Searching Near and Far: The Attentional Template Incorporates Viewing Distance

Surya Gayet[1, 2], Elisa Battistoni[3], Sushrut Thorat[2, 4], and Marius V. Peelen[2]
[1] Experimental Psychology, Helmholtz Institute, Utrecht University
[2] Donders Institute for Brain, Cognition and Behaviour, Radboud University
[3] Center for Mind/Brain Sciences, University of Trento
[4] Institute of Cognitive Science, Osnabrück University

According to theories of visual search, observers generate a visual representation of the search target (the "attentional template") that guides spatial attention toward target-like visual input. In real-world vision, however, objects produce vastly different visual input depending on their location: your car produces a retinal image that is 10 times smaller when it is parked 50 compared to 5 m away. Across four experiments, we investigated whether the attentional template incorporates viewing distance when observers search for familiar object categories. On each trial, participants were precued to search for a car or person in the near or far plane of an outdoor scene. In "search trials," the scene reappeared and participants had to indicate whether the search target was present or absent. In intermixed "catch-trials," two silhouettes were briefly presented on either side of fixation (matching the shape and/or predicted size of the search target), one of which was followed by a probe-stimulus. We found that participants were more accurate at reporting the location (Experiments 1 and 2) and orientation (Experiment 3) of probe stimuli when they were presented at the location of size-matching silhouettes. Thus, attentional templates incorporate the predicted size of an object based on the current viewing distance. This was only the case, however, when silhouettes also matched the shape of the search target (Experiment 2). We conclude that attentional templates for finding objects in scenes are shaped by a combination of category-specific attributes (shape) and context-dependent expectations about the likely appearance (size) of these objects at the current viewing location.

Public Significance Statement
When searching for an object in our surroundings, traditional theories of visual search posit that we generate a mental picture of the object we are looking for (the "attentional template"). Depending on *where* we look (e.g., further away), however, an object will produce a vastly different (i.e., smaller) image on the retina. Here we show that observers flexibly adjust their attentional template, based on their current search location, effectively accounting for viewing distance by searching for a smaller version of the object when searching further away. These findings reconcile traditional theories of visual search with the challenges imposed by naturalistic vision.

Keywords: visual search, scene perception, object recognition, predictive processing, naturalistic vision

Supplemental materials: https://doi.org/10.1037/xhp0001172.supp

Every moment in time our retinae collect unfathomable amounts of information from the world around us. Because the vast majority of this visual input is irrelevant to our current behavioral goals, our visual system is equipped with means to favor behaviorally relevant visual input over irrelevant visual input. One such means lies at the heart of most leading theories of visual search: these theories posit that observers generate a visual representation of the object they are looking for (a so-called attentional template), thus optimally preparing the visual processing stream to favor visual input that resembles the template (such as the target object), at the expense of visual input that does not (Desimone & Duncan, 1995; Duncan & Humphreys, 1989; Eimer, 2014; Kastner & Ungerleider, 2001; Wolfe, 1994; Wolfe & Horowitz, 2004; for reviews, see Battistoni et al., 2017; Beck & Kastner, 2009). Evidence for template-based visual search mostly comes from lab-based studies using impoverished visual displays, which stand in stark contrast with the complexity of naturalistic visual environments. Therefore, it remains a matter of debate to what extent well-established mechanisms of visual search generalize to real-world vision (Peelen & Kastner, 2014; Wolfe, 2021; Wolfe & Horowitz, 2004; Wolfe, Võ, et al., 2011).

Human observers are particularly proficient in detecting objects in naturalistic scenes (F. F. Li et al., 2002; Peelen et al., 2009; Potter, 1975; Thorpe et al., 1996; Wolfe, Võ, et al., 2011), despite their inherent complexity and clutter, as compared to the typical impoverished displays that are used in most studies investigating visual search. This proficiency suggests that mechanisms of visual search are particularly well-adapted to complex naturalistic vision (Peelen & Kastner, 2014). Natural scenes provide a rich source of information that observers can capitalize on during a search, by constraining the likely locations and identity of objects in the scene (i.e., contextual guidance; Boettcher et al., 2018; Droll & Eckstein, 2008; Malcolm & Henderson, 2010; Neider & Zelinsky, 2006; Spotorno et al., 2014; Torralba et al., 2006; for a review, see Castelhano & Krzyś, 2020). Naturalistic environments, however, pose a fundamental challenge to the core principle of template-based visual search: the image that any given target object will produce on the retinae is unknown in advance because it varies with the (unknown) location of the target object. Its color or brightness depends on the illumination (e.g., in the sun, in the shade, or under artificial lighting), its shape depends on the viewpoint (e.g., viewed from the side, from above, or at an angle), and—most dramatically—its size can vary by orders of magnitudes depending on the distance between the target object and the observer. Consequently, it remains unknown to the observer what template needs to be generated to effectively search for a given target object, which calls into question the usefulness of template-based visual search during real-world vision.

In this study we test one key mechanism that could solve this problem, focusing on the predictable relationship between viewing distance and retinal object size. We test the hypothesis that human observers account for viewing distance when searching for a given object. This would entail that observers effectively search for a smaller projection of the object when searching far away (generating a smaller attentional template), and for a larger projection of the object when searching nearby (generating a larger attentional template). In favor of this hypothesis, it has been shown that attentional templates can be flexibly adjusted to match the current task demands during a naturalistic search (Yu et al., 2023). For instance, observers can adjust the tuning (or precision) of the attentional template, to account for the uncertainty of the target object's appearance (Bravo & Farid, 2012; Hout & Goldinger, 2015; Lleras et al., 2022; Witkowski & Geng, 2022), or adjust the feature content of the attentional template to optimally distinguish the target object from anticipated distractor objects (Boettcher et al., 2020; Howard et al., 2011; Lerebourg et al., 2023). Moreover, priming the upcoming target object with word-cues or semantically congruent scenes benefits subsequent search (Malcolm & Henderson, 2009; Robbins & Hout, 2020; Stein & Peelen, 2017), suggesting that observers adjust their attentional template to account for the provided context. Most specifically, we recently showed that when participants prepare to search for a target object nearby (compared to far away), patterns of neural activity emerge in the visual cortex that are similar to activity patterns evoked by viewing large (compared to small) images of this target object (Gayet & Peelen, 2022). This shows that the human visual system anticipates the size of an object depending on the viewing distance. But does this visual-like activity evoked during search preparation benefit search behavior in any way? In other words: do human observers generate distance-dependent (i.e., size-specific) attentional templates to aid visual search? One finding supporting this possibility is that observers sometimes fail to identify an object that is disproportionally large compared to its background (Eckstein et al., 2017). Going against this hypothesis, however, are results from studies showing that attentional templates can be invariant to such visual attributes as orientation (Reeder & Peelen, 2013) and size (Bravo & Farid, 2009). This invariance may particularly apply to highly familiar real-world object categories (cars, people), for which detection is highly efficient (e.g., F. F. Li et al., 2002; Thorpe et al., 1996; see also Stein & Peelen, 2017). According to this view, an object-specific attentional template (e.g., of a car) would benefit search irrespective of its orientation or size. Here, we ask whether the attentional template incorporates (retinal) object size during naturalistic visual search, when size can be directly inferred from the scene context (i.e., viewing distance).

To answer this question, we conducted a series of behavioral lab-based experiments, in which participants were searching for one of two possible object categories (a person or a car), at different viewing distances within outdoor scene photographs. The viewing distance informed participants of the (retinal image) size of the target object, allowing them to incorporate size information in their preparatory attentional template. To test whether the attentional template indeed contained size information we used a dual-task design. In "search trials" participants searched for a precued object category (a car or person) and reported which of two briefly presented scenes contained the target object. Critically, the size of the target object was—in principle—predictable, based on the layout of the search scene (Experiment 1) or on a cue instructing where to search (in depth; Experiments 2–3). The goal of these trials was to motivate participants to instill a preparatory attentional template that could potentially incorporate size information. In intermixed "catch trials," we used a dot-probe task that allows for probing attentional biases (MacLeod et al., 1986), and has been used to reveal the contents of the attentional template (Gayet & Peelen, 2019; Reeder & Peelen, 2013; Reeder et al., 2015). In this task, the search cue is unexpectedly followed by two task-irrelevant silhouettes (on both sides of fixation), of a car or person of differing sizes. Participants are tasked with responding to a simple target stimulus presented to the left or right of fixation, immediately after the presentation of the silhouettes. The idea is that, if one silhouette matches the

attentional template to a better extent than the other silhouette (e.g., a car vs. a person silhouette), attention will be directed to the location of the matching silhouette, thus improving target reports at that location. In the current study, this approach allowed us to measure a specific aspect of the search template that is key to naturalistic visual search (whether it incorporates the size of the target object, as predicted from viewing distance) while preserving the experimental control of reductionist experiments.

To preface the results, we demonstrate that attentional templates are retinal size-specific (Experiments 1–3). These size-specific attentional templates, however, only favor size-consistent visual objects that resemble the search target; they do not favor all objects of the predicted size (Experiments 1–2). The data further show that observers could infer the predicted retinal size of the search target from the viewing distance in the scene, following a location cue, even when the viewing distance changed trial-by-trial (Experiments 2–3). This showcases the ability of observers to flexibly change the size of their attentional template when searching at different locations of a visual scene. Importantly, visual discrimination performance (on an orthogonal task) was better at the location of size-consistent compared to size-inconsistent silhouettes (Experiment 3), which implies that size-consistent objects attracted spatial attention. Together, these findings show that observers infer the predicted retinal size of a search target from the viewing distance in a scene to favor target-like visual input during naturalistic visual search.

Experiment 1

Method

Transparency and Openness

The current study adheres to all Transparency and Openness Promotion guidelines regarding research transparency; in the OSF project dedicated to this study (https://osf.io/84tbv/) we provide (a) the experiment scripts and stimuli that were used for data collection, (b) the raw data, (c) the data preprocessing and analysis scripts, and (d) the complete output of all statistical analyses. The experiments in this study were not preregistered. Nonetheless, we believe that the risk of false positive inflation caused by the degrees of freedom in data analysis choices is minimized by (a) applying minimal data exclusion, by (b) presenting three internal (conceptual) replications of the main finding, by (c) using the exact same analysis pipeline in all studies, and by (d) showing consistent statistical outcomes across different types of statistical tests. The years of data collection were 2017 (Experiment 1), 2019 (Experiment 2), and 2021 (Experiment 3).

Participants

Thirty healthy students from the University of Trento participated in Experiment 1, which comprised two experimental sessions conducted on different days. All participants (25 women; $M_{age} =$ 23.3 years, $SD = 3.8$) had normal or corrected-to-normal vision and provided written informed consent to take part in the study. Most participants received monetary compensation (€8/session), but three participants took part for course credits. The experiment was approved by the Ethics Committee of the University of Trento. The sample size for Experiment 1 was based on resource availability; formal power analyses were conducted for all subsequent experiments (see Method section of Experiment 2).

Setup

Stimuli were presented on a 19″ Philips 109P monitor with a screen resolution of 1,024 × 768 pixels and a refresh rate of 100 Hz. Stimulus presentation and response registration were done with Matlab 8.0 using Psychtoolbox-3 (Brainard, 1997; Pelli, 1997). All stimuli were presented on a uniform gray background, with a black plus-sign ("+") at the center serving as a fixation point. The viewing distance was fixed at 55 cm from the monitor using a chin-rest.

Natural Scene Stimuli (Search Trials)

A total of 378 outdoor scene photographs were found via Google Image search or retrieved from previous studies. Of those, 162 had target objects (i.e., people or cars) in the foreground (near location), which were thus relatively large: 54 scenes with cars, 54 scenes with people, and 54 scenes with cars and people. Another 162 scenes had target objects in the background (far location), which were thus relatively small: again, this comprised 54 scenes with cars, 54 scenes with people, and 54 scenes with cars and people. The remaining 54 scenes contained no target objects. In order to increase the number of scene stimuli, each of these 378 scenes was horizontally mirrored, amounting to a total of 756 unique scene stimuli. The 324 scenes with near/large target objects were used in one experimental session (the Near Target session), the 324 scenes with far/small target objects were used in another experimental session (the Far Target session), and the remaining 108 scenes without target objects were used in both sessions (see Figure 1a).

All scenes were converted to grayscale and rescaled to 427 (horizontal) by 320 (vertical) pixels, subtending 15.8° by 11.7° of visual angle. The average height of the target objects was 52 pixels for "far" persons, 240 pixels for "near" persons, 56 pixels for "far" cars, and 287 pixels for "near" cars. Of note, the largest "far" object of the stimulus set was smaller than the smallest "near" object, thus ensuring the validity of the session-specific manipulation of the expected object size.

Silhouette Stimuli (Catch Trials)

The stimuli used in the catch trials were black silhouettes of cars and people, presented on a uniform gray background. A total of 576 silhouettes were selected from stimuli used in previous experiments, or created based on images of cars and people found via Google Image search, using GIMP (https://www.gimp.org). This resulted in 144 unique silhouette stimuli in each size (large, small) and category (person, car) condition (see Figure 1b). These silhouettes were scaled to match the sizes of the target objects presented within the natural scenes that are used in the search task.

Experimental Procedure

The experiment consisted of two sessions of 45 min each; a "Near Targets" session in which all target objects in the scenes were relatively nearby (and thus subtended a large retinal image), and a "Far Targets" session in which all target objects in the scenes were relatively far away (and thus subtended a small retinal image). Each participant completed both sessions on separate days, and the second session was completed within a week of the first session. The order of sessions ("Near Targets" first or "Far Target" first) was counterbalanced across participants. Each session comprised nine blocks of 64 trials each, of which 48 search trials (75%) and 16 catch trials. The silhouettes were large in half of the

Figure 1

Example of Stimuli Used in the Different Experimental Conditions of Experiment 1
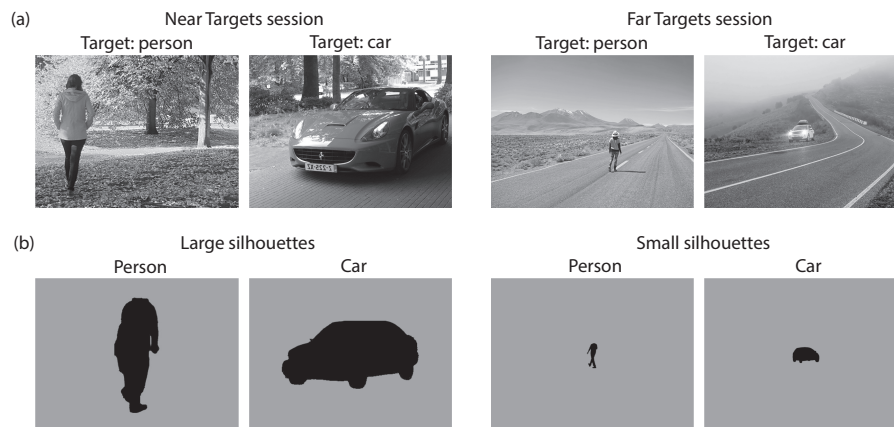
Note. (a) Scene stimuli used in the search task. During the Near Targets session, target objects (person or car) were located in the foreground, and their retinal image size was therefore relatively large. During the Far Targets session, target objects were located in the background, thus producing a relatively small retinal image. (b) Silhouette stimuli used in the catch trials. The sizes of the silhouettes were matched to the sizes of the target objects presented within the search trial scenes.

catch trials, and small in the other half. Therefore, each block comprised catch trials with two size-consistent silhouettes (i.e., large silhouette in "Near Targets" session, small silhouette in "Far Targets" session) and trials with two size-inconsistent silhouettes (i.e., large silhouette in "Far Targets" session, small silhouette in "Near Targets" session). The order of trials within a block was pseudo-randomized, so that search trials, catch trials with large silhouettes, and catch trials with small silhouettes were intermixed. The only restriction was that the first three trials of each block were always search trials, to ensure that participants were engaged in the (size-specific) search task before the first catch trial appeared. At the start of each session, participants performed one practice block to familiarize with the task.

Search Trials

The order of events in search trials is depicted in the top row of Figure 2. Each search trial started with a central fixation cross (500 ms), followed by the letter "C" or "P" (500 ms), which instructed participants to search for a car or person in the upcoming scene images (for Italian-speaking participants, this was replaced with an "M" or "P," for "macchina" and "persona," respectively).

After another fixation cross (1,000 ms), during which observers could prepare for the search task, two scenes were simultaneously presented for 67 ms on either side of fixation, in one of four possible combinations: (a) car in the left scene, person in the right scene; (b) person in the left scene, car in the right scene; (c) both person and car in the left scene, no target objects in the right scene; and (d) no target objects in the left scene, both person and car in the right scene. These combinations ensured that viewing one object (e.g., a car) in a scene was not predictive of the location of the other object, hence inciting participants to search for the cued object (rather than inferring its location from the location of the other object).

The scenes were followed by a blank screen of variable duration (range [10 ms, 300 ms]), and two backward masks that covered the same presentation area as the scenes (350 ms). The duration between scene offset and mask onset was titrated using an adaptive staircase procedure, aiming at a search task performance of 75% correct in both ("Near Targets" and "Far Targets") experimental sessions. This was done by reducing the duration of the blank screen by 20 ms when accuracy (from the sixth trial onward) rose above 75% and by increasing its duration by 20 ms when accuracy dipped below 75% correct.

The masks were followed by a fixation cross (1,660 ms), during which observers reported which scene (left or right of fixation) contained the target object, using the "z" and "n" arrow keys (for left or right scene, respectively). Finally, a feedback screen (500 ms) indicated whether they were correct ("+1") or incorrect ("+0").

To test whether the staircase procedure was successful in equating search task difficulty between the Near Targets session and Far Target session, we conducted a 2×2 repeated-measures ANOVA with the factors object (person vs. car) and distance (near vs. far), on both accuracy and response times. A main effect of distance on accuracy showed that participants were more accurate in localizing target objects in the Near Targets session ($M = 87.9\%$, $SD = 4.2$) than in the Far Target session ($M = 74.5\%$, $SD = 6.5$), $F(1, 29) = 179.12$, $p < .001$, $\eta^2 = .729$. Similarly, a main effect of distance on reaction times showed that participants were faster in localizing target objects in the Near Targets session ($M = 549$ ms, $SD = 93$) than in the Far Target session ($M = 609$ ms, $SD = 100$), $F(1, 29) = 11.18$, $p = .002$, $\eta^2 = .257$. These results show that larger objects remained easier to find than smaller objects, despite the thresholding procedure that was aimed at equating performance between distance conditions. This probably reflects that localization of relatively large objects was too easy with a presentation time of 67 ms, even at the shortest scene-mask interval of 10 ms (which motivated us to use a different staircase procedure in Experiments 2 and 3).

Catch Trials

The order of events in catch trials (dot-probe task) is depicted in the second and third rows of Figure 2. The start of a catch trial was

Figure 2

Schematic Depiction of the Experimental Procedure of Experiment 1
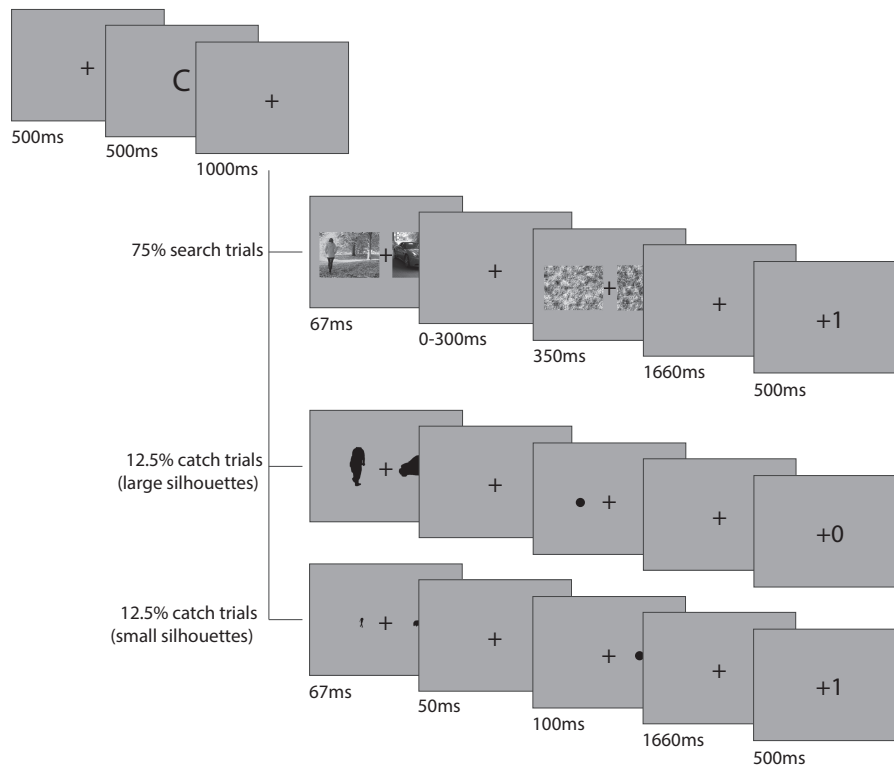
Note. Each block was made up of 64 trials presented in random order, comprising 48 search trials (75%) and 16 intermixed catch trials, half of which with small silhouettes and half of which with large silhouettes. In all trials, a letter cue instructed participants to search for a car or person. In search trials, participants reported which of two scenes (left or right of fixation) contained the cued target object. In catch trials, two task-irrelevant silhouettes appeared followed by a small target dot. Participants reported where (left of right of fixation) the target dot appeared.

indistinguishable from that of a search trial, thus inciting participants to generate an attentional template in anticipation of the search task. That is, the trial started with a fixation cross (500 ms), a letter cue (500 ms), and another fixation point (1,000 ms). Then, instead of two scenes, two silhouettes were presented on either side of the fixation (for 67 ms). The two silhouettes were either both small or both large (i.e., they were both either consistent or inconsistent with the size of search targets in the current session), and one silhouette was always of a car and the other of a person (i.e., one silhouette matched and the other silhouette mismatched the category of the search target).

After the silhouettes, a fixation point was briefly presented (50 ms), and a small circular target dot appeared on one side of fixation (100 ms); at the location of the silhouette that matched the category of the search target (*valid* trials) or at the location of the mismatching silhouette (*invalid* trials).

After the offset of the target dot, the fixation cross remained on screen for 1,660 ms, during which participants could report the location of the target dot (left or right of fixation), using the "z" and "n" arrow keys (for left or right scene, respectively). Participants were instructed to ignore the task-irrelevant silhouettes. Finally, a feedback screen (500 ms) indicated whether they were correct ("+1") or incorrect ("+0").

Data Analysis

We focus our analyses of catch trials on accuracy because pilot experiments revealed that our effects of interest were better captured by accuracy differences than reaction time differences between conditions. For transparency, and to verify that our reported effects are not the result of changes in speed-accuracy trade-offs, we report all reaction time analyses in S.1 in the online supplemental materials. Before performing the analyses, we collapsed the catch-trial data across all conditions of noninterest (e.g., the specific category of the silhouette); additional analyses in S.2 in the online supplemental materials show that none of the outcomes reported in the main manuscript depend on these conditions of noninterest.

All tests reported in the Results section and online supplemental materials are two-tailed within-subject tests with a significance threshold of 0.05. To compare between pairs of conditions, we use paired-samples *t*-tests when normality assumptions are met (according to a Shapiro–Wilk test, with a significance threshold of 0.05), and we use Wilcoxon signed-rank tests when they are violated. In case multiple factors are included in the analysis (e.g., Experiment 1), we always use repeated-measures ANOVAs, which are robust to

violations of normality (Blanca et al., 2023) and offer more flexibility than the nonparametric alternatives. Whenever parametric tests are used, we report parametric measures of central tendency (mean), effect sizes (d_z or η^2), and spread (standard deviation). Conversely, whenever nonparametric tests are used, we report nonparametric measures of central tendency (median), effect sizes (rank-biserial correlation), and spread (interquartile range). Finally, for all critical tests, we also conducted two-sided one-sample bootstrap tests (1×10^6 permutations) comparing the difference between conditions-of-interest to zero.

To address the main question of whether observers incorporate the predicted retinal size of a target object in the attentional template, we analyzed participants' average accuracy on catch trials. Catch trial data were analyzed as a function of two experimental factors: category-validity (of the target dot location relative to the silhouettes), and size-consistency (of the silhouettes with the search task session). In valid trials, the target dot appeared at the location of a silhouette that matched the search cue (i.e., a car silhouette when participants were cued to search for a car, or a person silhouette when participants were cued to search for a person). In invalid trials, the target dot appeared at the opposite location, where the silhouette mismatched the search cue (i.e., a car silhouette when participants were cued to search for a person, or a person silhouette when participants were cued to search for a car). In half of the trials, the silhouettes were size-consistent, which entails that the size of the silhouettes was consistent with the size of the search targets (i.e., large silhouettes in the "Near Targets" session, and small silhouettes in the "Far Targets" session). In the other half of the trials, the silhouettes were size-inconsistent, which entails that the size of the silhouettes was inconsistent with the size of the search targets (i.e., large silhouettes in the "Far Targets" session, and small silhouettes in the "Near Targets" session). Figure 3a illustrates the four conditions of the 2×2 factorial design. Mean accuracy scores were computed for each participant and for each of the four conditions of interest, only excluding trials in which no response was provided within the 1,660 ms time window.

Results

Catch Trial Analysis

If participants generate an attentional template that incorporates the predicted retinal size of a target object, the category-validity effect (higher accuracy for reporting target dots appearing at the category-valid location than the category-invalid location) should be more pronounced on trials with size-consistent silhouettes than with size-inconsistent silhouettes. This would imply that size-consistent silhouettes more closely resemble the attentional template than size-inconsistent silhouettes and, thus, that size information is incorporated in the attentional template.

Following size-consistent silhouettes, participants were more accurate on category-valid trials ($M = 98.6\%$, IQR = 2.8) than on category-invalid trials ($M = 89.6\%$, IQR = 10.8), $W = 406$, $p < .001$, *rank-biserial correlation* = 1.00 ($p_{bootstrap} < .001$, 95% CI [6.3%, 11.6%]). Following size-inconsistent silhouettes as well, participants were more accurate on category-valid trials ($M = 99.3\%$, IQR = 1.4) than on category-invalid trials ($M = 93.8\%$, IQR = 12.2), $W = 300$, $p < .001$, *rank-biserial correlation* = 1.00 ($p_{bootstrap} < .001$, 95% CI [4.5%, 9.3%]). The occurrence of

this validity effect shows that the attentional template contained category-selective information (i.e., distinguishing between car and person targets). Most importantly—and confirming our main hypothesis—this category-validity effect was larger for size-consistent silhouettes ($M = 7.6\%$, IQR = 9.0) than for size-inconsistent silhouettes ($M = 4.2\%$, IQR = 11.5), as showcased by a significant interaction effect between category-validity and size-consistency on response accuracy, $F(1, 29) = 9.88$, $p = .004$, $\eta^2 = .009$ ($p_{bootstrap} = .002$, 95% CI [0.7%, 3.2%]). This pattern of results (visualized in Figure 3b) supports the hypothesis that participants incorporated the expected size of target objects in their attentional template.

Note that the main effect of size-consistency was also significant, $F(1, 29) = 9.50$, $p = .004$, $\eta^2 = .009$ ($p_{bootstrap} = .001$, 95% CI [0.5%, 2.1%]), which shows that—irrespective of the location of the target dot—presenting two size-consistent silhouettes interfered more with catch-trial localization performance than presenting two size-inconsistent silhouettes (i.e., the vertical offset between lines in Figure 3b).

Interim Discussion

The goal of Experiment 1 was to test whether observers incorporate the expected size of a target object in their attentional template. This hypothesis was confirmed. Category-specific silhouette stimuli influenced localization reports of the target dot *more* when they matched the expected size of the cued target object (e.g., small silhouettes in a "Far Target session") than when they mismatched the expected size (e.g., small silhouettes in a "Near Target session"). This implies that the expected size of the cued target object was used during search preparation, otherwise, the dot-probe performance for size-consistent and size-inconsistent silhouette conditions would not differ.

In this experiment, however, observers might not have predicted the size of the target object based on the viewing distance, but could have based their expectations of object size on the prevalence of (larger or small) target objects within an experimental session. As such, it remains unclear whether observers could also incorporate object size in their attentional template during real-world search, where size needs to be inferred from the viewing distance in the scene, on a moment-to-moment basis.

The goal of Experiment 2A was to test whether observers also incorporate object size in their attentional template when they need to infer the size of the target object from the current search location in a scene, as would be done during a real-world visual search. To this end, participants now previewed the search scene that contained a location cue, informing participants about the viewing distance to the object (and thus its retinal size). This approach also allows to test whether observers can incorporate a new predicted object size in their attentional template in a trial-by-trial manner, which would indicate that observers can flexibly alter their attentional template as a function of search location (e.g., from saccade to saccade during real-world visual search). Because event-based designs (such as Experiment 2) are typically less powerful than block-based designs (such as Experiment 1), we decided to directly pit the two conditions-of-interest against each other within each trial, by contrasting a size-consistent silhouette with a size-inconsistent silhouette (both of the target object category).

Figure 3

Experimental Design and Results for the Dot-Probe Task (Catch Trials) of Experiment 1
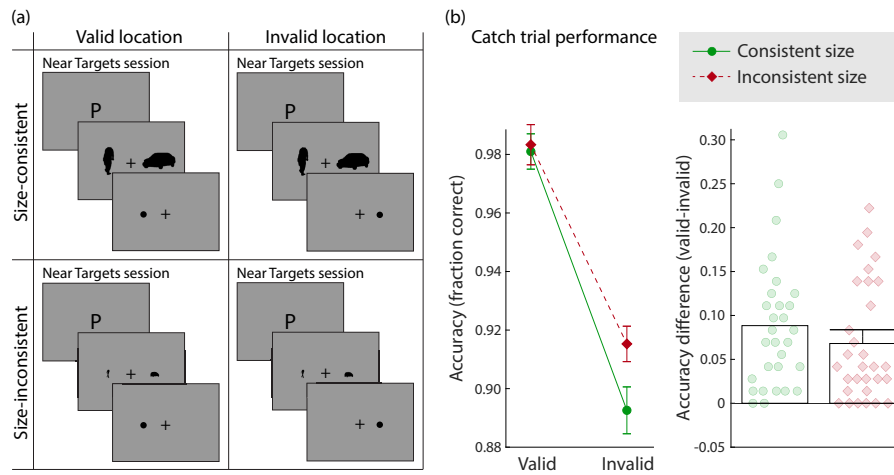
Note. (a) Visualization of the 2 × 2 factorial design (for simplicity, all four cells depict "Person" search, in a "Near Targets" session). The dot target appeared either at a valid location (i.e., at the location of a person silhouette following the "P" search cue, or at the location of a car silhouette following the "C" search cue) or an invalid location (vice versa). The size of the silhouettes was either consistent with the size of the search targets (i.e., large silhouettes in a "Near Targets" session, or small silhouettes in a "Far Targets" session) or inconsistent (vice versa). (b) Left: mean proportion correct in each of the 2 × 2 conditions depicted in panel a. Right: validity effect (performance on valid minus invalid trials) for the size-consistent and size-inconsistent conditions. Transparent dots are individual participant means; error bars in the interaction plot represent the within-subject standard error of the mean (Cousineau, 2005); the whisker on the right-most bar of the difference plot shows the 95% confidence interval of the paired difference between size-consistency conditions. See the online article for the color version of this figure.

Experiment 2

Method

Participants

Fifty-four healthy students from Radboud University participated in Experiment 2. Two participants were excluded for failing to perform above the chance level in the target probe localization task, according to a one-sided binomial test against 0.5. This resulted in a final sample of 26 participants in Experiment 2A (18 female, $M_{age} = 22.35$ years, $SD = 2.67$) and another 26 participants in Experiment 2B (22 female, $M_{age} = 22.58$ years, $SD = 3.19$).

The sample size of 26 was determined on the basis of a power analysis for a paired-samples *t*-test, conducted in G*Power. We aimed at 80% power for detecting an effect at least as large as that observed in our recent study (Experiment 1 of Gayet & Peelen, 2019; $d_z = 0.637$). In this study, we also compared performance on a dot-probe task between targets appearing at the location of size-consistent versus size-inconsistent visual objects. Due to an error in our power analysis, we eventually had 88% power for detecting said effect, as the required sample size for 80% power was actually 22.

All participants had normal or corrected-to-normal vision and provided written informed consent to take part in the study. Participants either received monetary compensation (€10/session) or course credits (one participant). The experiment was approved by the Ethics Committee of the Social Sciences Faculty of Radboud University Nijmegen, The Netherlands (ECSW2017-2306-517).

Setup

Participants were tested in a dark room where a chinrest kept their viewing distance fixed at 57 cm of a 24″ BenQ monitor with a screen resolution of 1,920 × 1,080 pixels and a refresh rate of 120 Hz. Stimulus presentation and response registration were done with Matlab 2015b using Psychtoolbox-3 (Brainard, 1997; Pelli, 1997). All stimuli were presented on a uniform gray background (30 Cd/m²), with a white outer circle (83.60 Cd/m²; 0.30 degrees of visual angle; dva) and a black inner circle (0.19 Cd/m²; 0.10 dva), serving as a central fixation dot.

Natural Scene Stimuli (Search Trials)

A total of 126 outdoor scenes were created for the purpose of this experiment, using an HD digital photo camera. Photographs were taken at 14 different locations, and nine different stimuli were created at each of these locations from the exact same viewpoint (using a tripod): scenes comprised either a car or a person, positioned either nearby or far away, and positioned either on the left or right half of the scene (to induce spatial uncertainty). Also, one "empty" scene was created, in which no target object was present.

All scenes were converted to greyscale (see Figure 4a) and were scaled to subtend 13.3 by 9.0 dva. Finally, based on each of the 14 empty scenes, four additional stimuli were created by superimposing a red or blue horizontal line indicating where the near or far objects touched the ground in that specific scene (i.e., the wheels of the car, or the feet of the person). These lines would serve as

Figure 4

Example of Stimuli Used in the Different Experimental Conditions of Experiments 2 and 3
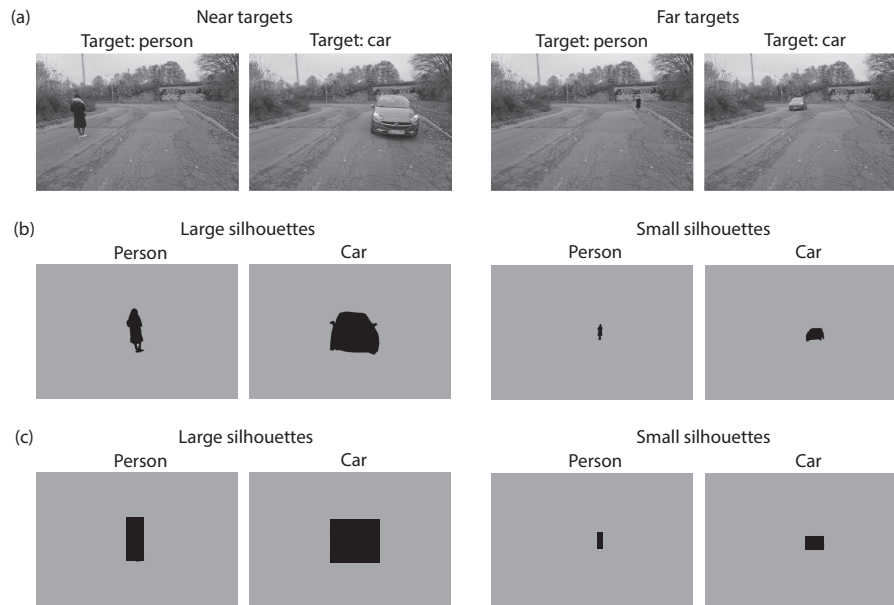
Note. (a) Scene stimuli used in the search task. Target objects (person or car) in each of the 14 scene families could be either located in the foreground or background (large or small target image). (b) Silhouette stimuli used in the catch trials of Experiment 2A, which were cropped out of the corresponding (near and far, car and person) search scenes. (c) Corresponding silhouette stimuli used in the catch trials of Experiment 2B, which were rectangles with the same height and width as the original silhouettes.

distance cues. Note that, in contrast to Experiment 1, distance (and therefore object size) was manipulated within-scene, and therefore distant objects were inherently smaller than nearby objects. As such, there was no need to compare the mean retinal object-sizes between near and far conditions.

Silhouette Stimuli (Catch Trials)

The stimuli used in the catch trials of Experiment 2A were black silhouettes of cars and people, cropped out of the scene stimuli described above, and presented on a uniform gray background (Figure 4b). This resulted in 112 silhouettes; 28 exemplars in each car or person, and near or far condition (i.e., two exemplars in each condition, for each specific scene). Importantly, because the silhouettes were cropped out of the scenes and because the distance cues were based on the positions of the objects in the scenes, the size of each silhouette corresponds exactly to the size of the target objects (that participants could expect) in the scenes.

The stimuli used in the catch trials of Experiment 2B were black rectangles, with the exactly same (maximum) height and width as the silhouettes used in Experiment 2A. Thus, the sizes and height-to-width ratios of these pseudo-silhouettes matched the sizes and height-to-width ratios of the target objects in each scene.

Experimental Procedure

The experiment consisted of a single session of approximately 60 min, starting with verbal and visual instructions, a practice block with search trials only (24 trials), a practice block with catch trials

only (24 trials), and a practice block with both trial types intermixed (24 trials total, of which six catch trials). Then, participants completed 16 experimental blocks of 32 trials each, of which 24 search trials (75%) and eight catch trials were randomly intermixed.
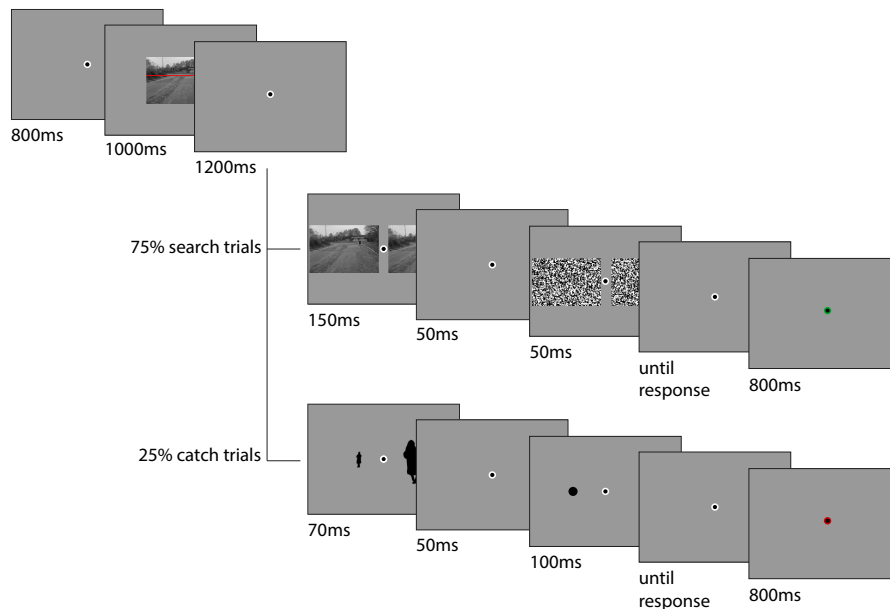
Search Trials

The order of events in search trials is depicted in Figure 5 (top row). Each search trial started with a central fixation dot (800 ms), followed by an empty scene (i.e., devoid of target objects) overlaid with a colored bar (1,000 ms). The color of the bar (blue or red) indicates the category of the target object (blue for car and red for person for even participant numbers, and the opposite for odd participant numbers). The vertical position of the bar indicates the location of the target object in depth, thus allowing observers to predict the size of the target object.

After another fixation cross (1,200 ms), the same outdoor scene that was previewed before was simultaneously presented for 150 ms on both sides of fixation, one of which comprised the target object while the other one contained no object at all. Scene offset was followed by a fixation screen (50 ms), a white-noise mask (50 ms), and another fixation screen that lasted until participants provided a response. Participants indicated by means of a key press which image (left or right of fixation) contained the target object. The white part of the fixation dot turned green or red to indicate whether the response was correct or not.

In order to equate task difficulty between the different search target conditions (near and far, car and person), we superimposed pink (i.e., 1/f) noise onto the scene stimuli, and adaptively adjusted the

Figure 5

Schematic Depiction of the Experimental Procedure of Experiment 2A

Note. Each block was made up of 32 trials presented in random order, comprising 24 search trials (75%) and eight intermixed catch trials, each containing a large and a small silhouette (both of the cued object category). In all trials, the vertical position of a colored bar instructed participants where (in depth) the target object would appear. The color of the bar (red or blue) indicated which target object to search for (car or person). Participants reported which of two versions of the same scene (left or right of fixation) contained the cued target object. In catch trials, two task-irrelevant silhouettes appeared, followed by a small target dot. Participants reported where (left or right of fixation) the target dot appeared. The procedure of Experiment 2B was identical to that of Experiment 2A, but the silhouettes were replaced by rectangles (see Figure 4c). See the online article for the color version of this figure.

percentage of noise using accelerated stochastic approximation (Kesten, 1958), separately for each search target condition. Unlike traditional up-down staircase procedures, accelerated stochastic approximation adjusts the step sizes by taking into account the stability of the estimated threshold. In doing so, we expected to stabilize performance levels at 75% correct in all search conditions (Faes et al., 2007). In contrast to Experiment 1, the onset asynchrony between the scene stimuli (mixed with pink noise) and the mask stimuli (white noise) remained fixed at 50 ms.

To test whether this staircase procedure was successful in equating difficulty between search target conditions, we conducted a repeated-measures ANOVA with the factors object (person vs. car) and distance (near vs. far), on both accuracy and response times. In Experiment 2A, far target search and near target search differed neither in terms of accuracy, $F(1, 25) = 0.024$, $p = .879$, $\eta^2 < .001$, nor in terms of reaction times, $F(1, 25) = 0.041$, $p = .841$, $\eta^2 < .001$. Similarly, in Experiment 2B, far target search and near target search differed neither in terms of accuracy, $F(1, 25) = 1.425$, $p = .244$, $\eta^2 = .032$, nor in terms of reaction times, $F(1, 25) = 0.197$, $p = .661$, $\eta^2 = .003$. Accuracies in all search task conditions ranged between 77.2% and 79.5% correct, and reaction times ranged between 590 and 631 ms. Taken together, the staircase procedure of Experiment 2 was successful in equating task difficulty across near and far search conditions, in terms of both accuracy and reaction times.

Catch Trials

Figure 5 (bottom row) illustrates the order of events in catch trials (dot-probe task) of Experiment 2A. The start of a catch trial was indistinguishable from that of a search trial, thus inciting participants to generate an attentional template in anticipation of the search task. Instead of the two search scenes, however, two silhouettes were presented for 70 ms, on either side of fixation: a large and a small silhouette of the cued target object, that was cropped out of the corresponding search scene. These silhouettes were vertically centered on the fixation dot, and presented on the left and right side of fixation at equal eccentricity (the eccentricity was varied on a trial-by-trial basis, to match the horizontal position of the objects in the scenes). After a fixation screen (50 ms), a black target dot would appear (19 Cd/m^2; 0.3 dva in diameter) at the center of one of the two previously presented silhouettes. In half of the trials, the target dot appeared at the location of the large silhouette, on the other half of the trials the target dot appeared at the location of the small silhouette (see Figure 6b). Participants reported the location of the target dot (left or right of fixation) by keypress, after which the white part of the fixation dot turned green or red to indicate whether they reported the location of the target dot correctly or not.

Experiment 2B was identical to Experiment 2A, except that the silhouettes in the catch trials were replaced by rectangles encompassing

Figure 6
Experimental Design and Results for the Dot-Probe Task (Catch Trials) of Experiment 2
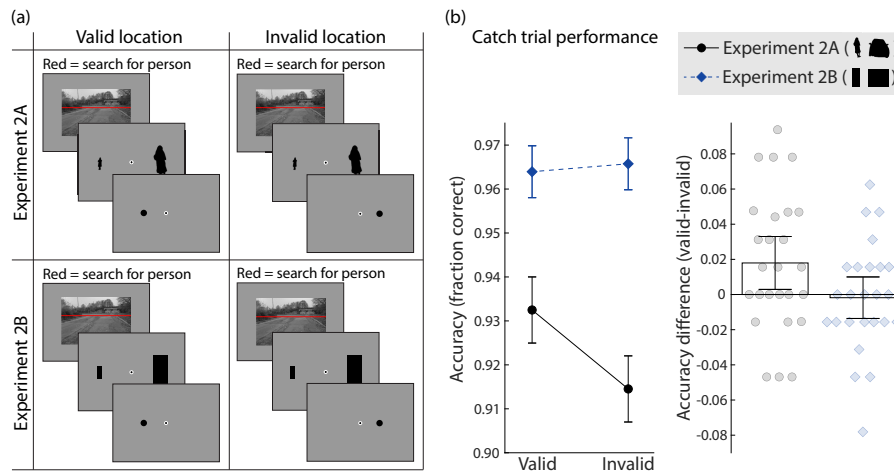
Note. (a) Visualization of the within-subject designs of Experiments 2A and 2B. For illustrative purposes, we here only depict trials in which participants were cued to search for a distant (i.e., relatively small image of a) person. There were two validity conditions: the target dot either appeared at the location of the size-consistent (valid; here: small) or the size-inconsistent silhouette (invalid; here: large). (b) In Experiment 2B, the silhouettes were replaced by filled rectangles, encompassing the (maximum) height and width of each silhouette. (c) Mean proportion correct for the size-consistent versus size-inconsistent locations in Experiment 2A (silhouettes) and 2B (rectangles). Error bars represent the 95% CI of the paired difference between conditions. CI = confidence interval. See the online article for the color version of this figure.

the maximum height and width of each silhouette. As such, the silhouette was still size-valid or size-invalid with regard to the current search task (and even comprised height-to-width ratios that could distinguish between car and person silhouettes), but lacked the target object-specific shape contours (see Figure 6b).

Data Analysis

The analysis approach is identical to that of Experiment 1, unless otherwise specified. To address the main question of whether the predicted retinal size of target objects is incorporated in the attentional template, we analyzed how accurately participants reported the location of the target dot. Data was analyzed as a function of one experimental factor, size-validity: On half of the trials, the target dot appeared at the location of a size-consistent silhouette (i.e., the large silhouette during near search, or the small silhouette during far search). In the other half of the trials, the target dot appeared at the location of the size-inconsistent silhouette (i.e., the large silhouette during the far search, or the small silhouette during the near search). Mean accuracy scores were computed for each participant and for both conditions of interest. No trials were excluded from the analysis.

If participants generate an attentional template that incorporates the predicted retinal size of a target object, we expect to observe a size-validity effect in Experiment 2A. More accurate responses to targets appearing at the location of a size-consistent silhouette than a size-inconsistent silhouette implies that size information was extracted from the viewing distance in the scene and incorporated in the attentional template. If a size-validity effect is found in Experiment 2A but not in Experiment 2B, this would show that the size information in the attentional template only applies to visual

input that matches the category-specific shape of the target object (i.e., of a car or person). If, instead, a size-validity effect is observed in both Experiment 2A and 2B, this would show that attentional templates favor category-matching and size-matching visual input independent of one another (i.e., as if there were multiple attentional templates biasing search in parallel).

Results

Catch Trial Analysis

To test the main hypothesis that attentional templates change as a function of viewing distance, we performed a paired-samples *t* test on catch trial accuracy, contrasting accuracy for target dots appearing at the location of size-consistent silhouettes with target dots appearing at the location of size-inconsistent silhouettes (see Figure 6c). In Experiment 2A, participants were more accurate in locating the dots appearing at the position of a size-consistent silhouette (M = 93.1%, SD = 7.6) compared to a size-inconsistent silhouette (M = 91.2%, SD = 7.8), t(25) = 2.54, p = .018, d_z = 0.498 (p_bootstrap = 0.010, 95% CI [0.5%, 3.4%]). This was not the case in Experiment 2B, where the silhouettes were replaced by rectangles, p > .7, d_z = −0.059 (p_bootstrap = .796, 95% CI [−1.3%, 1.0%]). Following the general approach of equivalence testing (Lakens et al., 2018), we established that the effect in Experiment 2B was negligible, as it was significantly smaller than half the effect size observed in Experiment 2A, p = .035. An independent-samples *t* test contrasting the validity effect of both experiments confirmed that the validity effect was larger in Experiment 2A (M = 1.9%, SD = 3.9) than in Experiment 2B (M = −0.2%, SD = 3.1), t(50) = 2.17, p = .034, d_z = 0.603

($p_{bootstrap} = .014$, 95% CI [0.4%, 3.9%]). Together, these data show that observers incorporate the expected size of a target object (as inferred from the viewing distance in the scene) in their attentional template. This attentional template, however, does not prioritize all visual input of the expected object size, but only visual input of the expected object size that also matches the visual characteristics of the object category.

We also noted that performance was generally lower in Experiment 2A ($M = 92.2\%$, $SD = 8.7$) than in Experiment 2B ($M = 96.5\%$, $SD = 8.7$), $W = 217$, $p = .027$, Hodges–Lehmann Estimate $= .023$ (Mann–Whitney test used due to the violation of assumption of equal variances). This might reflect that—overall—presenting visual stimuli that are more relevant to the participant (i.e., silhouettes compared to rectangles) interferes more with the localization of the dot target.

Interim Discussion

The goal of this experiment was to test whether observers flexibly incorporate the expected size of a target object in their attentional template. The observation that target dots were more accurately reported at the location of a size-consistent silhouette compared to a size-inconsistent silhouette of the target object demonstrates that size information was incorporated in the attentional template. Moreover, the present results extend the results of Experiment 1, by showing that participants predicted the retinal size of the cued search target, based explicitly on the viewing location in the scene. This demonstrates how observers could incorporate size information for efficient template-based search under naturalistic conditions, by updating the expected object size during search (e.g., across eye-movements).

The present results do not show, however, whether these size-specific attentional templates influence visual search by shifting spatial attention toward the location of size-matching visual input. There are two distinct accounts that could explain the accuracy difference in localizing target dots that appear at the location of size-consistent versus size-inconsistent silhouettes. One possibility is that participants mistook the size-consistent silhouette (more often than the size-inconsistent silhouette) for the search target; if participants report the location of the size-consistent silhouette, this gives a correct target localization response in size-consistent (i.e., valid) trials and an incorrect target localization response in size-inconsistent (i.e., invalid) trials. The other possibility is that the size-consistent silhouette attracted spatial attention due to its match with the attentional template, causing improved visual discrimination performance at the attended location and, consequently, better target localization performance.

Experiment 3 was designed to directly test this second possibility. Here, the target dot is replaced with a triangle pointing upward or downward (see Figure 7a), and participants are instructed to report the orientation of the arrow target (up vs. down). In this case, mistakenly responding to the size-consistent silhouette (left vs. right localization) would not influence discrimination performance on the arrow target (up vs. down discrimination). As such, better discrimination performance of the arrow target at the location of the size-consistent compared to the size-inconsistent silhouette would unequivocally demonstrate that spatial attention was drawn toward the size-consistent silhouette, thereby enhancing target discrimination performance.

Experiment 3

Method

Differences With Experiment 2

All methods were identical to that of Experiment 2A, except for (a) the use of upward and downward-pointing target triangles instead of a target dot in the catch trials, (b) the ensuing use of an up-down response instead of a left-right response in catch trials, and (c) the set-up on which the experiment was conducted.

Search Trials

In Experiment 3, far target search and near target search differed neither in terms of accuracy, $F(1, 25) = 1.159$, $p = .292$, $\eta^2 = .015$, nor in terms of reaction times, $F(1, 25) = 2.943$, $p = .099$, $\eta^2 = .028$. Accuracies in all search task conditions ranged between 78.4% and 82.9% correct, and reaction times ranged between 653 and 676 ms. As such, the staircase procedure of Experiment 3 was also successful in equating task difficulty across near and far search conditions, in terms of both accuracy and reaction times.

Participants

Another 26 healthy students from Radboud University participated in Experiment 3 (20 female, $M_{age} = 20.51$ years, $SD = 2.82$). The sample size was based on the same power analysis as Experiments 2A and 2B. Participants received monetary compensation (€10/session). The experiment was approved by the Ethics Committee of the Social Sciences Faculty of Radboud University.

Results

Catch Trial Analysis

To test the main hypothesis that size-specific attentional templates guide spatial attention, we performed a Wilcoxon signed-rank t test on catch trial accuracy, contrasting up-down discrimination performance for target triangles appearing at the location of size-consistent silhouettes with that of size-inconsistent silhouettes (Figure 7b). Participants were more accurate in reporting the orientation of the triangles when they appeared at the location of a size-consistent silhouette ($M = 98.4\%$, IQR $= 3.0$) compared to a size-inconsistent silhouette ($M = 96.1\%$, IQR $= 4.7$), $W = 108.5$, $p = .038$, *rank-biserial correlation* $= .596$ ($p_{bootstrap} = .014$, 95% CI [0.3%, 2.7%]). This shows that size-specific attentional templates cause spatial attention to shift toward template-matching visual input.

General Discussion

According to the idea of template-based search, observers generate a visual representation of the target object prior to search onset, which favors target-like visual input at the expense of nontarget visual input. Here, we investigated whether human observers adjust the size of the attentional template to account for viewing distance during the search, capitalizing on the predictable relationship between retinal object size and viewing distance. This would entail that observers effectively search for a smaller "image" of an object when searching further away, and for a larger "image" of that same object when searching closer by. We used a dot-probe task (MacLeod et al., 1986) to probe

Figure 7
Experimental Design and Results for the Dot-Probe Task (Catch Trials) of Experiment 3
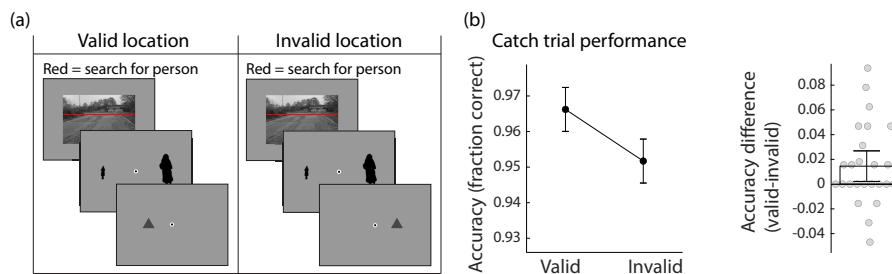
Note. (a) The procedure and experimental design of Experiment 3 are identical to that of Experiment 2A, but now the target was a triangle, and participants reported whether it was pointing upward or downward. Note that the intermixed search task trials still required left/right responses. (b) Mean proportion correct for the size-consistent (valid) versus size-inconsistent (invalid) locations. Error bars represent the 95% CI of the paired difference between conditions. CI = confidence interval. See the online article for the color version of this figure.

the content of the search template (Reeder & Peelen, 2013), intermixed with a search task that incited participants to search for a given object (person or car), at a specific viewing distance in a scene photograph. To summarize our findings: (a) in Experiment 1 we confirm earlier results that observers incorporate category-specific shape information in the attentional template, allowing observers to favor car-like visual input over person-like visual input and vice versa, depending on the search target; (b) most importantly, in Experiments 1, 2A, and 3, we demonstrate that observers incorporate the expected retinal size of the target object in the attentional template, favoring visual input of the expected retinal size over differently sized visual input; (c) the results of Experiment 3 show that the template causes a shift of spatial attention toward size-matching objects; (d) in Experiments 2A and 3, we demonstrate that observers flexibly predict the retinal size of a target object from the real-world viewing distance in a scene; and finally, (e) Experiments 1 and 2 show that distance-dependent size information and category-specific shape information are entangled, yielding a single attentional template that is both shape and size specific.

Confirming earlier work (Reeder & Peelen, 2013; Reeder et al., 2015), the results show that participants incorporate the category of the target object (i.e., person or car) in their attentional template. This conclusion stems from the finding that participants in Experiment 1 were more accurate in reporting the location of a target dot at the location of a category-valid (e.g., person silhouette following a "person" search cue) than a category-invalid silhouette (e.g., person silhouette following a "car" search cue). This effect was observed in virtually all participants and was between 2 and 4 times larger than the size-based validity effect observed in Experiment 2A. Since the dot-probe task used uniform black silhouettes, category-specific attentional templates (at least partly) rely on differences in shape attributes. Earlier work showed that category-specific attentional templates (for cars and persons) consist of category-diagnostic object parts (e.g., the wheel of a car or an arm of a person), and that these are rotationally invariant (Reeder & Peelen, 2013). Similarly, here the difference in behavioral responses to target dots following category-valid versus category-invalid silhouettes implies that category-specific information was maintained during search preparation. Accordingly, we conclude that observers incorporate target object-specific attributes in the attentional template, thus favoring target-like visual input during naturalistic visual search.

The key finding of the present study is that participants incorporate the expected size of a target object in their attentional template. Our conclusion stems from the observation that the category-specific effect on the target-dot report (discussed above) increased when silhouettes were of the expected size (within an experimental session) compared to the unexpected size. This is consistent with the idea that the attentional template is a visual representation of the object category that is scaled to the expected size of the target object (Gayet & Peelen, 2022). Can this finding explain how observers search for objects at different distances within a three-dimensional real-world environment? During real-world search, the expected size of a target object does not vary on a day-by-day basis, however, but rather depends on (a) the viewing distance that observers extract from the search scene, which (b) varies on a moment-to-moment basis. The data of Experiments 2A and 3 show that, indeed, when participants are cued (on a trial-by-trial basis) to search at a particular location in a natural scene photograph, an attentional template is generated with a size that corresponds to the viewing distance at the current search location. Specifically, when cued to search for a relatively distant target object, observers are better at reporting the target-dot following a small silhouette of the target object, but when cued to search for a relatively nearby target objects, observers are better at reporting the target-dot following a large silhouette of the target object. The present study provides the first behavioral evidence that human observers take into account the predicted size of search targets (as inferred from the viewing distance) when generating attentional templates to search for objects in a naturalistic scene. This finding could explain why observers sometimes fail to recognize objects that are inappropriately sized given the surrounding scene context (Eckstein et al., 2017). Based on the present study alone, it remains unknown whether the size of the attentional template is adjusted continuously, to match the specific viewing distance at the current search location, or whether it is adjusted categorically, favoring *relatively* larger objects over smaller objects during near search (e.g., Becker et al., 2010, 2013; Bravo & Farid, 2009). Nonetheless, the findings that attentional templates incorporate viewing distance contribute to the literature on mechanisms of attentional selection in naturalistic visual search (Eimer, 2014; Peelen & Kastner, 2014).

Embedding the present findings in the broader literature on attentional selection in visual search, we can ask how observers go about finding their keys on a cluttered desk or searching for their friend at a

crowded festival. The answers to this question distinguish between two types of search strategies: environmental cues that guide attention, and feature-based guidance (Wolfe & Horowitz, 2017). A large body of work has shown how—during naturalistic search—participants quickly direct their gaze toward locations that are likely to contain the target object, such as shoes on the floor, a phone on the desk, or a toothbrush near the bathroom sink (e.g., Boettcher et al., 2018; Droll & Eckstein, 2008; Neider & Zelinsky, 2006), even when set sizes are very large (Wolfe, Alvarez, et al., 2011). In parallel, other studies have shown that feature-based attention benefits naturalistic visual search (Bahle & Hollingworth, 2019; Bahle et al., 2018; Hollingworth & Bahle, 2020), by drawing attention to target-specific features such as color or shape across the visual field (Maunsell & Treue, 2006; Nuthmann & Malcolm, 2016; Thorat & Peelen, 2022). Such spatially global effects of attention have also been observed for category-level (car, person) search in natural scenes (Peelen et al., 2009). The present study shows one way in which these two mechanisms (i.e., scene guidance and feature-based guidance) interact: when observers are searching for their phone, they use a template comprising phone-specific visual features (small, black, rectangular), some of which are adjusted according to environmental cues (in this case, the size is adjusted based on the viewing distance extracted from the scene).

We considered two ways in which size information in the attentional template could have affected behavior in the dot-probe task. Either participants mistakenly responded to the silhouettes instead of responding to the target-dot (erroneously recognizing size-valid silhouettes as target objects, more often than size-invalid silhouettes). Alternatively, size-valid silhouettes attracted attention, thereby enhancing visual detection of the target-dot at the attended location. Experiment 3 was designed to discriminate between these possibilities, by replacing the target-dot detection task with a target-triangle (up-down) discrimination task. Here, we capitalized on the fact that performance on a variety of visual tasks should be better at the attended location compared to the unattended location (Carrasco et al., 2000). The data showed that participants were more accurate in reporting the orientation of the target-triangle when it followed a size-valid silhouette (e.g., a small silhouette of the target object during a distant search) compared to when it followed a size-invalid silhouette (e.g., small silhouette of the target object during nearby search). Under these circumstances, mistakenly responding to the silhouette (as if it were the search target) would not affect the accuracy of reporting the orientation of the triangle. Instead, if size-valid silhouettes attracted spatial attention (due to their match with the attentional template), participants should be better at discriminating the orientation of the briefly presented triangle-target. Considering that the magnitude of the size-validity effect was virtually identical between Experiment 2A (target-dot) and Experiment 3 (target-triangle), it can be argued that size information (extracted from viewing distance) mainly impacts visual search performance by attracting attention toward target objects of the predicted retinal image size.

The present results show how the attentional template can incorporate multiple aspects of the search target; in this case, its category-specific shape and its context-dependent size. Importantly, these two aspects of the attentional template are codependent: object-selectivity is more pronounced for objects of the expected target size (Experiment 1) and size-selectivity is more pronounced for objects of the expected target shape (Experiment 2). This argues against the existence of two independent attentional templates (a size-specific template and a shape-specific template) and demonstrates

that a single attentional template incorporates both category-specific shape information and context-dependent size information (see also Gayet & Peelen, 2022). The contributions of these two aspects to visual search performance seem asymmetrical, however. The findings of Experiment 1 show that category-selectivity is observed not only for size-consistent silhouettes, but also for size-inconsistent silhouettes. Thus, when searching for a car at a particular distance, car-like visual input is favored over noncar-like visual input, even when it does not match the predicted object size. The findings of Experiment 2, on the other hand, show that size-selectivity is observed only for target object silhouettes, and not for rectangular silhouettes that only preserved the height and width of the target objects. Can we then conclude that, when searching for a car, visual input of the predicted object size is only favored over differently sized visual input when it contains car-like visual shape properties? This might be too simplistic. Here, the (rectangle) silhouettes were very crisp and clearly lacked the shape attributes of a car. During real-world vision, the exact shape (or color, etc.) of a visual object might be more uncertain, for instance, because it is out of focus, occluded, or viewed peripherally. When the shape (or other property) of an object is uncertain, this can be accounted for by widening the tuning of attentional templates accordingly (Bravo & Farid, 2012; Hout & Goldinger, 2015; Lleras et al., 2022; Witkowski & Geng, 2022). From this perspective, we might expect that visual input of the predicted object size is favored as long as its shape (or some other property) does not provide sufficient evidence *against* it being the search target.

Nonetheless, within our experimental paradigm, shape information guided attention irrespective of whether the silhouette size was consistent with the search distance, whereas size information only guided attention when the shape of the silhouettes was consistent with the search target. We consider four possible (nonexclusive) explanations for this asymmetry. First, it could be that shape information generally dominates size information in visual search, akin to how color information tends to dominate over other stimulus attributes (e.g., Williams, 1966; Wolfe & Horowitz, 2004). Arguably, visual shape properties are more diagnostic of a target object, and less variable over time, compared to size information. Second, the dominance of one attribute over another could depend on the specific task; here, participants were instructed to report the location of a person or car (as defined by its visual shape properties), they were not instructed to search for an object of a particular size. Reversing the instructions might reverse the relative dominance of the two features in driving visual search. Third, the relative dominance of shape over size (and orientation; Reeder & Peelen, 2013) might be specific to highly familiar object categories, for which detection is particularly efficient (F. F. Li et al., 2002; Stein & Peelen, 2017; Thorpe et al., 1996; Treisman, 2006). When searching for less familiar objects, observers might rely more heavily on context-dependent attributes (such as inferred retinal size), because they fail to extract the most distinctive category-specific visual features. Fourth, the relative dominance of the shape and size properties might depend on the diagnosticity of shape-information and size-information for distinguishing between the two types of target objects (cars and persons), and between the target objects and their surroundings (size might be more relevant when distractor objects have similar shapes as the target; e.g., searching for a soccer ball among basket balls and tennis balls). The idea that the contents of the template are context-dependent is very common in the literature about attentional templates (Geng & Witkowski, 2019; Navalpakkam & Itti, 2007). It is known that templates are influenced by prior knowledge about,

for example, the scene layout (C. L. Li et al., 2018), distractor identity (Howard et al., 2011; Lerebourg et al., 2023), and object cooccurrence (Mack & Eckstein, 2011): different tasks and set-ups result in different templates. Taken together, it is likely that the exact way in which different (object-specific or context-dependent) features are combined in the attentional template depends on both task and stimulus context (for a recent discussion, see Yu et al., 2023).

The current results imply that attentional templates incorporate the expected retinal (or proximal) size of target objects, not their perceived (or distal) size. This provides support to behavioral studies showing that, in naturalistic visual search, an object's predicted proximal size (Eckstein et al., 2017; Sherman et al., 2011) or proximal shape (Aldegheri et al., 2023; Morales et al., 2020) contributes to attentional guidance and object recognition. Using proximal rather than distal features of target objects to guide visual search makes sense, when considering feed-forward accounts of visual perception; from a feed-forward perspective, retinal size is extracted faster than perceived (or veridical) size. Thus, biasing visual input based on proximal features would allow for earlier selection of target-like visual input than selection based on distal features. On the other hand, scene context modulates representations of object size even in the primary visual cortex (Fang et al., 2008; Murray et al., 2006; Sperandio & Chouinard, 2015; Sperandio et al., 2012), although this may reflect delayed feedback processes (Schmidt & Haberkamp, 2016; Zeng et al., 2020). Moreover, the visual system as a whole seems to preferentially represent the perceived size of objects rather than their retinal size (Amit et al., 2012; Cate et al., 2011; Chouinard & Ivanowich, 2014; Fang et al., 2008; Gabay et al., 2016; Konkle & Oliva, 2011, 2012; Liu et al., 2009; Murray et al., 2006; Pooresmaeili et al., 2013; Schwarzkopf et al., 2011; Sperandio et al., 2012; Sterzer & Rees, 2006). As such, using distal stimulus features to guide visual search would also allow for relatively early and effective visual selection. In line with this, we previously showed that memory templates favor perceptually size-matching objects over perceptually size-mismatching objects, even when both objects have the exact same retinal size (Gayet & Peelen, 2019). Considering human observers' proficiency in naturalistic visual search, it is very well possible that search preparation simultaneously capitalizes on proximal (retinal image-based) as well as distal (perceived) features, thus favoring target-like visual input over irrelevant visual input at multiple steps of the visual processing hierarchy.

Conclusion

During real-world visual search, any given object that we are searching for can produce a wide variety of visual input, depending on where it is located in the world. The eventual appearance of the object therefore remains unknown during the search, complicating template-based visual search. Conversely, however, the specific location in the scene at which we currently search for an object strongly constrains the appearance of the object. Notably, when the real-world size of the object is known, the viewing distance directly informs the observer of the retinal image size that the object produces. Here, we show that observers predict the appearance of the target object from the current search location in the scene. Specifically, participants formed predictions about the retinal size of the object, given the (cued) viewing distance. This size information is then incorporated in the attentional template, so that target-like visual input is favored—in particular—when its retinal size is

consistent with the viewing distance. Put simply: we provide direct behavioral evidence that the attentional template is scaled to account for viewing distance (in line with recent neuroimaging evidence, Gayet & Peelen, 2022). Finally, we show that visual input that matches this category-selective and size-specific attentional template attracts attention. Together, these findings demonstrate how preparatory attentional templates operate during naturalistic visual search.

Inclusion and Diversity Statement

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure that the study questionnaires were prepared in an inclusive way. While citing references scientifically relevant to this work, we also actively worked to promote gender balance in our reference list.

Constraints on Generality Statements

Our participant sample was recruited among the student population of the University of Trento (Italy), and Radboud University (The Netherlands), and therefore consists of highly educated, predominantly Caucasian, subjects with a predominantly Western background. Also, females are overrepresented in the sample. Based on this, we advocate caution in generalizing our findings to other populations. At the same time, the present study investigates fundamental properties of visual search that are commonly studied across species, including human primates, nonhuman primates, and even nonprimate mammals. Therefore, we do not expect the general principles studied here to vastly differ between primate species, let alone between human gender, ethnic, or cultural groups.

References

Aldegheri, G., Gayet, S., & Peelen, M. V. (2023). Scene context automatically drives predictions of object transformations. *Cognition*, *238*, Article 105521. https://doi.org/10.1016/j.cognition.2023.105521

Amit, E., Mehoudar, E., Trope, Y., & Yovel, G. (2012). Do object-category selective regions in the ventral visual stream represent perceived distance information? *Brain and Cognition*, *80*(2), 201–213. https://doi.org/10.1016/j.bandc.2012.06.006

Bahle, B., & Hollingworth, A. (2019). Contrasting episodic and template-based guidance during search through natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(4), 523–536. https://doi.org/10.1037/xhp0000624

Bahle, B., Matsukura, M., & Hollingworth, A. (2018). Contrasting gist-based and template-based guidance during real-world visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(3), 367–386. https://doi.org/10.1037/xhp0000468

Battistoni, E., Stein, T., & Peelen, M. V. (2017). Preparatory attention in visual cortex. *Annals of the New York Academy of Sciences*, *1396*(1), 92–107. https://doi.org/10.1111/nyas.13320

Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, *49*(10), 1154–1165. https://doi.org/10.1016/j.visres.2008.07.012

Becker, S. I., Folk, C. L., & Remington, R. W. (2010). The role of relational information in contingent capture. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1460–1476. https://doi.org/10.1037/a0020370

Becker, S. I., Folk, C. L., & Remington, R. W. (2013). Attentional capture does not depend on feature similarity, but on target-nontarget relations. *Psychological Science*, *24*(5), 634–647. https://doi.org/10.1177/0956797612458528

Blanca, M. J., Arnau, J., García-Castro, F. J., Alarcón, R., & Bono, R. (2023). Non-normal data in repeated measures ANOVA: Impact on Type I error and power. *Psicothema*, *35*(1), 21–29. https://doi.org/10.7334/psicothema2022.292

Boettcher, S. E., Draschkow, D., Dienhart, E., & Võ, M. L. H. (2018). Anchoring visual search in scenes: Assessing the role of anchor objects on eye movements during visual search. *Journal of Vision*, *18*(13), Article 11. https://doi.org/10.1167/18.13.11

Boettcher, S. E., van Ede, F., & Nobre, A. C. (2020). Functional biases in attentional templates from associative memory. *Journal of Vision*, *20*(13), Article 7. https://doi.org/10.1167/jov.20.13.7

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

Bravo, M. J., & Farid, H. (2009). The specificity of the attentional template. *Journal of Vision*, *9*(1), Article 34. https://doi.org/10.1167/9.1.34

Bravo, M. J., & Farid, H. (2012). Task demands determine the specificity of the search template. *Attention, Perception, & Psychophysics*, *74*(1), 124–131. https://doi.org/10.3758/s13414-011-0224-5

Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: Support for signal enhancement. *Vision Research*, *40*(10–12), 1203–1215. https://doi.org/10.1016/S0042-6989(00)00024-9

Castelhano, M. S., & Krzyś, K. (2020). Rethinking space: A review of perception, attention, and memory in scene processing. *Annual Review of Vision Science*, *6*(1), 563–586. https://doi.org/10.1146/annurev-vision-121219-081745

Cate, A. D., Goodale, M. A., & Köhler, S. (2011). The role of apparent size in building- and object-specific regions of ventral visual cortex. *Brain Research*, *1388*, 109–122. https://doi.org/10.1016/j.brainres.2011.02.022

Chouinard, P. A., & Ivanowich, M. (2014). Is the primary visual cortex a center stage for the visual phenomenology of object size? *The Journal of Neuroscience*, *34*(6), 2013–2014. https://doi.org/10.1523/JNEUROSCI.4902-13.2014

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

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Droll, J., & Eckstein, M. (2008). Expected object position of two hundred fifty observers predicts first fixations of seventy seven separate observers during search. *Journal of Vision*, *8*(6), Article 320. https://doi.org/10.1167/8.6.320

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10.1037/0033-295X.96.3.433

Eckstein, M. P., Koehler, K., Welbourne, L. E., & Akbas, E. (2017). Humans, but not deep neural networks, often miss giant targets in scenes. *Current Biology*, *27*(18), 2827–2832.e3. https://doi.org/10.1016/j.cub.2017.07.068

Eimer, M. (2014). The neural basis of attentional control in visual search. *Trends in Cognitive Sciences*, *18*(10), 526–535. https://doi.org/10.1016/j.tics.2014.05.005

Faes, L., Nollo, G., Ravelli, F., Ricci, L., Vescovi, M., Turatto, M., & Pavani, F. (2007). Small-sample characterization of stochastic approximation staircases in forced-choice adaptive threshold estimation. *Perception & Psychophysics*, *69*(2), 254–262. https://doi.org/10.3758/BF03193747

Fang, F., Boyaci, H., Kersten, D., & Murray, S. O. (2008). Attention-dependent representation of a size illusion in human V1. *Current Biology*, *18*(21), 1707–1712. https://doi.org/10.1016/j.cub.2008.09.025

Gabay, S., Kalanthroff, E., Henik, A., & Gronau, N. (2016). Conceptual size representation in ventral visual cortex. *Neuropsychologia*, *81*, 198–206. https://doi.org/10.1016/j.neuropsychologia.2015.12.029

Gayet, S., & Peelen, M. V. (2019). Scenes modulate object processing before interacting with memory templates. *Psychological Science*, *30*(10), 1497–1509. https://doi.org/10.1177/0956797619869905

Gayet, S., & Peelen, M. V. (2022). Preparatory attention incorporates contextual expectations. *Current Biology*, *32*(3), 687–692.e6. https://doi.org/10.1016/j.cub.2021.11.062

Geng, J. J., & Witkowski, P. (2019). Template-to-distractor distinctiveness regulates visual search efficiency. *Current Opinion in Psychology*, *29*, 119–125. https://doi.org/10.1016/j.copsyc.2019.01.003

Hollingworth, A., & Bahle, B. (2020). Feature-based guidance of attention by visual working memory is applied independently of remembered object location. *Attention, Perception, & Psychophysics*, *82*, 98–108. https://doi.org/10.3758/s13414-019-01759-8

Hout, M. C., & Goldinger, S. D. (2015). Target templates: The precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics*, *77*, 128–149. https://doi.org/10.3758/s13414-014-0764-6

Howard, C. J., Pharaon, R. G., Körner, C., Smith, A. D., & Gilchrist, I. D. (2011). Visual search in the real world: Evidence for the formation of distractor representations. *Perception*, *40*(10), 1143–1153. https://doi.org/10.1068/p7088

Kastner, S., & Ungerleider, L. G. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia*, *39*(12), 1263–1276. https://doi.org/10.1016/S0028-3932(01)00116-6

Kesten, H. (1958). Accelerated stochastic approximation. *The Annals of Mathematical Statistics*, *29*(1), 41–59. https://doi.org/10.1214/aoms/1177706705

Konkle, T., & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 23–37. https://doi.org/10.1037/a0020413

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, *74*(6), 1114–1124. https://doi.org/10.1016/j.neuron.2012.04.036

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Lerebourg, M., de Lange, F. P., & Peelen, M. V. (2023). Expected distractor context biases the attentional template for target shapes. *Journal of Experimental Psychology: Human Perception and Performance*, *49*(9), 1236–1255. https://doi.org/10.1037/xhp0001129

Li, C. L., Aivar, M. P., Tong, M. H., & Hayhoe, M. M. (2018). Memory shapes visual search strategies in large-scale environments. *Scientific Reports*, *8*(1), Article 4324. https://doi.org/10.1038/s41598-018-22731-w

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, *99*(14), 9596–9601. https://doi.org/10.1073/pnas.092277599

Liu, Q., Wu, Y., Yang, Q., Campos, J. L., Zhang, Q., & Sun, H. J. (2009). Neural correlates of size illusions: An event-related potential study. *NeuroReport*, *20*(8), 809–814. https://doi.org/10.1097/WNR.0b013e32832be7c0

Lleras, A., Buetti, S., & Xu, Z. J. (2022). Incorporating the properties of peripheral vision into theories of visual search. *Nature Reviews Psychology*, *1*(10), 590–604. https://doi.org/10.1038/s44159-022-00097-1

Mack, S. C., & Eckstein, M. P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision*, *11*(9), Article 9. https://doi.org/10.1167/11.9.9

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, *95*(1), 15–20. https://doi.org/10.1037/0021-843X.95.1.15

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

Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, *10*(2), Article 4. https://doi.org/10.1167/10.2.4

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

Morales, J., Bax, A., & Firestone, C. (2020). Sustained representation of perspectival shape. *Proceedings of the National Academy of Sciences*, *117*(26), 14873–14882. https://doi.org/10.1073/pnas.2000715117

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience*, *9*(3), 429–434. https://doi.org/10.1038/nn1641

Navalpakkam, V., & Itti, L. (2007). Search goal tunes visual features optimally. *Neuron*, *53*(4), 605–617. https://doi.org/10.1016/j.neuron.2007.01.018

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621. https://doi.org/10.1016/j.visres.2005.08.025

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

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94–97. https://doi.org/10.1038/nature08103

Peelen, M. V., & Kastner, S. (2014). Attention in the real world: Toward understanding its neural basis. *Trends in Cognitive Sciences*, *18*(5), 242–250. https://doi.org/10.1016/j.tics.2014.02.004

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Pooresmaeili, A., Arrighi, R., Biagi, L., & Morrone, M. C. (2013). Blood oxygen level-dependent activation of the primary visual cortex predicts size adaptation illusion. *The Journal of Neuroscience*, *33*(40), 15999–16008. https://doi.org/10.1523/JNEUROSCI.1770-13.2013

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

Reeder, R. R., & Peelen, M. V. (2013). The contents of the attentional template for category-level search in natural scenes. *Journal of Vision*, *13*(3), Article 13. https://doi.org/10.1167/13.3.13

Reeder, R. R., van Zoest, W., & Peelen, M. V. (2015). Involuntary attentional capture by task-irrelevant objects that match the attentional template for category detection in natural scenes. *Attention, Perception, & Psychophysics*, *77*, 1070–1080. https://doi.org/10.3758/s13414-015-0867-8

Robbins, A., & Hout, M. C. (2020). Scene priming provides clues about target appearance that improve attentional guidance during categorical search. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(2), 220–230. https://doi.org/10.1037/xhp0000707

Schmidt, F., & Haberkamp, A. (2016). Temporal processing characteristics of the Ponzo illusion. *Psychological Research*, *80*(2), 273–285. https://doi.org/10.1007/s00426-015-0659-8

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience*, *14*(1), 28–30. https://doi.org/10.1038/nn.2706

Sherman, A. M., Greene, M. R., & Wolfe, J. M. (2011). Depth and size information reduce effective set size for visual search in real-world scenes. *Journal of Vision*, *11*(11), Article 1334. https://doi.org/10.1167/11.11.1334

Sperandio, I., & Chouinard, P. A. (2015). The mechanisms of size constancy. *Multisensory Research*, *28*(3–4), 253–283. https://doi.org/10.1163/22134808-00002483

Sperandio, I., Chouinard, P. A., & Goodale, M. A. (2012). Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. *Nature Neuroscience*, *15*(4), 540–542. https://doi.org/10.1038/nn.3069

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision*, *14*(2), Article 7. https://doi.org/10.1167/14.2.7

Stein, T., & Peelen, M. V. (2017). Object detection in natural scenes: Independent effects of spatial and category-based attention. *Attention, Perception, & Psychophysics*, *79*, 738–752. https://doi.org/10.3758/s13414-017-1279-8

Sterzer, P., & Rees, G. (2006). Perceived size matters. *Nature Neuroscience*, *9*(3), Article 302. https://doi.org/10.1038/nn0306-302b

Thorat, S., & Peelen, M. V. (2022). Body shape as a visual feature: Evidence from spatially-global attentional modulation in human visual cortex. *NeuroImage*, *255*, Article 119207. https://doi.org/10.1016/j.neuroimage.2022.119207

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522. https://doi.org/10.1038/381520a0

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

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*(4–8), 411–443. https://doi.org/10.1080/13506280500195250

Williams, L. G. (1966). The effect of target specification on objects fixated during visual search. *Perception & Psychophysics*, *1*(5), 315–318. https://doi.org/10.3758/BF03215795

Witkowski, P. P., & Geng, J. J. (2022). Attentional priority is determined by predicted feature distributions. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(11), 1201–1212. https://doi.org/10.1037/xhp0001041

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, *28*(4), 1060–1092. https://doi.org/10.3758/s13423-020-01859-9

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics*, *73*, 1650–1671. https://doi.org/10.3758/s13414-011-0153-3

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501. https://doi.org/10.1038/nrn1411

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), Article 0058. https://doi.org/10.1038/s41562-017-0058

Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84. https://doi.org/10.1016/j.tics.2010.12.001

Yu, X., Zhou, Z., Becker, S. I., Boettcher, S. E., & Geng, J. J. (2023). Good-enough attentional guidance. *Trends in Cognitive Sciences*, *27*(4), 391–403. https://doi.org/10.1016/j.tics.2023.01.007

Zeng, H., Fink, G. R., & Weidner, R. (2020). Visual size processing in early visual cortex follows lateral occipital cortex involvement. *The Journal of Neuroscience*, *40*(22), 4410–4417. https://doi.org/10.1523/JNEUROSCI.2437-19.2020