Visual search as an embodied process: The effects of perspective change and external reference on search performance

Huiyuan Zhang

Department of Psychology, Sun Yat-sen University, Guangzhou, China ✉

Jing Samantha Pan

Department of Psychology, Sun Yat-sen University, Guangzhou, China
Guangdong Provincial Key Laboratory of Social Cognitive Neuroscience and Mental Health, Guangzhou, China ✉

Traditional visual search tasks in the laboratories typically involve looking for targets in 2D displays with exemplar views of objects. In real life, visual search commonly entails 3D objects in 3D spaces with nonperpendicular viewing and relative motions between observers and search array items, both of which lead to transformations of objects' projected images in lawful but unpredicted ways. Furthermore, observers often do not have to memorize a target before searching, but may refer to it while searching, for example, holding a picture of someone while looking for them from a crowd. Extending the traditional visual search task, in this study, we investigated the effects of image transformation as a result of perspective change yielded by discrete viewing angle change (Experiment 1) or continuous rotation of the search array (Experiment 2) and of having external references on visual search performance. Results showed that when searching from 3D objects with a non-zero viewing angle, performance was similar to searching from 2D exemplar views of objects; when searching for 3D targets from rotating arrays in virtual reality, performance was similar to searching from stationary arrays. In general, discrete or continuous perspective change did not affect the search outcomes in terms of accuracy, response time, and self-rated confidence, or the search process in terms of eye movement patterns. Therefore, visual search does not require the exact match of retinal images. Additionally, being able to see the target during the search improved search accuracy and observers' confidence. It increased search time because, as revealed by the eye movements, observers actively checked back on the reference target. Thus, visual search is an embodied process that involves real-time information exchange between the observers and the environment.

Introduction

Imagine a common scenario, on a rainy day, students leave their umbrellas outside the university's cafeteria on their way to lunch. After their meal, they walk out and swiftly pick up their own umbrellas. This everyday action requires visual search, which encompasses the observer, search items, and motions of either or both. Cognitive psychologists have been studying visual search using laboratory-based experiments, in which participants report the presence or absence of a specific object (the target) among other items (the distracters) displayed on computer monitors. The targets and distracters are typically simple (e.g. letters or color graphics) and distinguished on several dimensions (Treisman & Gelade, 1980; Koch & Ullman, 1987; Duncan & Humphreys, 1989; Wolfe & Gancarz, 1997).

However, there are key differences between a laboratory-based visual search and visual search in the real world. Classic laboratory-based visual search contains static image information, which is sufficient for accomplishing the laboratory-based experimental tasks. In a classic laboratory-based search task, the search array normally contains 2D stimuli (such as contours, line drawings, or silhouettes) that have countable distinct features between targets and distracters (for example, shape, color, size, orientation, etc., and their combinations), occupy the central part of the visual field, and do not change their appearances during the search (Zhang, Feng, Ma, Lim, Zhao, & Kreiman, 2018). But in real life, the appearances of 3D search items change as the spatiotemporal relations between search objects and observers change. Factors such as viewing distances, angle of observation, lighting and shadows, occlusion, and relative motion between the observer and search items may alter the appearances of the

search items and yield uncountable combinations of features that may or may not be saliently distinguishable (Seidl-Rathkopf, Turk-Browne & Kastner, 2015). Although these changes of appearance are regular (meaning they are regulated by the natural laws of physics), they are difficult to predict. For example, when looking for a red silicone spatula in a utensil drawer, how this spatula appears depends on color of light (e.g. natural light versus warm yellow light from a lightbulb), its orientation (flat or side-up), its spatial relations with other items in the drawer (occlusion), the spatial relations between it and the observer (e.g. the viewing distance and eye height of the observer affect viewing angle), and so on. Some of these factors (e.g. lighting condition) may be known to the observer before searching; some (e.g. the spatula's orientation inside the drawer) may not. Some factors (e.g. viewing angle) may change during the search; and some (e.g. occlusion among search items) may not. Either way, in real life situations, forming an a priori mental model of the search target is complex and it is questionable whether this is necessary or sufficient for successful visual search in natural environments. Therefore, is visual search resistant to changes of the objects' images?

Recent empirical studies on real world searches have raised that image-based information alone might be limited in predicting natural search activities (Kingstone, Smilek, Ristic, Kelland, Friesen, & Eastwood, 2003). For example, when participants walked to pick up targets and avoid obstacles in virtual reality (VR), the presence of salient distracters, which were objects with bright colors and high luminance, did not alter the participants' performance. Gaze duration on salient distractors accounted for only 0.2% of total trial time; in other words, participants did not attend to the task-irrelevant objects, despite their image saliency (Rothkopf, Ballard, & Hayhoe, 2007). In another real-world search experiment (Foulsham, Chapman, Nasiopoulos, & Kingstone, 2014), participants were to walk from the laboratory through the building to the faculty mailroom, locate a target mailbox among 120 boxes fixed on a wall, and pick up an envelope from it. In half of the trials, the target mailboxes were outlined with bright pink borders. This manipulation should make the search faster because of the high image saliency as predicted by classic visual search theories (Treisman & Gelade, 1980; Wolfe & Gancarz, 1997; Zhao & Koch, 2011). However, the response time was not shortened with the saliently outlined targets. Convergingly, in a series of experiments (Hayes & Henderson, 2019), when participants searched for the letter "L" embedded in real-life scene photographs, they tended to look at the semantically meaningful regions than the salient regions of the scene. The authors carved the search scenes into small patches and rated the patches' semantic richness to create meaning maps. The authors also generated saliency maps according to the

distinctiveness of local features. They found that when looking for an "L" in real-life scenes, 30% of gazes fell in the meaningful regions, and only 8% fell in the salient regions. Thus, it is crucial to look beyond image-based information to investigate real-world goal-directed search tasks.

Traditionally, a visual search has been widely used as a method to study attention and results from visual search experiments have provided invaluable theoretical explanations for the processes of attention capture, guidance, and allocation. Understandably, perceptual information in these tasks is simplified to contain only highly distinguishable static image from an exemplar view, for example, looking for a martini glass, which is symbolized by ⍟, from trophies, which are symbolized by ♔. Nonetheless, implicit in visual search is the perceptual task of object recognition and both visual search and object recognition involve recognizing and classifying visual patterns. For object recognition, observers integrate many dimensions or features to classify one object. For a visual search, the observers use a few dimensions or features to classify many objects. According to Nakayama and Martini (2011), object recognition and visual search are on the "two extremes of plausible trade-offs between dimensions versus objects" (page 8). Hence, a visual search and object recognition may share common issues, and it may be worthwhile to seek solutions for issues of a visual search from object recognition literatures.

As part of the natural viewing process, appearances of objects vary. So, does visual search remain robust and functional given the constant image change? This is similar to the problem of invariance in object recognition, which is concerned with how objects are recognized with various lighting, viewing perspectives, exemplars, etc. Central to both, the problem is whether and how accurate and stable perception is achieved despite changes of object appearance.

In natural viewing, progressive transformation of retinal images is yielded by the continuous relative motions between the observer and objects. The progressive transformation is regular and lawful and contains information that specifies the spatial relations and motions of the observers and objects. Images and the lawful transformation of them, known as optic flow, are two sources of information in the optics that specify spatial layout and 3D structures in the world. In other words, the change of images does not create a problem but solves it, at least when an observer is allowed to view the continuous changing process. The act of perception is spatiotemporal.

Integrating static image structure and dynamic optic flow information, accurate and stable perception of the objects' spatial relations and recognition of visual events were achieved in a robust fashion, despite perturbations of image information caused

by visual blur (Pan, Bingham, & Bingham, 2013), occlusion (Pan, et al., 2013), or orientation change (Pan, Bingham, Chen, & Bingham, 2017). This is because the co-existing optic flow and image structure information form a synergetic relationship. On the one hand, optic flow is strong in specifying 3D structures. For example, when multiple randomly textured surfaces that were separated in depth, without motion, their depth relations were not perceptible and they all appeared to be one surface. But when they rotated and precessed rigidly, the spatial relations or depth orders between them were readily perceptible and observers unmistakenly identified their locations in 3D (Pan, Li, Chen, Mangiaracina, Connell, Wu, Wang, Bingham, & Hassan, 2017). However, optic flow is temporally unstable and disappears when the motion stops. On the other hand, image structure information, such as hue and contrast, is projected by the opaque surfaces in the world. Although it does not immediately specify depth structure, it is available as long as the objects are visible and thus is temporally stable. When image structure and optic flow interact, optic flow specifies spatial structures and calibrates image structures (that is, assigns spatial meanings to the image information) at the same time. The spatial relations perceived with optic flow is hence preserved in the stable image structure information and the interaction between optic flow and image structure allow the observers to continue perceiving spatial structures with ongoing motion and after the motion stops. This accurate and stable perception is a result of an embodied process that spans over the observer and the surfaces/objects in the world.

In the current study, we treat visual search as an embodied process and focus on two perception-related aspects to learn how effective the search of objects leverages on the interaction between the observer and the external structures in the world. First, we explore whether visual search withstands image transformations as a result of either discrete or continuous perspective change. In everyday viewing, often we are looking at objects from directions that are not along the line of sight, yielding non-parallel projection of visual targets, whose images are different from the 2D exemplar views typically used in classic visual search studies. Hence, we introduce a discrete perspective change by simulating viewing with a 35-degree angle of declination, as if someone is sitting at a desk and looking down at the items on the desk surface, and test visual search performance with the corresponding image transformations. Additionally, characteristic of everyday viewing, the observer and/or the objects may be moving, yielding continuous perspective changes and progressive transformation of the search items' projected images. Accordingly, we introduce a continuous perspective change where the entire search array rotates and test its effects on visual search

performance. Overall, the perspective change, be it discrete or continuous, gives rise to distorted images of objects projected on the retina and relates to the issue of invariance. We compare search performances in these conditions to search based on static 2D exemplar views of objects to study whether and how systematic variation of image information affects visual search.

Second, an embodied process entails active exchange of information between an observer and the environment, including picking up information from the surroundings in real time. Everyday activities, such as scanning the room to find a spot to fit some new furniture, counting fingers to do math, and listing work deadlines on Google Calendar, are all cases of parasitizing the local environment and acquiring just-in-time information to alleviate internal cognitive load and boost cognitive performance (Clark & Chalmers, 1998). In an experiment, when participants selected and assembled colored blocks to replicate a model, their eye movements suggested that participants did not memorize the whole model before selecting the individual pieces; instead, they looked back and forth at a part of the model and copied it piecewise (Pelz, Hayhoe, & Loeber, 2001). Similarly, in a brick sorting experiment (Droll, & Hayhoe, 2007), observers were to pick up or put down one of five bricks according to features, such as color, width, height, and texture. Instead of memorizing the features and mentally comparing features of the brick to be sorted with features that indicated picking up or putting down, participants re-fixated on the bricks each time before picking them up or putting them down. In both experiments, the participants' behaviors demonstrated a tendency to acquire information in real time to reduce memory load and avoid memory-related errors. Similarly, when searching for an object or a person, one could memorize their appearance and then look through the search array for the target; or one could hold a picture of the search target in hand and look and compare the target and search array items in real time. The questions are when the external reference is available, whether searchers use the just-in-time information and whether the reference affects visual search performance. To answer these questions, we manipulate the presence of the target during the search and compare search performance and eye movements with or without the visible reference object.

Visual search is a conscious cognitive activity that relies on perception (Treisman, 1982; Theeuwes, Kramer, & Belopolsky, 2004) and is modulated by attention (Wolfe & Horowitz, 2017). Ultimately, the purpose of a visual search is to guide actions or facilitate decision making. For this purpose, metacognitive sensitivity, defined as being able to judge one's own performance outcome as correct or

incorrect, is critical (Fleming, Weil, Nagy, Dolan, & Rees, 2010). A reliable visual search should not only be accurate and fast, but also require a searcher to have veridical judgment of their own search performance. However, in many perceptual tasks, behavioral response and metacognitive judgment are incongruent, such as feeling confident about an incorrect response or vice versa (Kunimoto, Miller, & Pashler, 2001; Washburn, Smith, & Taglialatela, 2005; Lau & Passingham, 2006; Lau & Passingham, 2007; Szczepanowski, & Pessoa, 2007; Fleming & Dolan, 2010). For example, when discriminating grating bars' orientations, the observers' felt that confidence was independent of their task performance and furthermore it biased the subsequent perceptual decisions (Samaha, Switzky, & Postle, 2019). It has been proposed that this dissociation between perceptual performance and metacognition was because performance depends on weighing and comparing multiple stimuli but subjectively felt confidence depends on signal strength of the chosen stimulus (Samaha, Lemi, & Postle, 2017). Given that perceiving stimuli is the first step after which visual search can only take place, it is possible that, in a visual search, individuals' metacognitive sensitivity is also dissociated from their task performance. To verify this, we collect participants' self-rated confidence after each search trial and compare it with their search performance. Furthermore, we also test whether metacognitive sensitivity may be improved when introducing factors such as motion and reference targets.

In sum, there are two goals of this study. First, we investigate whether a visual search in 3D environments, which typically contain perspective change of the search items and hence image distortions, follows the same behavioral trends as revealed by classic laboratory-based experiments using static exemplar 2D views of stimuli. Specifically, we compare searching for 3D objects when they were presented with a slanted view (35 degrees angle of declination) versus with a frontal view (objects' silhouettes; Experiment 1) and study the search performances when the observers look down at stationary 3D search objects (discrete perspective change) or when they look down at rotating 3D search objects (continuous perspective change; Experiment 2). Second, we examine the effect of having external references that allow acquiring information in real time on the visual search performance. To do so, we manipulate the presence of search targets during search and compare search performances when an observer must remember the targets (as in classic visual search experiments) versus when they can make reference to the targets during the search (as an embodied process). The former taxes one's cognitive resources and the latter utilizes real-time perceptual information. Overall, we manipulate search stimuli and their motions to create closer approximations of a visual search in everyday life and study search behaviors and the metacognition

of them when an embodied visual search is or is not attainable.

Experiment 1

In this experiment, we introduced two key manipulations to a classic visual search task to simulate a more representative search process in the real world. First, we manipulated the viewing angle or projection of search items onscreen and created two types of stimuli. In classic visual search experiments, typically the search items were either 2D search items (e.g. contours or letters) or frontal displays of 3D objects, which were placed at eye height and hence showed their exemplar views (e.g. silhouettes). In this experiment, we continued to display frontal views of search items in one condition, but we also added a slanted view condition, where pictures of 3D search items were shot by a camera pointing down with a 35 degree angle of declination. The slanted view condition represented a more general case of search in real life, because often when observers look out to find something, the line of sight is not orthogonal to the target. It is thus meaningful to test whether search is resistant to image distortions yielded by a change of viewing angle.

The second key manipulation was the availability of the target throughout the search. In one condition, similar to traditional search tasks, observers first learned and remembered the target and then looked for it from an array of various items. In the other condition, observers also learned the search target first, but when they were searching, the target appeared on the screen along with the search array items. The first kind of search taxed memory (memory search). The second kind of search allowed observers to refer to the target during the search (reference search), offload a memorized target to structures in the world, and acquire real time information during the search.

With manipulations of stimulus type (frontal views versus slanted views of objects) and search type (memory search versus reference search), we created one condition that closely resembled a classic visual search experiment – searching with frontal views of objects and memorizing a target before search (the frontal-memory search condition), and one condition that was representative of search in real life – looking down at objects from a perspective and being able to look back and forth between the target and the search array (the slanted-reference search condition). Comparing performances across these manipulations, we aim to uncover whether behavioral trends found in classic visual search experiments were similar to behaviors in more complex and more natural searches.

Methods

Participants

To determine the appropriate sample size, we used G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) with four within-subject factors and set the effect size at Cohen's d = 0.25 and the alpha value at 0.05. Results indicated that a sample size of 10 would produce a power of 0.95. Accordingly, we recruited 13 adults (18—25 years old, 8 women) with normal or corrected-to-normal vision and each was reimbursed at ¥30/hour for their time and effort. All of them had normal or corrected-to-normal vision. This study was approved by the Institutional Review Board of Sun Yat-sen University. Informed consent was obtained from all participants.

Stimuli and apparatus

Composite LEGO blocks were used as search items. Four types of LEGO pieces were put together to create 24 uniquely shaped LEGO blocks. The four types of LEGO pieces (0.8 H × 0.8 D × 0.8 L cm^3 pieces with 2 studs, 0.8 H × 0.8 D × 1.6 L cm^3 pieces with 4 studs, 0.8 H × 0.8 D × 2.4 L cm^3 pieces with 6 studs, and 0.8 H × 0.8 D × 3.2 L cm^3 pieces with 8 studs) had the same height and depth, but varied in length. The 24 composite LEGO blocks differed on two dimensions

(all had the same depth of 0.8 cm) and none of them was symmetrical. All LEGO blocks were spray-painted to be gray. See Figure 1.

Images of the LEGO blocks were randomly selected to be the search targets or distracters. In one condition, silhouettes of the LEGO blocks were used. These were the exemplar views or frontal views of the LEGO blocks with viewing angle being 0 degrees, as if one is looking at the LEGO blocks that were placed on the line of sight and perpendicular to it. In the other condition, LEGO blocks were placed on a flat table and pictures (resolution = 1000 × 685 pixels) of them were taken with a Canon camera (E470) that was held at 40 cm above the table surface with a tripod and pointed down with a 35 degree angle between the camera lens and the table surface. We took professional photographic measures (e.g. using a reflector and adjusting the shutter speed), to control for light reflection and shadows. These were the slanted views of the LEGO blocks. Note that although the LEGO blocks were 3D, the distinguishing features existed only in two dimensions and the distinguishing features were detectable in both the frontal views and the slanted views of the LEGO blocks. The two viewing perspectives were tested in separate experimental sessions with counterbalanced orders between subjects.

Experimental stimuli were rendered using MATLAB Psychophysics Toolbox (Brainard & Vision, 1997;
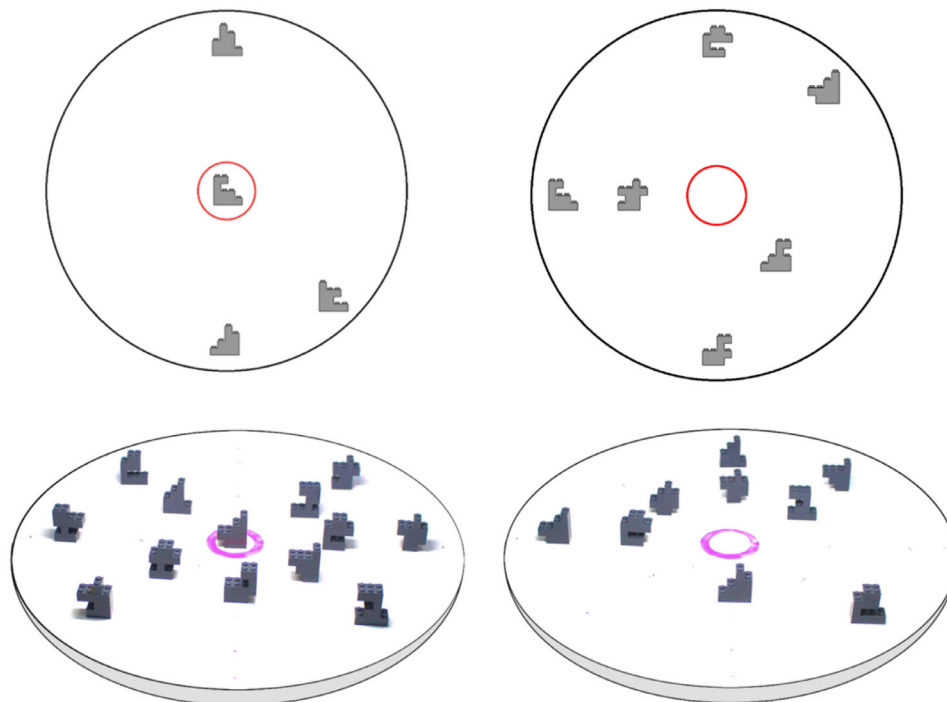

Figure 1. Stimuli and display in Experiment 1. Examples of search array in the frontal view reference search trial (top left), the frontal view memory search trial (top right), the slanted view reference search trial (bottom left), and the slanted view memory search trial (bottom right).

Kleiner et al., 2007) and displayed on a 27-inch ASUS PG279Q monitor with resolution of 2560 × 1440 pixels, refresh rate of 144 Hz and brightness set at 165 cd/m^2. The eye-to-screen viewing distance was 60 cm. Each trial had a learning phase and a search phase. A red circle (diameter = 3 degrees) was visible in the center of the screen throughout all trials. During the learning phase of all trials, the search target (the display size was 2 degrees × 2 degrees) was presented inside the red circle; during the search phase of half of the trials, the search target (the display size was 2 degrees × 2 degrees) was presented inside the red circle (the reference search trials) and in the other trials, there was nothing inside the red circle (the memory search trials). There were 16 possible locations for search items to appear during the search phase and these possible locations were evenly distributed on two concentric rings (with radii of 7.8 degrees and 10.4 degrees) around the center red circle (see Figure 1). The probability of a target appearing in any of the possible target locations was equal. There was no occlusion between objects. In half of the trials, there was a target in the search array (target present trials) and, in the other half, there was no target in the search array (target absent trials). Note that for the slanted view condition, the target appeared slightly different in the learning phase and in the search phase, because the target was always at the center of the turntable in the learning phase, but in the search phase it could be anywhere amid other items occupying a non-centered location on the turntable.

Eye movement during the search was tracked using an EyeLink 1000 Plus Desktop Mount eye tracker (SR Research Inc., Mississauga, Canada), which was controlled by a native EyeLink host PC. Monocular eye position was collected from the eye with higher uncorrected visual acuity at a sampling frequency of 1000 Hz.

Procedures

Participants came to the laboratory and were explained about the purpose and procedures of the experiment. After signing the consent forms, they were seated in front of the testing computer and placed their heads on a forehead-and-chin rest. A nine-point calibration was first done to validate the eye tracker. During the experiment, when there was a ≥0.5 degrees deviation, the eye tracker was recalibrated. Drift correction was done before stimulus onset for each trial. Participants received verbal instructions that they were going to see one target LEGO block and subsequently report the presence or absence of it from many LEGO blocks.

A trial started with a target displayed for 3 seconds, followed by the onset of the search array containing 3, 6, 9, or 12 items (set size). The search array items were randomly scattered around the center red circle, inside

Figure 2 . Procedures of a typical slanted view memory search trial in Experiment 1.

which the target might be visible in half of the trials (the reference search trials). The search array remained onscreen until a response was given. Participants used a standard QWERTY keyboard to make responses. They pressed the "F" key to report the presence and pressed the "J" key to report the absence of a target in the search array. Afterward, participants rated the confidence level of their previous search on a 4-point scale by pressing the "R," "T." "Y," or "U" key (R for least confident and U for most confident). After confidence rating, the fixation cross reappeared, and the next search trial began. See Figure 2.

There were altogether four experimental blocks for the two factors of stimulus type (frontal versus slanted views) and search type (memory versus reference search) and the order of the blocks were counterbalanced between subjects. In each block, we tested four set sizes crossed with target presence/absence (the probability of each was 50%), yielding eight unique conditions and each condition was repeated four times. It took approximately 15 minutes to complete one block. Altogether, each participant spent about 1 hour to complete 128 trials in four experimental blocks.

Sixteen practice trials were given to each participant before the actual experiment. With the presence of the experimenter, participants in the practice trials searched for LEGO blocks from a small search array (3 items). Procedures in the practice trials were identical to those in the actual experiment. Stimuli used in the practice trials never appeared again in the actual experiment.

Data analysis

This experiment used a 2 (stimulus type, blocked) × 2 (search type, blocked) × 2 (target present/

absent) × 4 (set size) within-subject factorial design. We collected data on search accuracy (defined as percentage of correct trials over total trials, with 50% being chance-level performance), response time (RT), search efficiency (defined as the slope of the fitted least-squared line between RT and set size; Duncan & Humphreys, 1989), subjective confidence ratings, and eye movement patterns.

In this and the second experiment, to compare behaviors in different search conditions, we conducted Bayesian analyses (Kruschke, 2010) using the JASP software (JASP Team, 2020 [https://jasp-stats.org/]; Wagenmakers et al., 2018). Statistical evidence for ANOVA was reported using BF_{incl}, the value of which indicated strength of evidence for including a specific variable in the ANOVA/regression model and the larger the BF_{incl} value, the stronger the evidence (van den Bergh, Van Doorn, Marsman, Draws, Van Kesteren, & Derks, 2020). As a general rule, $BF_{incl} > 1$ suggested to include a particular factor; $BF_{incl} < 1$ suggested not to include a particular factor (van Doorn, et al., 2019). Statistical evidence for post hoc pairwise comparisons was reported using BF_{10}, which indicated the strength of evidence for H_1 and the larger the BF_{10} value, the stronger the evidence. As a general rule, $BF_{10} > 1$ suggested strong evidence for H_1, and $BF_{10} < 1$ suggested weak evidence for H_1 but strong evidence for H_0 (van Doorn, et al., 2019). Other data treatments, such as fitting regression lines and computing means and 95% confidence intervals (CIs), were computed in R (R version 4.0.5, RStudio Team, 2021). Statistically significant main and interactive effects were reported. Occasionally, informative nonsignificant effects were mentioned and discussed because they suggested lack of difference between the conditions being contrasted.

Eye movement data were parsed into saccades and fixations using the Eyelink Data Viewer (Data Viewer 3.2.48, SR Research, 2018). Eye movements that had an acceleration threshold of 8000 degrees/second^2, a velocity threshold of 30 degrees/second, and deflections>0.1 degrees were classified as saccades. All fixation points that were within 0.6 degrees of each other and within a 120 ms temporal window of each other were counted as a single fixation. Oculomotor variables, including the number of fixations, the mean fixation duration, and the mean saccadic amplitude were analyzed to study the search behavior (Zelinsky & Sheinberg, 1997). Blinks were removed.

The regions of interest (ROI) were entailed with each LEGO block on the screen and an area of 2 degrees diameter around it (Spotorno, Malcolm, & Tatler, 2014). In the reference search blocks, we outlined a reference item's ROI and then calculated the total duration of fixations that fell within this ROI as the dwell time on the reference item. Similarly, in

target-present trials, we outlined a target's ROI and the target was considered to be fixated when at least one fixation fell within its ROI. Two ROI-related measures were used, including the proportion of trials where participants' fixations landed on the target over all target-present trials (*p*), the proportion of trials where targets were identified over trials where targets were fixated upon (p_{id}). In the literatures, *p* was treated as measures of selection; p_{id} was measures of identification (Godwin, Menneer, Liversedge, Cave, Hollimane, & Donnelly, 2020).

Results

We individually compared search accuracy, RT, self-rated confidence, and eye movement patterns between the two stimulus types and two search types with various set sizes and target presence/absence and found participants' performance was generally similar when search from frontal versus slanted views, with differences in eye movement patterns. Comparing between the search types, participants completed the reference search with longer search time, higher accuracy, and higher level of confidence than the memory search. Next, we examined each aspect of the search behaviors in details.

Accuracy

In all conditions, participants were able to search with an overall accuracy rate between 84% and 96%. We performed a Bayesian repeated measures ANOVA analysis, with stimulus type, search type, targets' presence, and set size as the within-subject variables. First, accuracy was not affected by stimulus type ($BF_{incl} = 0.084$). When search items were objects' frontal views, the mean accuracy was 93.0% (95% within-subject CI = 90.4%, 95.5%, calculated using Morey's 2008 method); when the search items were 2D projections of objects from 35 degrees viewing angles, the mean accuracy was 90.4% (95% within-subject CI = 87.8%, 93.0%). Second, search accuracy was affected by search type ($BF_{incl} = 36.82$). When the target was visible throughout search, the mean accuracy was 94.6% (95% within-subject CI = 92.4%, 96.8%); in trials with no target displayed in the middle of the search array, the mean accuracy was 88.8% (95% within-subject CI = 85.3%, 92.3%). Third, target presence/absence did not affect search accuracy ($BF_{incl} = 0.083$; target presence: mean = 90.7%, 95% within-subject CI = 88.6%, 92.8%; target absence: mean = 92.8%, 95% within-subject CI = 90.7%, 94.9%). Finally, set size affected accuracy ($BF_{incl} = 1553.21$), and as the set size increased, the accuracy dropped. Post hoc planned comparisons with Bonferroni correction showed that the accuracy was

higher when looking for a target from three items (95.8%) or six items (95.8%) than from nine or 12 items (87.4% and 87.9%, respectively; BF_{10} ranged from 8.03 to 225.63).

Response time and search efficiency

Analyzing response time in the correct trials with Bayesian repeated measures ANOVA, we found that search RT was affected by search type ($BF_{incl} = 6300.42$), target presence/absence ($BF_{incl} = $ infinite), set size ($BF_{incl} = $ infinite), and the interaction between set size and target presence/absence ($BF_{incl} = 3.80 \times e^{11}$); it was not affected by stimulus type ($BF_{incl} = 0.027$).

Looking at the individual effects, first, search RT was longer in reference search trials (mean = 3.82 seconds, 95% within-subject CI = 3.44, 4.21) than in the memory search trials (mean = 3.37 seconds, 95% within-subject CI = 2.85, 3.89). Second, the search RT increased with set size (set size 3: mean = 1.79 seconds, 95% within-subject CI = 1.51, 2.07; set size 6: mean = 3.04 seconds, 95% within-subject CI = 2.68, 3.40; set size 9: mean = 4.16 seconds, 95% within-subject CI = 3.66, 4.67; and set size 12: mean = 5.39 seconds, 95% within-subject CI = 4.65, 6.14). Post hoc analysis showed that all the pairwise comparisons reached significance (all $BF_{10} > 100$). Third, consistent with findings in the classic visual search literature, search RT was longer in target absent trials (mean = 4.30 seconds, 95% within-subject CI = 3.73, 4.87) than in the target present trials (mean = 2.90 seconds, 95% within-subject CI = 2.53, 3.27). Finally, RT was affected by the interaction between set size and target presence/absence (Figure 3). As the set size became larger, the RT became increasingly longer in the target absent search than in the target present search. In the target present trials, the search RT increased at the rate of 0.29 second/item. The RT – set size relation was linear (RT = 0.29 × set size + 0.74, $r^2 = 0.64$, $F(1, 50) = 88.96$, $p < 0.001$) and monotonically increasing, with slope significantly different from 0 ($BF_{10} = 56,721.50$). In the target absent trials, the search RT increased at the rate of 0.51 second/item. The RT – set size relation was linear (RT = 0.51 × set size + 0.50, $r^2 = 0.72$, $F(1, 50) = 131.5$, $p < 0.001$) and monotonically increasing, with slope significantly different from 0 ($BF_{10} = 467,518.16$).

In the classic visual search literature, it has been repeatedly suggested that efficiency, which was defined as the search time increment per additional search array item, of the target present search was twice as high as that of the target absent search or the increment was half in target present search (Treisman & Gelade, 1980; Wolfe, 1998). In each of the eight conditions in this experiment (target present/absent, stimulus type, and search type) and for every participant, we fit a linear trend between the search RT and set size and analyzed the slopes (which reflected search efficiency)
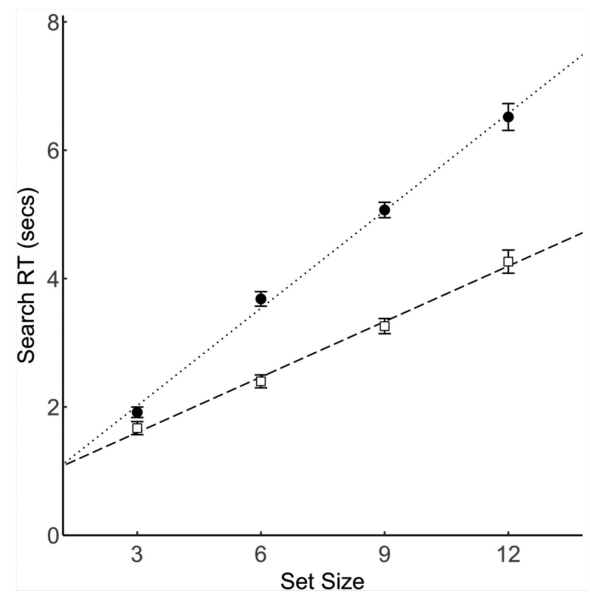

Figure 3. Behavioral results of Experiment 1. RT was affected by target presence/absence and set size. Filled circle = target absent; and open square = target present. Error bars (some were small and occluded by the markers) represent 1 *SE*.

with Bayesian repeated-measures ANOVA. Results showed that search efficiency was only affected by target presence/absence ($BF_{incl} = 2.08 \times e^{10}$). As expected, the mean slope in the target absent trials (mean = 0.53 second/item, 95% within-subject CI = 0.45, 0.60) was equivalent to two times the mean slope of the target present trials (mean = 0.28 second/item, 95% within-subject CI = 0.46, 0.70, $BF_{10} = 0.62$); or the search efficiency was half in the target absent trials. Whether the search items were shown as frontal view or slanted view did not affect the search efficiency ($BF_{incl} = 0.21$; $mean_{frontal} = 0.38$ second/item, 95% within-subject CI = 0.32, 0.45; $mean_{slanted} = 0.41$ second/item, 95% within-subject CI = 0.35, 0.47). The presence of the reference targets did not affect the search efficiency as well ($BF_{incl} = 0.44$; $mean_{memory} = 0.38$ second/item, 95% within-subject CI = 0.31, 0.44; $mean_{reference} = 0.42$ second/item, 95% within-subject CI = 0.35, 0.48).

Confidence

We conducted Bayesian repeated-measures ANOVA on confidence rating with stimulus type, search type, targets' presence, and set size as the within-subject factors. Confidence rating was affected by search type ($BF_{incl} = 4.46 \times e^8$) and set size ($BF_{incl} = 4.66$). Whether the search items were shown as the frontal view or the slanted view did not affect subjective confidence ($BF_{incl} = 0.034$). Participants were more confident in the reference search trials (mean rating = 3.94, 95% within-subject CI = 3.90, 3.98), than in the memory
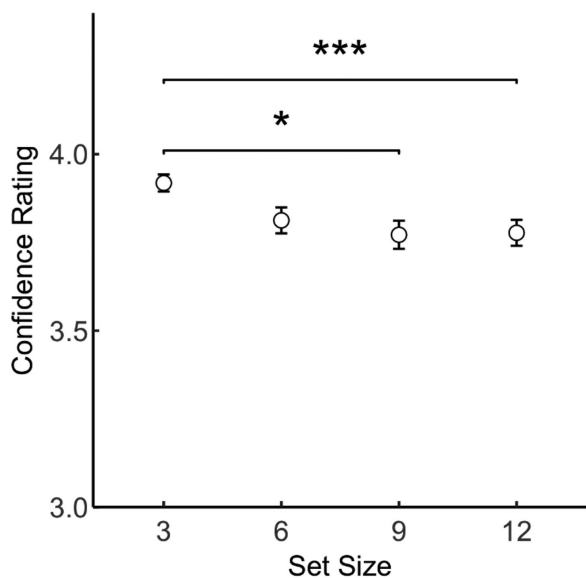
Figure 4. Behavioral results of Experiment 1. Confidence rating decreased with increasing set size. Error bars (some were small and occluded by the markers) represent 1 *SE*.

Predictor	B	SE	Wald	Sig.	Odds ratio
Stimulus type (frontal)	0.375	0.197	3.632	0.057	1.454
Search type (reference)	0.482	0.215	5.039	0.025	1.620
Target pre/abs (absent)	0.480	0.199	5.799	0.016	1.617
Set size			25.817	0.000	
Set size 3	1.092	0.321	11.551	0.001	2.979
Set size 6	1.068	0.306	12.214	0.000	2.910
Set size 9	−0.084	0.227	0.136	0.712	0.920
Confidence			65.811	0.000	
Confidence 2	2.531	0.757	11.184	0.001	12.565
Confidence 3	3.199	0.714	20.067	0.000	24.517
Confidence 4	4.265	0.679	39.491	0.000	71.189

Table 1. Results of logistic regression predicting accuracy from stimulus type, search type, target presence/absence, set size, and confidence rating.

search trials (mean rating = 3.72, 95% within-subject CI = 3.67, 3.76). Furthermore, as the number of search items increased, the confidence rating scores dropped. The differences between set size 3 (3.92) and 9 (3.77), and 3 and 12 (3.78) reached significance (BF_{10} = 12.31,>100, respectively; Figure 4).

Comparing the subjective confidence to the search response, although in the correct search trials, the confidence scores had higher mean and lower variance than in the incorrect trials (mean$_{correct}$ = 3.87, $SD_{correct}$ = 0.41; mean$_{incorrect}$ = 3.32, $SD_{incorrect}$ = 0.97), the mean confidence score in the incorrect trials was still significantly higher than 2.5, the mid-point of the 4-point scale (BF_{10} = 85.70). The participants were overly confident.

A binary logistic regression was performed to assess how confidence, stimulus type, search type, targets' presence, and set size predicted search outcomes (correct or incorrect). The results showed that the full logistic regression model containing all the five predictors was statistically significant, χ^2 = 133.12, df = 10, N = 1599, p<0.001, in other words, all five input factors were significant in predicting the search correctness (Table 1). The predictor confidence had an odds ratio of 71.19, which meant that everything else being equal, an individual who reported the highest confidence (confidence = 4) on a particular trial was 71.19 times more likely to be correct in their search than someone who rated the lowest confidence level (confidence = 1). The odds ratio for stimulus type was 1.45 or search with the frontal views was 1.45 times more likely to be correct than search with the slanted views of object. The odds ratio for search type was 1.62

or reference search was 1.62 times more likely to be correct than memory search. The odds ratio for target presence/absence was 1.62 or search in the target absent trials was 1.62 times more likely to be correct than search in the target present trials. Last, when searching from three or six items, the likelihood of being correct was approximately three times higher than searching from nine or 12 items.

Overall, the analyses of search performance indicated that participants searched faster but made more errors in conditions when they relied on memory, as compared to when they were able to use external reference. Their search performance (accuracy, RT, and confidence) was largely equivalent regardless of whether search items were presented with a frontal view or a slanted view. In particular, we compared performances between searches with the frontal views and no reference objects (frontal-memory condition), which was a typical laboratory-based visual search task setup and search with slanted views and reference objects (slanted-reference condition), which was designed to be representative of real-world search. First, the accuracy was equivalent between these conditions (BF_{10} = 0.46; the frontal-memory condition: mean = 91.1%, 95% within-subject CI = 87.8%, 94.3%; the slanted-reference condition: mean = 93.8%, 95% within-subject CI = 90.1%, 97.5%). However, the slanted-reference condition had longer search RT (BF_{10} = 7.67; mean$_{frontal-memory}$ = 3.41 seconds, 95% within-subject CI = 2.92, 3.90; mean$_{slanted-reference}$ = 3.80 seconds, 95% within-subject CI = 3.31, 4.29) and higher confidence rating (BF_{10} = 103.89; mean$_{frontal-memory}$ = 3.74, 95% within-subject CI = 3.62, 3.86; mean$_{slanted-reference}$ = 3.92, 95% within-subject CI = 3.86, 3.97). Because the search performance, in terms of search RT, search efficiency, and confidence rating, was equivalent between the frontal-reference condition and the slanted-reference condition (search RT: BF_{10}

= 0.33; search efficiency: $BF_{10} = 0.28$; and confidence rating: $BF_{10} = 0.36$), we can infer the longer RT and higher reported confidence level in the slanted-reference condition was due to the presence of reference items, not the perspective projection. In other words, in a more natural visual search task with perspective change, observers used the external reference to achieve accuracy as high as in a classic visual search task with exemplar 2D views.

Next, we examined the effects of search type, stimulus type, set size, and target presence/absence on eye movement patterns, including the number of fixations, the fixation duration, and the saccade amplitude.

Fixations and saccades

To compare the number of fixations between the memory and reference searches and between the frontal and slanted views, we conducted a Bayesian 2-way repeated-measures ANOVA and the results showed that there were more fixations in the frontal view than in the slanted view condition ($BF_{incl} = 24.73$; mean$_{frontal} = 14.82$, 95% within-subject CI = 13.98, 15.66; mean$_{slanted} = 14.05$, 95% within-subject CI = 13.16, 14.94). There were more fixations in reference search than in memory search ($BF_{incl} = 1191.85$; mean$_{reference} = 15.82$, 95% within-subject CI = 14.91, 16.73; mean$_{memory} = 12.96$, 95% within-subject CI = 12.17, 13.75). It is reasonable to attribute that the extra fixations in the reference search was due to repeatedly checking the reference item. To verify this speculation, we subtracted fixations on the reference item from the total number of fixations and then performed a Bayesian paired sample *t*-test comparing the number of fixations in the memory search with that in the subtracted reference search. The results showed that the remaining fixation counts (Figure 5) in the reference search (mean = 13.07, 95% within-subject CI = 12.32, 13.82) and in the memory search (mean = 12.87, 95% within-subject CI = 12.12, 13.62) were equivalent ($BF_{10} = 0.33$), which confirmed that participants repeatedly checked the reference item while searching. Furthermore, we ran a Bayesian 1-way repeated-measures ANOVA with the number

Figure 5. The fixation counts make-up in the reference and memory search conditions of Experiment 1.

of fixations on the reference target as the dependent variable and set size as the independent variable. Results showed that when the number of search items increased, the number of fixations on the reference target increased ($BF_{incl} = 16,241.69$; set size 3: mean = 1.91, 95% within-subject CI = 1.39, 2.42; set size 6: mean = 2.46, 95% within-subject CI = 1.84, 3.07; set size 9: mean = 2.78, 95% within-subject CI = 1.99, 3.56; set size 12: mean = 3.64, 95% within-subject CI = 2.60, 4.69). The differences between set sizes 3 and 6 ($BF_{10} = 5.95$), 3 and 9 ($BF_{10} = 47.71$), 3 and 12 ($BF_{10} = 214.47$), 6 and 12 ($BF_{10} = 21.54$), and 9 and 12 ($BF_{10} = 4.41$) all reached significance. This suggested that as the task became more demanding, the observers were more willing to seek aid from the reference target in the environment.

Next, we examined the effects of stimulus type and search type on fixation durations and saccade amplitude. Bayesian repeated-measures ANOVA analysis revealed shorter fixation durations in the frontal view condition than in the slanted view condition ($BF_{incl} = 23300.59$; mean$_{frontal} = 0.23$ seconds, 95% within-subject CI = 0.22, 0.24; mean$_{slanted} = 0.26$ seconds, 95% within-subject CI = 0.25, 0.27) and shorter fixation durations in reference search than in memory search ($BF_{incl} = 1,096.89$; mean$_{reference} = 0.24$ seconds, 95% within-subject CI = 0.23, 0.25; mean$_{memory} = 0.26$ seconds, 95% within-subject CI = 0.25, 0.27). Taking out fixations on the reference targets, we compared the remaining fixation durations on the memory and the reference search trials with a Bayesian paired samples *t*-test. Results showed that fixation durations on the search array items in the memory search and the reference search were equivalent ($BF_{10} = 0.47$; mean$_{memory} = 0.26$ seconds, 95% within-subject CI = 0.23, 0.28; mean$_{reference\_remain} = 0.25$ seconds, 95% within-subject CI = 0.22, 0.28). Therefore, the existence of the reference targets did not speed up the object identification process. Furthermore, Bayesian repeated measures ANOVA showed that the mean saccade amplitude was larger in the frontal view condition than in the slanted view condition ($BF_{incl} = 1.90 \times e^9$; mean$_{frontal} = 3.82$ degrees, 95% within-subject CI = 3.71, 3.93; mean$_{slanted} = 3.06$ degrees, 95% within-subject CI = 2.95, 3.17). Mean saccade amplitudes were marginally different between the search types ($BF_{incl} = 1.93$; mean$_{reference} = 3.52$ degrees, 95% within-subject CI = 3.19, 3.84; mean$_{memory} = 3.37$ degrees, 95% within-subject CI = 3.07, 3.68).

The results of fixations and saccades implied that although the total search RTs were similar between the frontal and the slanted view conditions, the search subprocesses differed between them. Namely, in the frontal view condition, observers made a greater number of quick fixations that covered the entire display; in the slanted view condition, observers made fewer fixations with smaller amplitudes (less "jumping
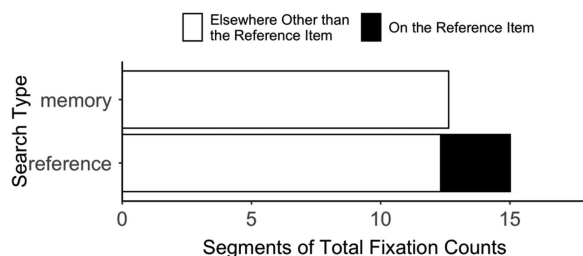
around"). This could be because of the objects in the slanted view condition were compressed along the vertical dimension, making the search array more compact than that of the frontal view condition. It could be inferred that observers were more scrupulous when searching in the slanted view condition but hastier when searching in the frontal view condition. When searching with references, observers checked back and forth between the reference item and the search array items, which reflected their strategy of leveraging on external landmarks and minimizing the use of internal memory. In other words, observers picked up information from the environment in real-time to accomplish the search task.

Perceptual selection and perceptual identification

The process of the visual search can be divided into two phases – the selection phase and the identification phases (Cain, Adamo, & Mitroff, 2013; Godwin, et al., 2020). The selection phase was defined as the process where target-like objects were selected for further determination. The relevant indicator for search accuracy was the proportion of trials where the target was ever fixated on over the total number of trials (p). In the identification phase, observers fixated on features of the target-like objects to determine whether they were indeed the target. The relevant measure for search accuracy was the proportion of trials where the targets were identified over the trials where targets were fixated upon (p_{id}). We tested how these two measures were affected by stimulus type and search type using separate Bayesian repeated-measures ANOVAs. First, the proportion of target fixation (p) was not affected by stimulus type ($BF_{incl} = 0.33$; $mean_{frontal} = 3.82$, 95% within-subject CI = 3.71, 3.93; $mean_{slanted} = 3.06$, 95% within-subject CI = 2.95, 3.17), search type ($BF_{incl} = 0.25$; $mean_{reference} = 0.80$, 95% within-subject CI = 0.75, 0.85; and $mean_{memory} = 0.82$, 95% within-subject CI = 0.76, 0.87), or their interaction ($BF_{incl} = 0.15$). Second, the proportion of identification, p_{id}, was marginally affected by search type ($BF_{incl} = 1.03$) and p_{id} was higher in the reference search (mean = 95.0%, 95% within-subject CI = 90.8%, 99.3%) than in the memory search (mean = 90.0%, 95% within-subject CI = 84.6%, 95.4%). Neither the stimulus type ($BF_{incl} = 0.28$; $mean_{frontal} = 0.93$, 95% within-subject CI = 0.88, 0.98; $mean_{slanted} = 0.92$, 95% within-subject CI = 0.87, 0.96) nor the stimulus type times the search type interaction ($BF_{incl} = 0.27$) affected p_{id}.

RT segmentation

The total search RT in a search trial was divided into four segments – looking at the target, looking at the distracters, looking at the reference item, and looking elsewhere on the display. The proportion of each segment over the total looking time was computed.
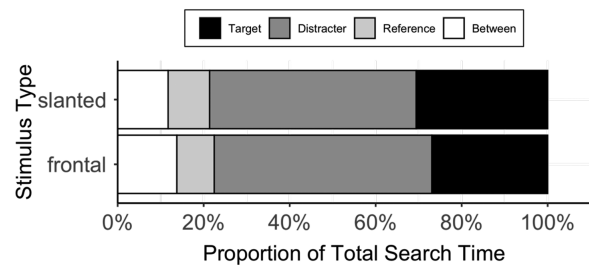

Figure 6. The proportions of looking time at the target, at the distracters, at the reference item, and at elsewhere on screen in the frontal/slanted view conditions.

Although the total search RT was not different between stimulus types, we further examined the proportion of each looking segment in the frontal/slanted view conditions. With the frontal or slanted views, the proportions of looking at the distracters ($BF_{10} = 0.28$), at the reference item ($BF_{10} = 0.52$), or at elsewhere ($BF_{10} = 0.82$) were equivalent (Figure 6). However, searchers took longer to verify the target in the slanted view condition (mean = 0.82 seconds, 95% within-subject CI = 0.57, 1.07) than in the frontal view condition (mean = 0.70 seconds, 95% within-subject CI = 0.49, 0.91), $BF_{10} = 4.08$.

Taken together, these results suggested that (1) although the total search time was similar between the frontal view and the slanted view conditions, it took observers more time to verify that a target-like object was the target in the slanted view condition, and (2) the proportion of identifying a target after spotting it was higher and the identification took less time in reference search.

Discussion

In Experiment 1, we primarily studied the effects of stimulus type, search type, target presence/absence, and set size on search accuracy, search RT and efficiency, subjective confidence about search performance, and eye movements during the search. We manipulated the viewing perspectives to simulate a more general circumstance where observers were looking with a non-zero angle of declination and searching from non-exemplar views of objects. We were particularly interested in the degree to which their search performance was affected by image distortions resulting from the perspective change. Visual search outcomes, such as accuracy, search RT, efficiency, and self-reported confidence level, were not affected by perspective manipulation. However, the process of searching was different with the slanted view versus with the frontal view, in that observers made fewer but longer fixations and smaller saccades when searching, and spent more time verifying a target after spotting it in the slanted view condition.

We introduced a reference item that remained visible throughout the search in some trials to create an embodied situation of the reference search and to contrast with the traditional memory search. The reference search allowed observers to exploit the stable environment and acquire task-relevant information in real-time. The memory search required observers to remember the target and use their memory in the traditional sense to search. Results showed that in the reference search, participants were more accurate but slower in the reference search because they looked back and forth between the reference item and the search array items before making a response. Perhaps because of that, they were more confident about their search responses. This was reflected in the eye movement patterns that in the reference search, participants made more fixations with shorter durations. With the reference item, participants also had a higher chance of correctly identifying the target after spotting it and needed less time to do so.

Experiment 2

In this experiment, we introduced motion to a visual search task and generated a dynamic search in VR. In classic visual search experiments, the search items were typically static. In this experiment, we continued to display the static condition. We also added a dynamic condition, where all search items were placed on a turntable, and the turntable rotated in depth around the vertical axis. The dynamic search simulated a realistic search scenario where an observer was moving around the search items and looking for a particular one. If the visual search was based on comparisons of projected images of targets and search array items, then motion was a perturbation and the motion-introduced image transformation would undermine search accuracy and/or efficiency. If the visual search was based on recognizing 3D objects, then the vast literatures on "structure from motion" (Norman & Todd, 1993; Tittle & Braunstein, 1993; Bingham & Lind, 2008; Lee, Lind, Bingham, & Bingham, 2012; Lind, Lee, Mazanowski, Kountouriotis, & Bingham, 2014); suggested that continuous motion would not hinder recognition and therefore not affect search performance. In this experiment, we compared visual search performance when the search array items rotated (dynamic search) versus when they remained still (static search) and studied whether the benefit of having external reference items extended to aiding the dynamic search.

Methods

Participants

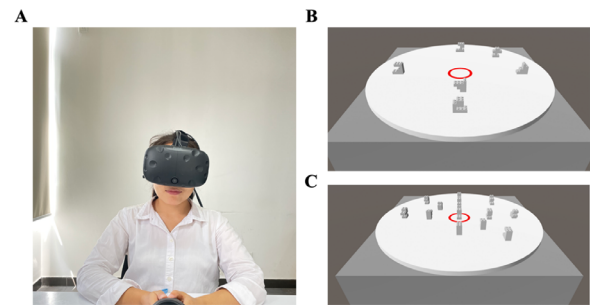To determine the appropriate sample size, we used G∗Power (Faul, Erdfelder, Lang, & Buchner, 2007)

Figure 7. Stimuli and apparatus in experiment 2. (**A**) A participant performed visual search in VR wearing an HMD and holding a controller to make response. (**B**) Experimental display in the static and frontal-dynamic conditions. In the frontal-dynamic condition, rotation of the turntable always began with the frontal view of the search items facing an observer. (**C**) Experimental display in the sideview-dynamic condition (the control condition), where rotation always began with the side view of the search items facing an observer.

with four within-subject factors and set the effect size at 0.25 and the alpha value at 0.05. Results indicated that a sample size of 10 would produce a power of 0.95. We recruited 20 adults (aged between 18 and 35 years, 11 women) in this experiment. All participants had normal or corrected-to-normal vision and received ¥30/hour to compensate for their time and effort. Five participant's data were removed from analysis due to technical errors in eye movement recording. The henceforth analysis was based on 15 participants' data. This study was approved by the Institutional Review Board of the Department of Psychology, Sun Yat-sen University. Informed consent was obtained from all participants.

Stimuli and apparatus

Participants wore an HTC Vive head-mounted display (HMD; New Taipei City, Taiwan) equipped with a Qingtech eye tracker (version V1S; Shanghai Qingyan Technology Co., Ltd., Shanghai, China) and held an HTC Vive controller in their dominant hand (Figure 7A). The two 1080×1200 px OLED screens had a refresh rate of 90 Hz and a combined field of view of approximately 100 degrees (horizontal) × 110 degrees (vertical). The integrated Qingtech eye tracker (version V1S; Shanghai Qingyan Technology Co., Ltd.) recorded the eye movements binocularly with a refresh rate of 100 Hz and a spatial resolution around 0.5 degrees within a 20 degrees window centered in the viewports. The experiment was implemented in C# in the Unity 3D game engine (version 2018.4.14; Unity Technologies, San Francisco, CA, USA) using SteamVR (version 1.16.10; Valve Corporation, Bellevue, WA, USA), and Qingtech eye-tracking software libraries (version 2.1.82; Shanghai Qingyan Technology Co., Ltd.) on a computer operated with Windows 10.

The search items were similar those in Experiment 1 but instead of actual LEGO blocks, they were modeled using SketchUp Pro (version 18.0.16976; Trimble Inc., Sunnyvale, CA, USA). The display size for each LEGO block was 2 degrees × 2 degrees, which was the same as that in Experiment 1 (Figures 7B, 7C). Created in VR, all search items were placed on a round turntable and they did not occlude one another. In the center of the turntable, there was a red ring (diameter = 4 degrees), inside which there was the reference item on some but not all trials. The graphics simulated a 35 degrees viewing angle, as if a seated observer was looking down at the search items on a table. In the static display condition, the turntable and the search items were stationary; in the dynamic display condition, the turntable, with the search items it supported, rotated in depth. The turntable began rotating with the frontal views search items directly facing the observer and it rotated with the speed of 12 degrees/seconds and the maximum displacement of 45 degrees to the left and to the right of the starting position (see Figure 7B).

Procedures

Participants came to the laboratory and were explained about the purpose and procedures of the experiment. They then signed the consent forms, sat down, and put on the HMD. Head movements were not restricted. A five-point calibration was performed to validate the eye tracker. Participants received verbal instructions that they were going to see one LEGO block first, and then report the presence or absence of the LEGO target from many LEGO blocks. The procedures were similar to those in Experiment 1 with exceptions of how responses were made. In this experiment, participants reported the presence of a target by pulling the trigger on the controller with the index finger and reported the absence of a target by pressing the trackpad on the controller with the thumb. Participants selected a confidence level by pressing the left and right button on the controller.

There were altogether four experimental blocks for the two factors of display type (static and dynamic) and search type (memory search and reference search) and the order of experimental blocks was counterbalanced between subjects. Within each block, there were eight conditions (4 set sizes × target-present/absent) and each condition was repeated four times. It took approximately 15 minutes to complete the 32 trials in one block. Altogether, each participant completed four blocks of 128 trials.

The control condition

In the dynamic display condition, the starting view might affect search performance because distinguishing features of the LEGO blocks were only found in two dimensions and rotation began with the frontal view of the LEGO blocks directly facing the observer. In other words, the distinguishing features of search items were utterly presented when rotation was about to start. We add a control condition in which rotation began when the sideview of the LEGO blocks were facing the observer (see Figure 7C). In this case, no distinguishing features of search items was presented when rotation began but distinguishing features gradually came to view as the rotation went on. In the control condition, rotation speed was also 12 degrees/second but the turntable only rotated to one direction for 120 degrees and then turned back. The 120 degrees rotation allowed both sides of the objects to be seen by the observers. We hence referred the control condition as the "sideview-dynamic condition," in contrast to the "frontal-dynamic condition." Another group of 16 participants (aged between 18 and 33 years, 8 women) did the sideview-dynamic condition, which consisted of one memory search block and one reference search block (block order counterbalanced). Within each block, there were eight conditions (4 set sizes × target-present/absent) and each condition was repeated four times.

Data analysis

The main experiment consisted of a 2 (display type, blocked) × 2 (search type, blocked) × 2 (targets' presence) × 4 (set size) within-subject factorial design. We analyzed the outcome measures of search accuracy, response time, search efficiency, confidence rating, and eye movement patterns. Same as in Experiment 1, the means were compared using Bayesian methods conducted in the JASP software and regressions and other analysis were conducted in R software. Raw eye tracking data were preprocesses in an in-house program developed by Qingtech to generate object-based outcome. The eye tracker mounted in the helmet tracked the observer's gaze patterns in real-time, and the VR processor mapped the gazes to the VR scene. The duration of gazes dwelled on an object was the target identification time. The eye tracking program recorded the identification time on each object throughout the search trials.

Results

Three lines of analyses were conducted for three goals – comparing search performance between the slanted condition in Experiment 1 and the static condition in Experiment 2 to validate the VR setup; comparing the search performance between the static and the frontal-dynamic conditions to test for the effect of motion and continuous perspective change on visual search; and examining the eye movement patterns in

the static and the frontal-dynamic conditions to explore how motion affected participants' search process in the selection and identification phases.

Validating the VR setup

To validate the experimental setup in VR, we first compared performance of the static condition in Experiment 2 with that of the slanted condition of Experiment 1. The only differences between the two experiments were the participants and whether the display was in VR with free head movement or onscreen with their head on a chinrest. All other settings were identical. Bayesian repeated-measures ANOVA with experiment (screen versus VR) as the between-subject factor and search type (memory search and reference search) as the within-subject factor were conducted to individually assess search accuracy, search RT, and self-reported confidence. Accuracy was not different between the screen and the VR display conditions ($BF_{incl} = 0.56$; screen: mean = 90.4%, 95% within-subject CI = 87.1%, 93.6%; VR: mean = 93.0%, 95% within-subject CI = 91.3%, 95.7%). Search RT was not different between screen and VR conditions ($BF_{incl} = 0.24$; screen: mean = 3.54 seconds, 95% within-subject CI = 2.93, 4.14); VR: mean = 4.24 seconds; 95% within-subject CI = 3.68, 4.81). The mean confidence rating was not different between the screen and VR conditions ($BF_{incl} = 0.43$; screen: mean = 3.81; 95% within-subject CI = 3.73, 3.88; VR: mean = 3.87, 95% within-subject CI = 3.80, 3.94). Accuracy, search RT, and confidence rating were not affected by any interactive effects. Furthermore, Bayesian repeated-measures ANOVA with experiment (screen versus VR) as the between-subject factor and the search type (memory versus reference search) and target presence/absence as the within-subject factors were conducted to individually assess search efficiency. Results showed that search efficiency was not different between screen and VR conditions ($BF_{incl} = 0.24$; screen: mean = 0.41 seconds/item, 95% within-subject CI = 0.34, 0.48); VR: mean = 0.40 seconds/item; 95% within-subject CI = 0.34, 0.47). Search efficiency was not affected by any interactive effects. Thus, with the current setup, the experiment conducted in VR produced congruous results with the screen-based experiment.

Comparing visual search performance with and without motion

After validating the VR setup, we introduced the search array rotations and examined the search accuracy, RT, efficiency, self-rated confidence, and eye movement patterns in the static/dynamic search and in the memory/reference search with various set sizes when targets were present or absent. Results of these analyses directly answered the questions of whether continuous perspective change during visual search affects the performance and whether having references facilitated visual search with motion.

First, we examined search accuracy using a Bayesian repeated-measures ANOVA with display type, search type, target presence/absence, and set size as the within-subject factors. First, the search accuracy was not affected by display type ($BF_{incl} = 0.14$). Mean accuracy when the search array was static (mean = 93.4%, 95% within-subject CI = 91.5%, 95.4%), was similar to that when the search array was rotating (mean = 90.9%, 95% within-subject CI = 88.4%, 93.3%). Consistent with the results in Experiment 1, when the target was visible in the middle of the search array throughout search, the accuracy was higher than otherwise ($BF_{incl} = 676.89$; reference search: mean = 95.1%, 95% within-subject CI = 93.1%, 97.1%; memory search: mean = 89.2%, 95% within-subject CI = 86.3%, 92.2%). Finally, as expected, accuracy declined as the set size increased ($BF_{incl} = 12.18$) and the declination was linear at the rate of 0.81% per additional item ($F(1, 58) = 11.25$, $p = 0.001$, slope>0 with $BF_{10} = 1.63$). Post hoc planned comparisons with Bonferroni correction showed that the accuracy was higher when looking for a target from three (95.5%, $BF_{10} = 93.65$) or six (93.4%, $BF_{10} = 232.97$) items than from 12 items (87.9%). Accuracy was not affected by the target presence/absence ($BF_{incl} = 0.79$; target presence: mean = 90.4%, 95% within-subject CI = 86.4%, 95.4%; target absence: mean = 94.0%, 95% within-subject CI = 91.3%, 96.2%), neither were there significant interactive effects.

Next, the effects of display type, search type, target presence/absence, and set size on the search RT in the correct trials were analyzed using Bayesian repeated measures ANOVA. The search RT was affected by target presence, set size, and the interaction between target presence and set size, and all three BF_{10} were greater than 100. The search RT was shorter in target present trials than in target absent trials (mean$_{present}$ = 3.69 s, 95% within-subject CI = 3.03, 4.35; mean$_{absent}$ = 5.06 seconds, 95% within-subject CI = 4.34, 5.79). The search RT increased with set size (set size 3: mean = 2.46 seconds, 95% within-subject CI = 2.31, 2.61; set size 6: mean = 3.73 seconds, 95% within-subject CI = 3.52, 3.94; set size 9: mean = 5.05 seconds, 95% within-subject CI = 4.76, 5.34; set size 12: mean = 6.28 seconds, 95% within-subject CI = 5.87, 6.69). Post hoc analysis showed that all the pairwise comparisons reached significance (all BF_{10}>100). Finally, as shown in Figure 8, the slope in target absent trials was larger than that in the target present trials. Specifically, in the target present trials, the search RT increased at the rate of 0.30 seconds/item. The RT – set size relation was linear ($RT = 0.30 \times set\ size + 1.48$, $r^2 = 0.39$, $F(1, 58) = 36.79$, $p < 0.001$) and monotonically increasing,
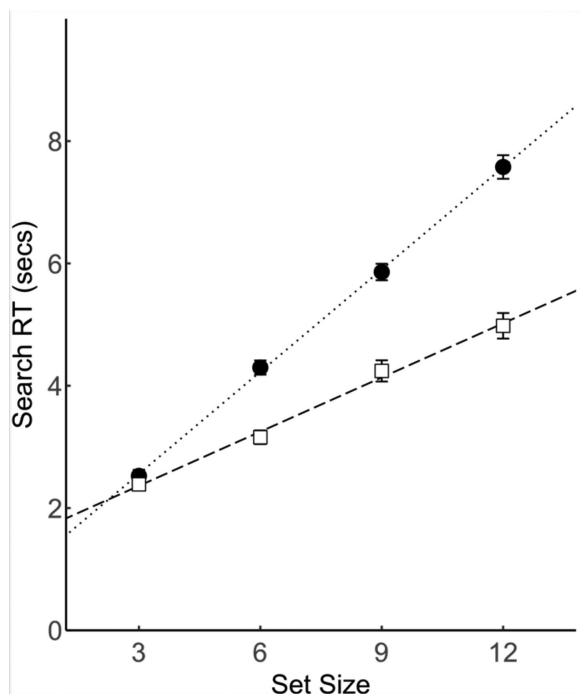
Figure 8. Behavioral results of Experiment 2. Search RT increased with set size, but at a different rate for the target present and target absent searches. Filled circle = target absent trials, and open square = the target present trials. Error bars represent 1 SE (some were too small to show).

with the slope significantly different from 0 (BF_{10} = 834,062.74). In the target absent trials, the search RT increased at the rate of 0.54 seconds/item. The RT – set size relation was linear (RT = 0.54 × set size + 0.90, r^2 = 0.65, $F(1, 50)$ = 107.3, $p < 0.001$) and monotonically increasing, with the slope significantly different from 0 (BF_{10} = 2.01 × e^7). The search RT was not affected by display type (BF_{incl} = 0.33) and the mean search RT when the search array was rotating was 4.50 seconds (95% within-subject CI = 4.18, 4.82) and the mean search RT when the search array was stationary was 4.26 seconds (95% within-subject CI = 3.97, 4.55). Furthermore, the search RT was equivalent for the memory search and the reference search (BF_{incl} = 0.05; $mean_{memory}$ = 4.32 seconds, 95% within-subject CI = 3.96, 4.68; $mean_{reference}$ = 4.44 seconds; 95% within-subject CI = 4.10, 4.78).

Next, we fit a linear trend between the search RT and the set size in each of the eight conditions (static/dynamic, memory/reference, and target present/absent) and for every participant and the slopes indicated search efficiency. A Bayesian repeated measure ANOVA with display type, search type, and targets' presence as within-subjects factors was conducted. Search efficiency was different between the target-present and the target-absent trials (BF_{incl} = 2.98 × e^{12}; target-present: mean = 0.89 seconds/item; 95%

within-subject CI = 0.72, 1.05; target-absent: mean = 1.68 seconds/item, 95% within-subject CI = 1.44, 1.92). A follow-up Bayesian paired t-test indicated that slopes in target absent trials was equal to twice the slopes in target present trials (BF_{10} = 0.36). There was an interaction effect of display type × search type (BF_{incl} = 3.00) on search efficiency. Post hoc pairwise comparison showed that in memory search blocks, dynamic search had a lower search efficiency (mean slope = 0.50 seconds/item; 95% within-subject CI = 0.41, 0.60) than the static search (mean slope = 0.38 seconds/item; 95% within-subject CI = 0.28, 0.48), BF_{10} = 2.30. But, in the reference search, efficiency did not differ between the static and dynamic display conditions (BF_{10} = 0.29; $mean_{static}$ = 0.42 seconds/item, 95% within-subject CI = 0.34, 0.51; $mean_{dynamic}$ = 0.40 seconds/item, 95% within-subject CI = 0.32, 0.49).

A four-way Bayesian repeated-measures analysis on subjectively reported confidence revealed that the confidence rating was affected by the search type (BF_{incl} = 7.21 × e^{11}), set size (BF_{incl} = 17,478.31), and their interaction (BF_{incl} = 38.88). Participants were more confident in the reference search (mean = 3.95, 95% within-subject CI = 3.91, 3.98) than in the memory search (mean = 3.77, 95% within-subject CI = 3.71, 3.82). Confidence dropped as the set size increased and the drop between three and 12 (BF_{10} = 132.69; set size 3: mean = 3.93, 95% within-subject CI = 3.89, 3.97; set size 12: mean = 3.77, 95% within-subject CI = 3.70, 3.84), and between six and 12 search items reached statistical significance (BF_{10} = 30.29; set size 6: mean = 3.90, 95% within-subject CI = 3.85, 3.95). When searching from larger arrays (6 items or more), the confidence rating was higher in the reference search than in the memory search (set size = 6: BF_{10} = 21.24, set size = 9: BF_{10} = 28.00, set size = 12: BF_{10} = 31.02); when searching among three items, the confidence rating did not differ between search types (BF_{10} = 0.32; Figure 9). Importantly, confidence did not differ between the dynamic search and the static search (BF_{incl} = 0.42; dynamic search: mean = 3.84, 95% within-subject CI = 3.80, 3.89; static search: mean = 3.87, 95% within-subject CI = 3.83, 3.91).

Comparing the subjective confidence to the search response, although the confidence rating was higher and less variable in the correct trials ($mean_{correct}$ = 3.90, $SD_{correct}$ = 0.39; $mean_{incorrect}$ = 3.37, $SD_{incorrect}$ = 0.95), mean confidence score in the incorrect trials was still higher than 2.5, the mid-point of the 4-point scale (BF_{10} = 743.49). This reflected that participants were overly confident about their performance. Finally, the confidence rating was correlated with the search RT in the correct trials (Pearson correlation r = - 0.23, BF_{10} = 3.27 × e^{18}), and the more confident the participants felt, the less time it took for them to search.

A binary logistic regression was performed to assess how confidence, display type, search type, target
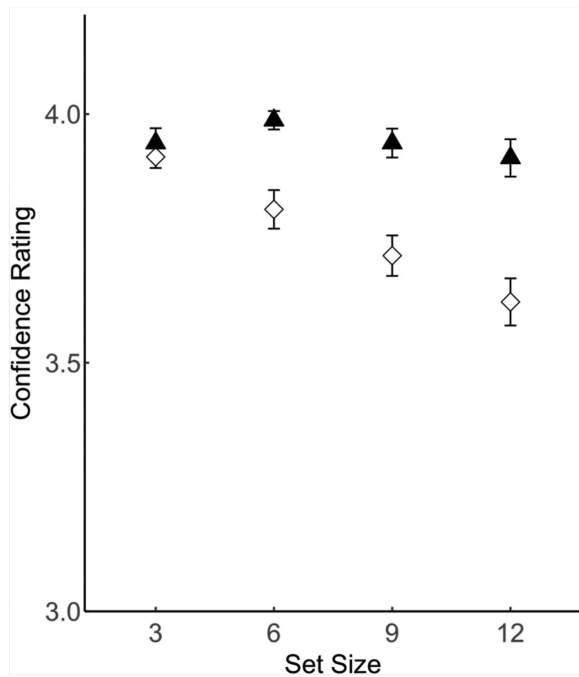
Figure 9. Behavioral results of Experiment 2. Confidence rating across the varied levels of search type and set size. Open diamond = memory search trials, and filled triangle = reference search trials. Error bars represent 1 *SE* (some were too small to show).

presence/absence, and set size predicted the likelihood of making a correct search response. The full logistic regression model was significant ($\chi^2 = 149.28$, *df* = 9, *N* = 2431, *p* < 0.001), which indicated that all these factors were involved in predicting search accuracy (Table 2). Confidence strongly predicted the search accuracy and the odds ratio of 19.21 indicated that all other predictors being equal, an individual who reported a level 4 confidence was 19.21 times more likely to be correct than someone who reported level 1 confidence on a particular trial. Furthermore, as shown

in Table 2, the smaller the set size, the more likely that the search was correct; a static display condition was 1.38 times more likely to be correct than a dynamic display trial; a reference search trial was 1.69 times more likely to be correct than a memory search trial; and, surprisingly, a target absent trial was 1.83 times more likely to be correct than a target present trial.

Control condition performance

We designed the control condition to contrast the effects of initial viewing perspectives on the search performance. Separate Bayesian repeated-measures ANOVAs with initial view condition (frontal-dynamic or sideview-dynamic) as the between-subject factor and search type (memory-search or reference-search) as the within-subject factor were conducted on outcome measures of accuracy, search RT, and self-reported confidence. First, accuracy was equivalent in the frontal-dynamic condition (mean = 91.1%, 95% within-subject CI = 88.3%, 93.9%) and the sideview-dynamic condition (mean = 90.7%, 95% within-subject CI = 87.2%, 94.3%), $BF_{10} = 0.57$. Second, there was strong evidence for longer search RT in the sideview-dynamic condition (mean = 7.10 seconds, 95% within-subject CI = 6.67, 7.54) than in the frontal-dynamic condition (mean = 4.43 seconds, 95% within-subject CI = 3.94, 4.91). This was predicted because the search items were differentiated by features in two dimensions and could not be distinguished solely based on the side views (see Figure 7C). So, in the sideview condition, participants had to wait for the search items to rotate before they could possibly search. Finally, the confidence rating was equivalent in the two view conditions ($BF_{10} = 0.39$; $mean_{frontal-dynamic} = 3.85$, 95% within-subject CI = 3.79, 3.90; $mean_{sideview-dynamic} = 3.81$, 95% within-subject CI = 3.71, 3.90). Hence, the initial orientation of the rotating search array, with search item's canonical view perpendicular (as in the frontal-dynamic condition) or parallel (as in the

Predictor	B	SE	Wald	*p* value	Odds ratio
Display type (static)	0.322	0.186	3.000	0.083	1.379
Search type (reference)	0.527	0.208	6.440	0.011	1.694
Targets' presence (absent)	0.604	0.190	10.076	0.002	1.829
Set size			10.801	0.013	
Set size 3	0.866	0.282	9.471	0.002	2.378
Set size 6	0.476	0.249	3.650	0.056	1.609
Set size 9	0.482	0.240	4.040	0.044	1.620
Confidence			81.414	0.000	
Confidence 2	0.694	0.585	1.408	0.235	2.001
Confidence 3	1.934	0.559	11.973	0.001	6.917
Confidence 4	2.956	0.499	35.092	0.000	19.214

Table 2. Logistic regression predicting accuracy from display type, search type, targets' presence, set size, and confidence rating.

sideview-dynamic condition) to the observer's line of sight, did not impact the search performance for the dynamic search condition.

Heretofore, the search performance in the experimental and control conditions suggested that rotation of the search array did not affect search accuracy, search RT, search efficiency, and subjective confidence. Having references throughout the search led to higher search accuracy and better metacognitive judgment of one's own performances.

Eye movement with static and dynamic displays

The analyses of behavioral measures suggested that being able to refer to the target during the reference search improved the accuracy and searching from the rigidly rotating arrays did not change the accuracy or search RT. In addition to search outcomes, we further analyzed eye movements during the search to uncover similarities and discrepancies in the search process when looking at a stationary or a rotating array with or without a reference item. Specifically, we examined the proportion of trials where the target was ever fixated on over total number of trials (p) to study the process of target detection, and the proportion of trials where targets were identified over trials where targets were fixated upon (p_{id}) to study the process of target identification. We ran Bayesian repeated measures ANOVAs with p and p_{id} as the dependent measures and display type and search type as the factors to answer the questions that how the rotation of the search array or having the reference items affected the target detection and target identification.

First, target detection p was not affected by display type ($BF_{incl} = 0.20$), search type ($BF_{incl} = 0.23$), or their interaction ($BF_{incl} = 0.13$). Observers were equally likely to fixate on the target in the dynamic display condition (mean = 73.7%, 95% within-subject CI = 66.6%, 80.8%) and in the static display condition (mean = 73.5%, 95% within-subject CI = 65.7%, 81.4%); observers were equally likely to fixate on the target in the reference search condition (mean = 75.1%, 95% within-subject CI = 67.2%, 83.0%) and in the memory search condition (mean = 72.1%, 95% within-subject CI = 62.9%, 81.4%), regardless of whether the search array was stationary or rotating.

Second, target identification p_{id} was higher in reference search than in memory search (mean$_{reference}$ = 92%, mean$_{memory}$ = 85%, $BF_{incl} = 3.73$). Target identification was not different between the static display condition (mean = 91.0%, 95% within-subject CI = 86.5%, 95.6%) and the dynamic display condition (mean = 86.5%, 95% within-subject CI = 82.4%, 90.6%, $BF_{incl} = 0.81$). The interaction of the display type and the search type did not affect the p_{id}.

The total search RT in a search trial was divided into four segments – looking at the target, looking
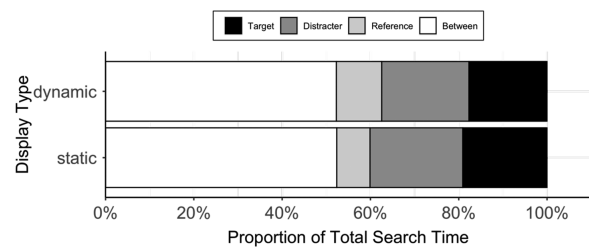

Figure 10. The proportions of looking time at the target, at the distracters, at the reference item and at elsewhere on screen in the static/dynamic display conditions.

at the distracters, looking at the reference item, and looking elsewhere on the display. The proportion of each segment over total looking time was computed. Although the total search RT was not different between display types (mean$_{static}$ = 4.26 seconds, mean$_{dynamic}$ = 4.50 s, $BF_{incl} = 0.33$) and between search types (mean$_{memory}$ = 4.32 seconds, mean$_{reference}$ = 4.44 seconds, $BF_{incl} = 0.05$), we further examined the proportion of each looking segment in the static/dynamic display conditions and in the memory/reference search conditions. First, with static or dynamic displays, the proportions of looking at the target ($BF_{10} = 0.30$), at the distracters ($BF_{10} = 0.31$), at the reference item ($BF_{10} = 0.46$), or at elsewhere ($BF_{10} = 0.26$) were equivalent (Figure 10). Second, with the memory or the reference search, the proportions of looking at the target ($BF_{10} = 0.37$) or at the distracters ($BF_{10} = 0.63$) were equivalent.

Discussion

In Experiment 2, we introduced rotation and hence continuous perspective change of the search array, but this did not affect the search performance in terms of accuracy, search RT, search efficiency, and subjective confidence rating. Furthermore, our eye movement analysis suggested that the rotation of the display did not affect any of the subprocesses of the search either. More oculomotor measures, such as fixation-based and saccade-based measures, will be collected in the upcoming experiments to further explore the effects of continuous perspective change on search performance and the more detailed search behaviors.

Consistent with the findings of Experiment 1, having reference items during the search resulted in higher search accuracy and higher confidence rating. Eye movement patterns revealed that in the reference search participants were better at identifying the target after detecting it, which might contribute to the overall higher accuracy. Unlike in Experiment 1, the search RT did not differ between memory search and reference search in Experiment 2. The memory

search in Experiment 2 was no longer faster than the reference search, probably because the memory search in the dynamic display condition involved complex internal processes like mental rotation that were more challenging to the observers than the memory search in Experiment 1. Furthermore, the presence of reference items in Experiment 2 did not speed up the target identification process like in Experiment 1, possibly because when participants looked back and forth between the reference item and the target object, the rotation of the search array in the dynamic display condition caused image disparities between these two, thus complicating the process of reference-making.

General discussion

In this study, we investigated visual search performance when there were image transformations of the search array items due to viewing perspective change (Experiment 1) and motion of the search array (Experiment 2), when observers must memorize the targets or when they could resort to external references in real time. In short, we found that perspective change and image distortions that followed did not change trends of search behaviors and that having external references made search slower but more accurate. More details are discussed below.

First, participants were able to find the targets despite image distortions. Search was equally accurate with 2D exemplar views, with 2D slanted views, and with continuous image transformations whether starting from the frontal view or the side view ($BF_{10} = 0.16$). Therefore, as far as accuracy was concerned, the visual search was not view-point dependent.

Second, in accurate search trials, the RTs were equivalent when searching from the frontal or slanted views of objects; neither did it increase when searching from rotating arrays as compared to from the static arrays. This implies that participants could efficiently search for objects despite image transformations and, in some cases, with changing retinal locations. The RT was longer when rotation began with the sideview of the search items than when rotation began with the frontal views as a result of the experimental setup, because in the sideview-dynamic condition, objects' distinguishing features gradually came to view. Gauging by the RTs, the current experimental task was harder than classic laboratory-based search tasks (such as searching for a "T" among "L"s) that typically lasted for approximately 10^{-1} seconds, and easier than complex search tasks (such as finding an abnormal cell from a mammogram image) that typically lasted for approximately 10^{2} seconds (Wolfe & Van Wert, 2010).

Consistent with the findings of traditional visual search tasks, search RTs in this study increased linearly with set size and the rate of increment, which reflects search efficiency, was equivalent when searching with the frontal or slanted views, from rotating or stationary arrays. The increment of approximately 300 ms/item in the target-present trials of current study was higher than the typical rate of increment approximately 50 ms/item) when searching for semantically meaningful symbols, such as numbers or letters, which indicated that searching for irregularly shaped objects was likely more difficult. Furthermore, in this study, search efficiency was lower by half in target-absent trials than in target-present trials. This implied that when searching for a target or confirming its absence, an observer examined each displayed item until the target was found or went through all items before claiming target absence. Implied by the target-present/absent slope relation, with N items in the search array, averaging over a large number of trials, observers looked at all N items on target-absent trials but at $(N + 1)/2$ items on target-present trials, which yielded the 2:1 slope ratio when plotting search RT over set size for target-absent and target-present searches. Hence, search in this study was serial and self-terminating (Wolfe, 2021).

We also examined the self-rated confidence level to see if the observers' metacognitive sensitivity was congruent with their search behavior. The confidence ratings were equivalent regardless of viewing perspectives, the presence of motion, or whether rotation began with objects' frontal or side views facing the observers. This echoed with the search accuracy and RT. Furthermore, in both experiments, the confidence rating was able to predict the search correctness and the higher the confidence, the more likely that the search was accurate. Interestingly, in the incorrect trials of both experiments, participants' mean felt confidence was still higher than the mid-point of the 4-point scale. Thus, observers' metacognition was not accurate, and they tended to be overly confident.

The difference of object appearances as a result of perspective projections in the various experimental conditions allowed us to uncover whether visual search was based on pixel-level image matching or object identification. The frontal view condition of Experiment 1 was similar to a classic visual search task, where objects in the search array were exemplar views of LEGO blocks and the target in the learning phase and in searching phase were identical. In the slanted view condition, with the non-parallel viewing, the target appeared slightly different in the learning phase and in the search phase. This was because when shooting pictures to create the search task stimuli, the camera was fixed and pointed down at the turntable; in the learning phase, the target was always in the center of the turntable, but in the search phase it could be anywhere amid other items occupying a non-centered location on the turntable. So, in the pictures, which were later loaded into MATLAB and used as test displays, the

projected images (sizes and shapes) of the target object varied depending on its physical location. The slanted view condition was, in fact, a better representation of real-world searches, where searchers typically would not know a target's location a priori to predict its projected image. Given that with or without an exact match of a target's appearance, participants' search performance, in terms of accuracy, RT, and efficiency, were equivalent, the visual search process was likely not based on matching of retinal images. Instead, recognizing objects based on their defining features and differentiating objects based on their distinguishing features seem to be prerequisites of finding randomly shaped objects.

Similarly, in Experiment 2, participants searched for LEGO blocks in an immersive VR environment where they looked down at an array of search items which were either stationary (static display condition) or rotating as a group (dynamic display condition). In the rotation condition, the target was stationary in the learning phase but it went through continuous image transformation in the search phase. In other words, images of the target were never identical, but this did not affect search performance either in terms of outcomes such as accuracy and search RT or in terms of the searching process such as time spent to reject distracters, identify targets, or check reference items, as revealed by eye movement analyses. These strongly suggest that humans do not apply pixel-level image matching to find targets in visual search tasks.

In real-life practices, airport security officers could benefit from seeing multiple views of items when searching for suspicious articles. In a study involving 5717 airport security screeners from more than 70 airports, when training them with multiple views of target objects, their detection sensitivity of the target item increased tremendously (Halberr, Schwaninger, Budgell, & Wales, 2013). Furthermore, professional searchers who worked with 3D rotatable images of items exhibited better detection performance than those who worked with 2D images (Hättenschwiler, Mendes, & Schwaninger, 2019). As compared with these results, the lack of improvement when searching from dynamic displays in the current study was possibly because participants in this study were not trained and only professional searchers notably benefited from moving 3D objects.

We attempted to unveil the search process by investigating eye movement patterns, paired with search outcomes. The eye movement data allowed us to parse a search trial's total RT into looking time at the target, looking time at the distracters, looking time at the reference object (when present), and looking time at elsewhere on the screen. In Experiment 2, when the participant performed the memory search, their search RT increased more with each additional item in the dynamic display condition than in the static display condition; but this difference was not found in the reference search. Parsing the total search RT into each segment, we found that when there was no reference object, the looking time at the target was equivalent when searching in the static and dynamic display conditions ($BF_{10} = 0.32$); looking time at the distracters was also equivalent between these conditions ($BF_{10} = 0.26$); but the time that the participants spent looking elsewhere on the screen was longer in the dynamic display condition than the static display condition ($BF_{10} = 1.85$). These implied that when the search array items were rotating, the participants spent more time locating a moving target but once they had located it, they spent a similar amount of time verifying it. So, the drop of search efficiency in the memory search with the dynamic displays was possibly due to the difficulty in locating a target from a moving array than recognizing it. This was not found in the reference search with the dynamic displays, in which case, the looking times at the target, at the distracters, at the reference item, and at elsewhere on the screen were all similar whether the search was from static or dynamic displays ($BF_{10} < 0.90$ in all contrasts). It is possible that the observers used the reference target object as a landmark, and used it to anchor locations of specific search array items (for example, potential targets) in the rotational display for further processing.

Results from the current study showed that perspective change, either being discrete (as a result of viewing angle change) or continuous (as a result of rotation), did not affect the search performance as a whole. But what about the subprocesses of target selection and target identification? Were they equally (un) affected? With a discrete perspective change in Experiment 1, the target identification process became slower when searching in the slanted view condition than in the frontal view condition ($BF_{10} = 3.59$). This was consistent with the findings of baggage screener studies that objects were more difficult to detect when viewed from an unusual, noncanonical viewpoint (Koller, Hardmeier, Michel, & Schwaninger, 2008; Biggs, & Mitroff, 2015). Moreover, the perspective effect on target identification time was consistent with finding in the classic object recognition literature that it was more difficult to identify objects from novel views as compared to from familiar or learned views, and yielding slower responses and more errors (Friedman, Vuong & Spetch, 2009; Friedman, Vuong, & Spetch, 2010).

However, when searching from a moving array (Experiment 2, dynamic display condition), the target identification subprocess was as fast ($BF_{10} = 0.46$) as in the static frontal view condition (Experiment 1, frontal view condition) and both were better than in the static slanted view condition. Continuous rotation did not impede the visual search possibly because the continuous image transformation in the 3D enabled

structure-from-motion (SFM) and therefore the immediate specification of object shapes that led to fast and accurate object identification. It has been shown that 3D objects that oscillated in depth with amplitudes ≥45 degrees generated sufficient information to specify their shapes and allow observers to distinguish them (Lee & Bingham, 2010; Lee et al., 2012; Lind et al., 2014). Thus, instead of being a source of perturbation, motion and the resultant continuous perspective change might provide extra information for discriminating objects and fostering visual search. This, again, suggested that the visual search was based on object identification instead of image-matching. Convergingly, previous object identification studies had reported facilitative effects of motion on object identification and recognition. For example, Papenmeier and Schwan (2016) showed that when observers first learned objects from videos of them rotating and then attempted to recognize them from static images, they performed better than when they first learned objects from static images and then recognized them from static images. Rotation brought continuous views of the objects, which, according to Tjan and Legge (1998), led to accurate recognition even when there were Gaussian noises in the images.

Finally, in both experiments, participants were more accurate in and more confident about their search when the target remained visible throughout search. Eye movement analyses revealed that 18.0% of search RT was spent on looking at the reference items in the reference search of Experiment 1, and 16.9% in Experiment 2. In the reference trials of both experiments, the average number of items on the screen was 8.5 and the looking time percentage at the reference target (18.0% or 16.9%) was significantly longer than the average of 11.8% ($BF_{10} = 5.95$, and $BF_{10} = 1.24$, respectively). This showed that participants actively used the external reference and sought just-in-time information. Visual search was not the only task that could benefit from an active interaction between an observer's internal cognitive resources and structures in the world. For example, Pan, Bingham, and Bingham (2013; Pan, Bingham & Bingham, 2017) showed that landmarks were extremely helpful in preserving spatiotemporal relations and aiding participants to locate and identify previous-seen-but-currently-hidden objects. Hayhoe and colleagues (Pelz, Hayhoe, & Loeber, 2001; Droll & Hayhoe, 2007) showed that in various tasks, such as sorting objects or copying colorful patterns, participants did not memorize the locations or patterns in their head; instead, they looked at the target locations or objects and acquired just-in-time information to lessen burdens on internal cognitive resources. In their words, there were "trade-offs between gaze and working memory use." This "trade-off" entailed making references to external structures and was an embodied process that led to

more accurate and efficient performances in visual search and other cognitive tasks than when observers had to rely entirely on their internal cognitive resources for memorizing and processing visual targets.

Conclusion

In this work, we studied the visual search with image transformations due to discrete or continuous perspective change when the target was or was not visible during the search. The introduction of image transformation and reference objects was to create closer approximations to a real-world visual search. Search performance was generally unaffected by image transformations but improved with the presence of reference objects. This suggests that visual search does not require pixel-level image matching but requires object identification and it could benefit from an embodied process, in which the observers acquire real-time information from the environment to improve their performance.

Acknowledgments

Commercial relationships: none.
Corresponding author: Jing Samantha Pan.
Email: panj27@mail.sysu.edu.cn.
Address: Department of Psychology, Sun Yat-sen University, Daxuecheng Guhe South Road Nearby, Guangzhou, Guangdong, China.

References

Biggs, A. T., & Mitroff, S. R. (2015). Improving the efficacy of security screening tasks: A review of visual search challenges and ways to mitigate their adverse effects. *Applied Cognitive Psychology, 29*(1), 142–148.

Bingham, G. P., & Lind, M. (2008). Large continuous perspective transformations are necessary and

sufficient for accurate perception of metric shape. *Perception & Psychophysics, 70*(3), 524–540.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision, 10*(4), 433–436.

Cain, M. S., Adamo, S. H., & Mitroff, S. R. (2013). A taxonomy of errors in multiple-target visual search. *Visual Cognition, 21*(7), 899–921.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19.

Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 1352.

Duncan, J. S., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433–458.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Fleming, S. M., & Dolan, R. J. (2010). Effects of loss aversion on post-decision wagering: Implications for measures of awareness. *Consciousness and Cognition, 19*(1), 352–363.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science, 329*(5998), 1541–1543.

Foulsham, T., Chapman, C. S., Nasiopoulos, E., & Kingstone, A. (2014). Top-down and bottom-up aspects of active search in a real-world environment. *Canadian Journal of Experimental Psychology, 68*(1), 8–19.

Friedman, A., Vuong, Q. C., & Spetch, M. L. (2009). View combination in moving objects: The role of motion in discriminating between novel views of similar and distinctive objects by humans and pigeons. *Vision Research, 49*(6), 594–607.

Friedman, A., Vuong, Q. C., & Spetch, M. (2010). Facilitation by view combination and coherent motion in dynamic object recognition. *Vision Research, 50*(2), 202–210.

Godwin, H. J., Menneer, T., Liversedge, S. P., Cave, K. R., Holliman, N. S., & Donnelly, N. (2020). Experience with searching in displays containing depth improves search performance by training participants to search more exhaustively. *Acta Psychologica, 210*, 103173.

Halbherr, T., Schwaninger, A., Budgell, G. R., & Wales, A. (2013). Airport security screener competency: a cross-sectional and longitudinal analysis. *The International Journal of Aviation Psychology, 23*(2), 113–129.

Hättenschwiler, N., Mendes, M., & Schwaninger, A. (2019). Detecting bombs in X-ray images of hold baggage: 2D versus 3D imaging. *Human Factors, 61*(2), 305–321.

Hayes, T. R., & Henderson, J. M. (2019). Scene semantics involuntarily guide attention during visual search. *Psychonomic Bulletin & Review, 26*(5), 1683–1689.

Kingstone, A., Smilek, D., Ristic, J., Kelland Friesen, C., & Eastwood, J. D. (2003). Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Science, 12*(5), 176–180.

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

Koch, C., & Ullman, S. (1987). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4*(2), 115–141.

Koller, S. M., Hardmeier, D., Michel, S., & Schwaninger, A. (2008). Investigating training, transfer and viewpoint effects resulting from recurrent CBT of X-Ray image interpretation. *Journal of Transportation Security, 1*(2), 81–106.

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences, 14*(7), 293–300.

Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition, 10*(3), 294–340.

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences, 103*(49), 18763–18768.

Lau, H. C., & Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *Journal of Neuroscience, 27*(21), 5805–5811.

Lee, Y. L., & Bingham, G. P. (2010). Large perspective changes yield perception of metric shape that allows accurate feedforward reaches-to-grasp and it persists after the optic flow has stopped! *Experimental Brain Research, 204*(4), 559–573.

Lee, Y. L., Lind, M., Bingham, N., & Bingham, G. P. (2012). Object recognition using metric shape. *Vision Research, 69*, 23–31.

Lind, M., Lee, Y. L., Mazanowski, J., Kountouriotis, G. K., & Bingham, G. P. (2014). Affine operations plus symmetry yield perception of metric shape with large perspective changes (≥45°): Data and model. *Journal of Experimental Psychology: Human Perception and Performance, 40*(1), 83.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason, 4*(2), 61–64.

Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research, 51*(13), 1526–1537.

Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics, 53*(3), 279–291.

Pan, J. S., Bingham, N., & Bingham, G. P. (2013). Embodied memory: effective and stable perception by combining optic flow and image structure. *Journal of Experimental Psychology: Human Perception and Performance, 39*(6), 1638–1651.

Pan, J. S., Bingham, N., & Bingham, G. P. (2017). Embodied memory allows accurate and stable perception of hidden objects despite orientation change. *Journal of Experimental Psychology: Human Perception and Performance, 43*(7), 1343–1358.

Pan, J. S., Bingham, N., Chen, C., & Bingham, G. P. (2017). Breaking camouflage and detecting targets require optic flow and image structure information. *Applied Optics, 56*(22), 6410–6418.

Pan, J. S., Li, J., Chen, Z., Mangiaracina, E. A., Connell, C. S., & Wu, H. (2017). Motion-generated optical information allows event perception despite blurry vision in AMD and amblyopic patients. *Journal of Vision, 17*(12):13, 1–16.

Papenmeier, F., & Schwan, S. (2016). If you watch it move, you'll recognize it in 3D: Transfer of depth cues between encoding and retrieval. *Acta Psychologica, 164*, 90–95.

Pelz, J., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research, 139*(3), 266–277.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision, 7*(14):16, 1–20.

RStudio Team. (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA, http://www.rstudio.com/.

Samaha, J., Iemi, L., & Postle, B. R. (2017). Prestimulus alpha-band power biases visual discrimination confidence, but not accuracy. *Consciousness and Cognition, 54*, 47–55.

Samaha, J., Switzky, M., & Postle, B. R. (2019). Confidence boosts serial dependence in orientation estimation. *Journal of Vision, 19*(4):25, 1–13.

Seidl-Rathkopf, K. N., Turk-Browne, N. B., & Kastner, S. (2015). Automatic guidance of attention during real-world visual search. *Attention, Perception, & Psychophysics, 77*(6), 1881–1895.

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):7, 1–21.

Szczepanowski, R., & Pessoa, L. (2007). Fear perception: Can objective and subjective awareness measures be dissociated? *Journal of Vision, 7*(4):10, 1–17.

Theeuwes, J., Kramer, A. F., & Belopolsky, A. V. (2004). Attentional set interacts with perceptual load in visual search. *Psychonomic Bulletin & Review, 11*(4), 697–702.

Tittle, J. S., & Braunstein, M. L. (1993). Recovery of 3-D shape from binocular disparity and structure from motion. *Perception & Psychophysics, 54*(2), 157–169.

Tjan, B. S., & Legge, G. E. (1998). The viewpoint complexity of an object-recognition task. *Vision Research, 38*(15–16), 2335–2350.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8*(2), 194.

Treisman, A., & Gelade, G. (1980) A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.

van den Bergh, D., Van Doorn, J., Marsman, M., Draws, T., Van Kesteren, E. J., & Derks, K. (2020). A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. *LAnnee Psychologique, 120*(1), 73–96.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., . . . Wagenmakers, E. J. (2019). The JASP guidelines for conducting and reporting a Bayesian analysis. manuscript submitted for publication. Retrieved from: psyarxiv.com/yqxfr.

Washburn, D. A., Smith, J. D., & Taglialatela, L. A. (2005). Individual differences in metacognitive responsiveness: Cognitive and personality correlates. *The Journal of General Psychology, 132*(4), 446–461.

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*(1), 35–57.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*(1), 33–39.

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, *28*(4), 1–33.

Wolfe, J. M., & Gancarz, G. (1997). *Guided Search 3.0. In Basic and clinical applications of vision science* (pp. 189–192). Dordrecht: Springer.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour, 1*(3), 1–8.

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology, 20*(2), 121–124.

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel-serial visual search. *Journal of Experimental Psychology Human Perception & Performance, 23*(1), 244.

Zhang, M., Feng, J., Ma, K. T., Lim, J. H., Zhao, Q., & Kreiman, G. (2018). Finding any Waldo with zero-shot invariant and efficient visual search. *Nature Communications, 9*(1), 1–15.

Zhao, Q., & Koch, C. (2011). Learning a saliency map using fixated locations in natural scenes. *Journal of Vision, 11*(3):9, 1–15.