object detection.
1154 *Journal of Vision*, 13(3):30, 1–20. <https://doi.org/10.1167/13.3.30>
1155
1156