# Knowing where to go: Spatial memory guides eye and body movements in a naturalistic visual search task

**M. Pilar Aivar**
Facultad de Psicología, Universidad Autónoma de Madrid, Madrid, Spain

**Chia-Ling Li**
Institute of Neuroscience, The University of Texas at Austin, Austin, TX, USA
Present address: Apple Inc., Cupertino, California, USA

**Matthew H. Tong**
Center for Perceptual Systems, The University of Texas at Austin, Austin, TX, USA
Present address: IBM Research, Cambridge, Massachusetts, USA

**Dmitry M. Kit**
Center for Perceptual Systems, The University of Texas at Austin, Austin, TX, USA
Present address: F5, Boston, Massachusetts, USA

**Mary M. Hayhoe**
Center for Perceptual Systems, The University of Texas at Austin, Austin, TX, USA

**Most research on visual search has used simple tasks presented on a computer screen. However, in natural situations visual search almost always involves eye, head, and body movements in a three-dimensional (3D) environment. The different constraints imposed by these two types of search tasks might explain some of the discrepancies in our understanding concerning the use of memory resources and the role of contextual objects during search. To explore this issue, we analyzed a visual search task performed in an immersive virtual reality apartment. Participants searched for a series of geometric 3D objects while eye movements and head coordinates were recorded. Participants explored the apartment to locate target objects whose location and visibility were manipulated. For objects with reliable locations, we found that repeated searches led to a decrease in search time and number of fixations and to a reduction of errors. Searching for those objects that had been visible in previous trials but were only tested at the end of the experiment was also easier than finding objects for the first time, indicating incidental learning of context. More importantly, we found that body movements showed changes that reflected memory for target location: trajectories were shorter and movement velocities were higher, but only for those objects that had been searched for multiple times. We conclude that memory of 3D space and target location is a critical component of visual search and also modifies movement kinematics. In natural search, memory is used to optimize movement control and reduce energetic costs.**

## Introduction

Visual search is a common daily activity. To perform any action, like getting dressed, the first step is often to find the necessary objects in the environment (such as clothes). Therefore it is not surprising to find an overwhelming abundance of articles based on visual search tasks (for reviews see: Eckstein, 2011; Wolfe, 1998). Much is known about how humans perform computer-based search tasks, but only in a relatively few cases has visual search been analyzed in natural settings (Foulsham, Chapman, Nasiopoulos & Kingstone, 2014; Franchak, McGee & Blanch, 2021; Howard, Pharaon, Körner, Smith & Gilchrist, 2011; Jiang, Won, Swallow & Mussack, 2014; Nachtnebel, Cambronero-Delgadillo, Helmers, Ischebeck & Höfter, 2023; Spotorno, Dragusin, Kirtley & Tatler, 2019). This issue is important because there are crucial differences

Citation: Aivar, M. P., Li, C.-L., Tong, M. H., Kit, D. M., & Hayhoe, M. M. (2024). Knowing where to go: Spatial memory guides eye and body movements in a naturalistic visual search task. *Journal of Vision*, *24*(9):1, 1–23, https://doi.org/10.1167/jov.24.9.1.

https://doi.org/10.1167/jov.24.9.1 Received August 25, 2023; published September 3, 2024 

between searching for elements on a computer screen and searching for objects in a three-dimensional space. First, to find an item on a screen you only need to move your eyes, at most. It is even possible to perform visual search tasks while maintaining central gaze fixation (for a review, see Carrasco, 2011). This work ignores the role of eye movements during search (for a discussion, see Hulleman & Olivers, 2017). In natural environments, however, eye and head movements, and very often also locomotion, are required to explore the visual field and to locate the target object (Land & Hayhoe, 2001; Land & Tatler, 2009). The response is also very different in both kinds of search tasks. In the laboratory, visual search tasks usually require just a key press to register target detection, while search in a natural setting normally occurs as part of a global action, so once found the target object would most likely be picked up and used for some other task (Tatler, 2014; Tatler, Hayhoe, Land & Ballard, 2011; Tatler & Land, 2016).

The different constraints imposed by these two types of search tasks might explain some of the discrepancies in our understanding of some aspects of visual search, like the role of memory resources (Kristjánsson & Draschkow, 2021). From results obtained in computer-based tasks it has been suggested that information does not need to be retained during visual search (Horowitz & Wolfe, 1998). However, contextual cueing experiments also show that repeatedly searching for the same targets on the same display facilitates search (Chun, 2000; Chun & Jiang, 1998), although this process is quite slow, often requiring thousands of trials. This means that the attentional system is sensitive to environmental regularities and takes advantage of co-variations to improve performance (Conci, Zellin & Müller, 2012; Zhang & Carlisle, 2023; Zinchenko, Conci, Müller & Geyer, 2018). Interestingly, when 2D images of natural scenes are used, other forms of memory come into play and facilitation occurs much faster (Vö & Wolfe, 2015). Similar results have been obtained under more ecological conditions, such as when repeatedly searching for objects in an immersive virtual reality apartment (Draschkow & Vö, 2017; Kit et al., 2014).

One aspect that might be important in this discussion is the *cost* associated with search (Christie & Schrater, 2015; Hayhoe, 2017; Moskowitz, Berger, Castelhano, Gallivan & Flanagan, 2022). Moving the eyes around to find an item on the screen is fast and easy, which might discourage the use of memory processes to determine target location information. However, relying on memory representations during natural actions is probably cost effective, given the benefits it provides at the motor level. The different effectors (eyes, head, hands, legs) have different kinematic properties that constrain motor planning. For example, head and hand movements are slower than eye movements, more energetically costly, and have longer latencies. Using spatial knowledge about the environment might facilitate the spatial and temporal synchronization of all the effectors (Binsted, Chua, Helsen & Elliott, 2001; Pelz, Hayhoe & Loeber, 2001; Sailer, Eggert, Ditterich & Straube, 2000), but these kinds of effects are hard to see in standard computer-based visual search tasks, which can be successfully performed with small amplitude saccades (Over, Hooge, Vlaskamp & Erkelens, 2007). This means that computer-based tasks have *low* motor costs because they are mostly *static*. When tasks are designed to require head movements, however, results suggest that participants try to avoid large movements of the head (Hardiess, Gillner & Mallot, 2008), probably to reduce motor costs. If a task has a higher motor cost, for example by requiring moving a mouse or locomotion, visual search strategies might adapt even further to the motor constraints (see, for example, Li, Chen, Wolfe & Olivers, 2023; Scrafton, Stainer & Tatler, 2017). The presence of obstacles in the hand path to the target has been shown to influence visual search as well (Moskowitz, Fooken, Castelhano, Gallivan & Flanagan, 2023). Therefore, to understand performance in a given task, we need to consider the tradeoff that occurs between cognitive and motor costs.

Some evidence supports these claims. For example, in a repeated visual search task, in which participants could use memory to direct gaze to the correct target location, they choose not to do so (Kunar, Flusberg & Wolfe, 2008). In this case performance only required saccadic eye movements. However, when a very similar repeated visual search task required changes in head orientation to locate the targets, the use of location information from previous trials increased (Solman & Kingstone, 2014). A similar result was found by Hardiess et al., (2008) in a comparative visual search task in which head movement amplitude was manipulated across trials. Incrementing search difficulty also increased reliance on memory for target location (Solman & Smilek, 2012) and incidental encoding of distractors characteristics (Guevara-Pinto, Papesh & Hout, 2020). In an object-copying tasks using Lego blocks, the need for larger head movements also led to a higher retention of information in working memory (Ballard, Hayhoe & Pelz, 1995). This result was replicated when a very similar task was implemented in virtual reality with whole body movements (Draschkow, Kallmayer & Nobre, 2021). Performing actions in real, natural contexts also facilitates the extraction and retention of detailed information about the properties of objects, even those that are irrelevant for the task (Tatler & Tatler, 2013). Overall, when the mode of exploration requires higher levels of activity the resulting spatial knowledge seems to increase (Chrastil & Warren, 2012).

In this context, Virtual Reality combined with eye-tracking provides a useful research tool to clarify some of the issues discussed above. For example, in a previous experiment we directly compared performance on a visual search task in two- (2D) and three-dimensional (3D) environments (Li, Aivar, Kit, Tong & Hayhoe, 2016). Participants searched for a series of geometric objects in three consecutive blocks and then had to find some contextual objects that were visible during previous searches. The 3D version of the task was performed in a virtual reality apartment, while the 2D version was an adaptation in which images of the apartment were presented on a computer screen and participants simulated movement by changing images through key presses. We found a greater use of memory in 3D compared to 2D.

These results show that studying search behavior in conditions that are more similar to those found in natural circumstances is useful to understand whether visual search strategies differ from those seen in simple, computer based tasks (Kristjánsson & Draschkow, 2021). But there is an additional benefit: it also allows us to explore whether searching in an environment has an effect at the motor level, for example modifying the kinematic properties of the movements. In a previous study (Li, Aivar, Tong & Hayhoe, 2018) we found that participants often planned their body movements using remembered information about target location and that trajectory length decreased with more experience in the task. The goal of this new experiment was to examine the nature of memory representations and their role in movement control. Our main goal was to analyze the effects of spatial memory on locomotion control (walking trajectories). In addition, we wanted to better characterize the relation between attention and memory by controlling visual exposure to the relevant objects. We also manipulated target locations to test whether search strategies varied depending on the *reliability* of the information (Zinchenko et al., 2018). We also wanted to test if co-occurrence facilitated search, for example providing a local context that could help participants locate a previously seen object (Mack & Eckstein, 2011). Finally, we also analyzed *incidental fixations* to a few objects that were always visible, but only became targets at the end of the experiment. All of these factors are present in most natural search tasks and might influence search strategies.

# Methods

## Participants

A total of 21 participants voluntarily took part in the experiment: three of them were authors, who did not know about the experimental conditions; five were colleagues, and the other 13 were volunteers (students from the UT Austin) who were paid for their time or received course credit for their participation.[1] They all gave their informed written consent. All reported normal vision and were naïve to the purpose of the study. Participants who normally used glasses or contact lenses were not accepted in the study to avoid problems tracking eye movements. Eight of the participants felt uncomfortable during the experiment, mostly due to motion sickness, and could not complete all the trials, so their data was removed from further analysis. The final group had 13 participants (10 males and three females, ages between 18 and 60). Unfortunately, eye-tracking data was not very reliable for 5 of these participants, with frequent signal losses and drifts (commonly found in this head-mounted display (HMD)/eye-tracking system). Therefore all reported measures dependent on gaze position are based on the data of the eight participants with good eye-tracking data (including the three authors). All other measures are based on the data of 13 participants. The research project followed the protocols of the World Medical Association Declaration of Helsinki (2008) and was approved by the Human Subjects and Institutional Review Board of The University of Texas at Austin (IRB: 2006-06-0085).

## Stimulus, apparatus, and set-up

The experiment was performed in an immersive virtual reality apartment. A top view of the apartment is shown in Figure 1. The apartment consisted of two rooms (a bedroom and a living room, approximately 3.6 meters wide by 4.5 meters long) connected by a corridor. A TV on the wall in the center of the corridor was used to present an image of the *target object* for each trial. The apartment was decorated to look natural (see left side of Figure 2) with objects and furniture chosen from a database of 3D models. There were doors from the corridor to the rooms. Doors slid up to open when participants approached them, to prevent seeing the rooms from the corridor. The apartment's size (including walls) was approximately 6.4 by 7.2 meters. Its virtual size corresponded to the real space that was tracked.

The apartment was created in FloorPlan 3D V11 (IMSI) and then rendered by Vizard 4 (WorldViz, LLC). The visual display was delivered via a nVisor SX111 (NVIS) wide field of view HMD. The nVisor SX111 is made of a pair of SXGA displays with fields of view of 76°H × 64°V per eye and a resolution of 1280 × 1024 pixels, which corresponds to horizontal and vertical fields-of-view of 102° and 64°, respectively. It weighs 1.3 kg, according to manufacturers. The stereo image was generated by Vizard using a standard windows computer. A ViewPoint monocular infrared video eye-tracker (Arrington Research, Scottsdale, AZ, USA) recording at 60 Hz was integrated into the helmet and monitored the left eye. Its accuracy was about

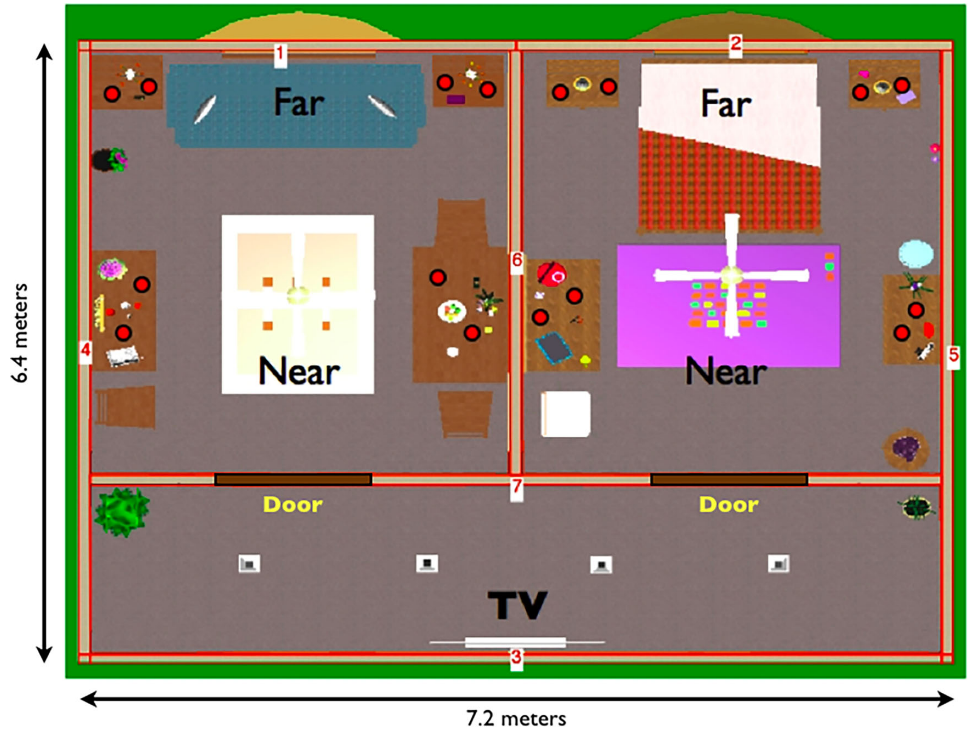


Figure 1. Bird's-eye view of the virtual apartment. The apartment included ceiling lamps, plants and multiple decorative objects, that can be seen in the image. Red dots indicate the locations in which the target objects could appear (defined as near or far depending on the distance from the door). The location of the doors and the TV in the corridor is also indicated.

1°. Because of a relocation of the laboratory halfway through the project, head position was recorded with two different systems. For six participants the HMD had six LED markers attached to it, which were tracked by a PhaseSpace motion tracking system (PhaseSpace, San Leandro, CA, USA). For the remaining seven participants head position was monitored with a wide-area ceiling HiBall motion Tracking System (3rd Tech, Chapel Hill, NC, USA) working at 600 Hz. Head position in the environment was used to update the stereo view at the 60 Hz refresh rate of the HMD and was also recorded for later analysis. The latency for updating the visual display after a head movement was about 50 to 75 ms.[2] A Wiimote Remote (Nintendo, Kyoto, Japan) was also given to the participants to indicate when the target object was found with a button press. Views of the helmet, eye-tracker and one of the authors wearing the portable system are shown in Figure 2. During the experiment the video records of the eye and scene camera were combined in a custom QuickTime digital format. Data from head, eye-position, and Wiimote presses, as well as the virtual simulation of the room and objects, was saved as synchronized metadata on each video frame. Synchronization occurred at the frequency rate of the eye-tracker (60 Hz).

## Task and procedure

The participants' task was to search for each target object until it was found. Each trial started with a new object appearing on the TV screen (see Movie 1). Participants could walk to any room they wanted first. A button press was used to indicate that the target was found, but to successfully end the trial participants had to be within 1.5 m from the target and looking at it to prevent them simply pressing without finding the target. When the response was correct, a message was displayed on the screen (*Trial Done*). At that point participants had to walk back to the TV to see the next search target. Objects on the TV were 2D versions of the 3D search targets.

Participants received written instructions and performed a nine-point calibration. After calibration was achieved, participants moved around a little to get familiarized with the HMD, the controller and the virtual environment. Then they were moved to the starting position, in front of the TV. At that point the virtual apartment was made visible and the first trial started. Participants performed a total of 52 trials, which normally took about 30-40 min. During the experiment one of the experimenters carried a backpack with the control computer and made sure that the cables did not get in the way of


Figure 2. Stimulus and apparatus. The left part of the figure shows examples of how the two rooms looked to participants during the experiment. The right part of the figure shows a participant performing the task, views of the apparatus and one of the authors wearing the portable system (top), and also an example of the video recording with the reconstructed fixation window (in red) intersecting one of the target objects (purple cylinder) (bottom).

the movements of the participant (see Figure 2). The quality of the eye tracking signal was checked one or more times (as needed), and also at the very end of the session. The eye-tracker was recalibrated if a drift was detected. The whole experiment, including initial calibration and recalibrations, normally took about an hour.

## Experimental conditions

To control for object size and saliency, ten different 3D geometrical objects from the same database were used as targets, similar to those used in previous experiments (Li et al., 2016; Li et al., 2018). Objects were about 5 × 5 × 5 cm in size and had different highly saturated colors chosen to make targets clearly different from other contextual objects in the rooms. Each color was used in two different target objects. Because color-shape combinations were unique and each shape was associated with one of the experimental conditions, at the time we did not anticipate that color repetition could be problematic. This issue will be discussed later.

Target objects were divided into four different groups, depending on visibility, reliability of target location, and moment they were tested (see Figure 3):

- *Stable objects* (red and blue ball): These objects always appeared at the same location, so participants could learn where to find them (Zinchenko et al., 2018).
- *Random objects* (red and green pyramid): These objects appeared at a new location in each trial, so it was not possible to anticipate where to find them. These objects allow us to separate the general effects of learning how to move around from specific knowledge about target location.
- *Paired objects* (purple cube and green cylinder, and yellow cube and blue cylinder): These objects appeared in pairs, always at the same locations. Each object of the pair was tested at a different time during the experiment, although both objects were

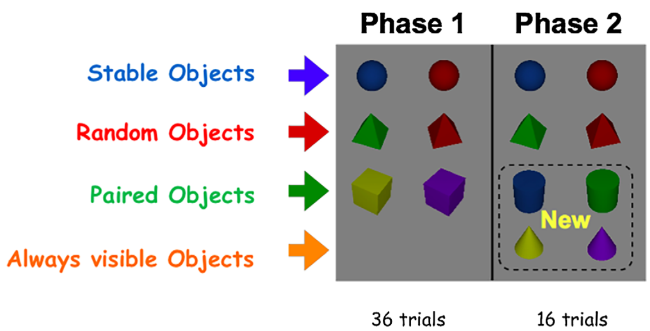


Figure 3. Target objects. Ten geometric objects were used as targets during the experiment. Objects belonged to different conditions depending on the reliability of their position (stable and paired versus random) and on the moment they were tested as targets (Phase 1 or Phase 2).

visible whenever one of them was tested. Since both objects appeared close to each other this condition allowed us to analyze the effect of local context in visual search (Mack & Eckstein, 2011).

– *Always visible objects* (yellow and purple cone): These objects had a constant location and were the only ones that were visible during the whole experiment, so they could receive incidental fixations while participants searched for other objects. They were tested at the end of the experiment, to analyze visual memory for previously seen objects.

The experiment was divided in two different phases (unbeknown to participants). During *Phase 1* participants searched for six of the target objects for a total of 6 times (36 trials in total). The objects tested during this phase were *stable objects*, *random objects* and two of the *paired objects* (see Figure 3 for a schematic representation). This phase allowed us to obtain information about the build-up of memory representations for relevant objects in reliable and unreliable locations, and also made it possible for participants to acquire information about other, still not relevant, objects in the environment. During *Phase 2* participants searched twice for eight target objects (16 trials in total): *stable* and *random objects* and four objects that had been visible during Phase 1 but had never been targets (the other two objects in the case of the *paired objects* and *always visible objects,* see Figure 3). This phase allowed us to analyze the effect of previous incidental fixations in finding the new target objects.

Except for random objects, which changed locations *each time*, all objects were distributed so that one object of each condition was placed in each of the rooms and one at each of two possible distances from the door (near/far). Near/far locations were used to evenly distribute objects in the rooms and to modulate task difficulty (far objects were harder to find). Distance from the door to the target objects, was about 2.5 meters for near locations and about 4 meters for far locations. Random objects varied location in each trial, but never occupied the locations of the other objects. Except for *always visible* objects, which were continuously present in the environment, all objects were only visible on the trial in which they were targets.

The sequence of targets was identical for all participants and was designed with the following constraints: in each block of trials (six and eight trials in Phases 1 and 2, respectively) participants searched for all the different objects; objects were never repeated in consecutive trials; and the sequence guaranteed that participants had to visit both rooms and had to turn around to return to the previously visited room equally often to find the target object. This last constraint was introduced because pilot runs had shown that participants had a strong bias to continue from the room they had just visited to the other room, so it was added to avoid an artificial increase of correct guesses.

## Pre-processing of eye-tracking and locomotion data

As in previous experiments (Li et al., 2016; Li et al., 2018) the video records of the eye and scene camera were combined in a custom QuickTime digital format which included data of head, eye-position and Wiimote presses on each video frame. The recorded metadata was analyzed offline to reconstruct the visual experience of the participants and to determine the 3D coordinates of each fixation and the identity of the fixated objects. For this, a rectangular window of 100 × 100 pixels (approximately 6.5° × 6.5° of visual angle) was centered around the recorded eye position location, and any relevant objects that projected onto this window were recorded as potential targets of a fixation (see Figure 2, bottom right for an example).

Eye position data was analyzed using custom-written programs in Matlab (MATLAB and Statistics Toolbox, The MathWorks, Inc., Natick, Massachusetts, United States), with the same parameters used in previous experiments (Li et al., 2016; Li et al., 2018). Details can be found in the Supplementary Material (Section 1). After segmenting the fixations, the contents of the fixation window allowed us to compute the total number of fixations that occurred in each of the target objects, and also a total number of *general fixations* without specific content (fixations on walls, furniture, irrelevant objects, etc.) and fixations on the TV screen.

After this, data was segmented into trials. *Complete trial* refers to the interval from the moment the target was presented on the TV to the Wiimote press. However, we also constrained our analysis to the moments in which visual search actually occurred. *In front of door* refers to the last segment of the trial, starting when the participant stopped in front of the door of the correct room, until finding the target object (more details can be found in Section 2 of the Supplementary Material).

## Data analysis

### *Temporal measures and errors*

The *total duration* of each trial was defined as the interval in seconds from the moment the target object was presented on the TV to the Wiimote press. From body position (x-y coordinates of the head) we also determined the time spent in the correct and the incorrect room. *Time at the correct room* was defined as the time spent inside the room in which the target object was present prior to finding it. *Time at the incorrect room* was calculated by taking total duration and subtracting the time at the correct room and the time spent going through the corridor. This measure includes all the time spent at the incorrect room (all visits together) and also the unsuccessful time spent at the correct room, if it were the case (on 7 trials participants did not see the target on their first visit to the correct room). The distinction between time in the correct and incorrect room also gave us a measure of the number of *errors*. Similarly to other authors (i.e., Solman & Kingstone, 2014) we also determined the actual search time in each trial. *Search time* was defined as the interval in seconds from the moment the participant was in front of the correct door (see above) to the Wiimote press. It is important to notice that search time is longer than time at the correct room. This accounts for the fact that participants start looking for the objects once they are visible through the door, that is, even before entering the room.

### *Eye movement measures*

From eye movement recordings we determined the *total number of fixations* that occurred *during visual search*. These fixations were separated according to their content (target object, other objects, and general fixations), and fixation *duration* was also analyzed. For each of the four objects that became targets during Phase 2 (green and blue cylinder and yellow and purple cone) we also determined the total number of *incidental fixations* received during Phase 1 of the experiment (that is, while participants searched for the other six target objects). For these calculations we only used the data from the eight participants who showed very good eye tracking signal.

### *Body movement measures*

Head position coordinates (x-y) were also used to characterize participants' movements in the apartment. *Movement paths* and *movement velocities* were calculated for each participant and trial. *Peak velocity* and *trajectory length* were also determined in each trial to evaluate possible changes in locomotion over the experiment. More details on how trajectories were standardized and movement velocities compared across conditions are provided in the Supplementary Material (Section 3).

After extracting the different measures, data from five trials were eliminated from further analysis because of errors in the recording (missing values). In one case data was removed because the participant took more than 15 minutes to complete one of the trials.

## Statistical analysis

We used linear mixed-effect models (LMMs) to evaluate and estimate the effects of the manipulated variables on the different aspects of performance. Because our task required repeated searches, we expected a high correlation between the responses of each participant to the same targets and variability between participants (as seen in previous experiments). LMMs are better than analysis of variance at handling correlated data (Brown, 2020), and they allow multiple correlation patterns to be explicitly modeled (Lachaud & Renaud, 2011; Seltman, 2012). Our goal when using LMMs was not to find the best-fitting model for our data but to use modeling to *estimate the effects* of the manipulated variables (Meteyard & Davies, 2020). For that reason, the only variables included in our models, as fixed predictors, were the manipulated variables: kind of target or target object, repetition and target location. All variables were included as categorical variables to obtain specific estimates for each level. No interactions were added to the models to reduce complexity and to avoid misinterpretation of the results (Brown, 2020). There were also no theoretical reasons to expect an interaction between the manipulated variables. Participants were included in the models as a random effect, with random intercepts. Model estimates were obtained with the *lme* function in Matlab (version R2022b), using the Restricted Maximum Likelihood estimation method and Cholesky parameterization to determine the variance-covariance structure (default method in Matlab). Residuals were analyzed to confirm that they followed model assumptions (homogeneity of variances and normal distribution). Details of the specific model used for each dependent variable and parameter estimates, with corresponding $t$ statistics and $p$ values, are reported in Section 4 of the Supplementary

Material. None of the models presented issues with convergence.

# Results

## General performance during the task

Participants took on average 71 seconds to perform the first trial of the experiment (from the moment they were looking at the TV to the moment the Wiimote key was pressed), while they only took 12 seconds to perform the last one (trial 52). Much of the duration of each trial was spent going through the corridor (see Movie 1). Over trials, time in the incorrect room decreased from an average of 22 sec on the first trial to an average of 3.4 sec on the last trial. Time in the correct room also decreased from an initial average value of 11 sec to an average of 4 sec on the last trial (see Figure 4). This suggests that over trials participants are quicker in determining whether the room they are visiting is the correct one. Movie 1 shows an example of this effect, comparing the first two trials of the experiment with trials that occurred at the end of the first phase (sixth repetition of target objects).

Having more experience with the task also reduced the number of errors or visits to the incorrect room: six of the 13 participants (46%) made an error on the first trial of the experiment, whereas only two participants (15%) made an error on the last trial. These results reflect a general improvement due to practice (faster trials and shorter times in each room) but also suggest memory for target locations (decrease in errors over trials).

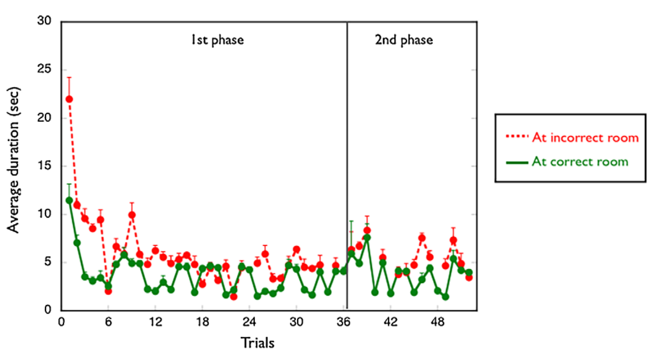


Figure 4. Time at each room. Dotted red and continuous green lines indicate the average time spent at the incorrect and the correct room on that trial. Error bars represent SEM. When the dotted red line shows no values it indicates that for that trial none of the participants visited the incorrect room. In the remaining trials the average time at the incorrect room was based on data from those participants that went to the incorrect room on that trial (which varied from trial to trial). Data from 13 participants.

## Effects of repeated search (Phase 1)

For objects at *reliable* locations we found that search time decreased over repetitions. Figure 5 shows raw data for each of these objects and a Cumming estimation plot of the differences over repetitions (Ho, Tumkaya, Aryal, Choi & Claridge-Chang, 2019). *Random* objects also showed a *decrease* in search time over repetitions, which can be seen in the upper part of Figure 6. Objects placed at the far location took between two and three seconds longer to be found. These differences were expected because participants had to be at a certain distance from the target to end the trial.

Search improvement over repetitions could be due to two factors: learning how to move in the apartment, and developing specific knowledge about target location. Two LMMs were used to estimate the improvement rate over repetitions for *stable* and *random* objects separately (see Tables S1 and S2 in Section 4 of the Supplementary Material). For *stable* objects search time had decreased by about *three seconds* after six repetitions. This reduction in search times replicates our previous findings (Li et al., 2016; Li et al., 2018). For *random* objects the model estimated that after six presentations search time had decreased by about 1.75 seconds. Search times for random objects were also slightly *faster* than for stable objects (see intercepts in Tables S1 and S2). This seems counterintuitive, but is most likely a result of target order in the first block of trials: random objects appeared as targets in trials 5 and 6, whereas stable and paired objects were presented as targets earlier in the first block. The first object presented (purple cube) had the longest search times (Figure 5). Overall, these results show that search performance improved between 1 and 1.75 seconds because of practice in the apartment, whereas specific information about target location reduced search times between 1 and 1.25 additional seconds.

Regarding number of fixations we found that over repetitions participants needed fewer fixations to find the target (Figure 7). Number of fixations was also higher for targets placed at the far end of the room (which required almost 2 additional fixations). The decrease due to repetition was more pronounced for paired objects (green circles in Figure 7) and occurred very early (see Section 5 of the Supplementary Material for more information). Interestingly, from the second block on, number of fixations was very similar for reliable and unreliable objects (see lower part of Figure 6 and Figure 7). A LMM comparing number of fixations over repetitions for *stable* and *random* objects showed that there were only very small variations due to repetition (see Table S3). Moreover, stable and random objects *did not differ*. This is in contrast with what we found for search time, so it is worth analyzing which elements are fixated in each trial.

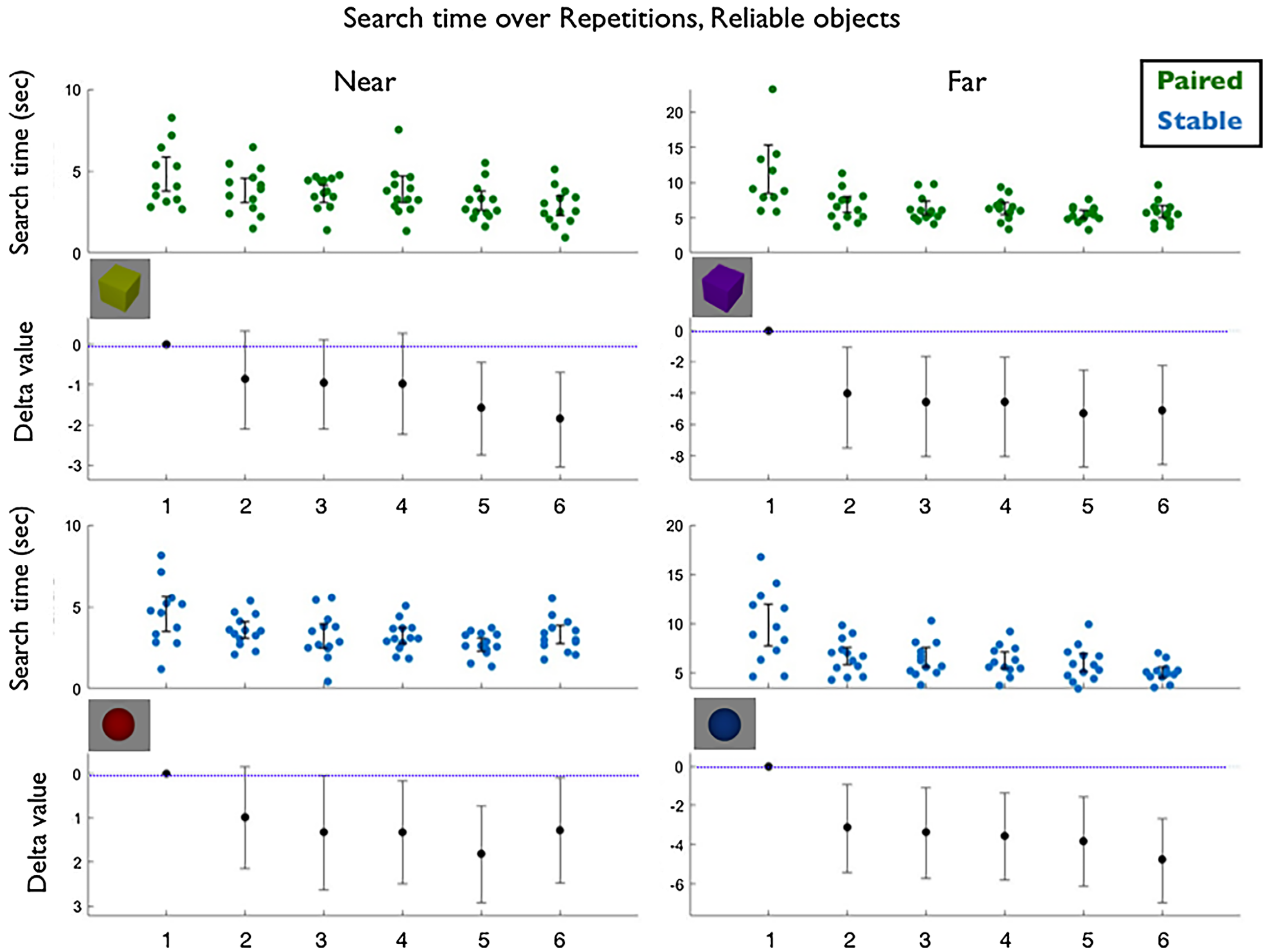


Figure 5. Search time for Reliable objects during Phase 1. Raw data and average search times are shown using a Cumming estimation plot for paired (green) and stable (blue) objects. On the lower axes mean differences are plotted as bootstrap sampling distributions. Black dots depict mean differences and the vertical error bars indicate 95% confidence intervals. The first time participants searched for each of the objects is used as reference. Measures are calculated from the moment the participant is in front of the door of the correct room. Data from 13 participants.

As Figure 8 shows, the most frequently performed fixations on all trials are *general fixations*, that is, those on walls, floor, surfaces, and other room elements (this can be seen in Movie 1). In fact, the decrease in number of fixations that occurs over the first few trials is mostly due to a reduction in general fixations and might be linked to more efficient locomotion. There are usually just one or two fixations on the target object per trial (averages: 0.8 fixations near, 1.6 fixations far). These do not seem to vary with more repetitions, which suggests that visual search strategies, once in the correct room, are very similar for all objects. On the other hand, other target objects that are present in the room (paired objects, always visible objects) seem to attract very little attention (averages lower than 1, see red lines). We will analyze this issue in more detail later on, when discussing incidental fixations.

Fixation durations also varied depending on what was fixated (Figure 9). Fixations on the target object were usually very long, with durations over 500 msec. and often up to 4500 msec. (almost as long as the time in the correct room). An example of this can be seen in Movie 1 (from 01:24). Fixations on other elements of the scene, however, were usually much shorter, with durations below or around 500 msec. Fixation durations were also systematically longer for objects placed at the far location. These results suggest that the target object served as the *heading point*: once it was found, it was fixated while participants walked toward it.

Knowing the locations of the target objects also led to a reduction in the number of errors: visits to the incorrect room varied depending on repetition and kind of object. On the first search episode proportion of errors for reliable objects was around 0.5, but by the last

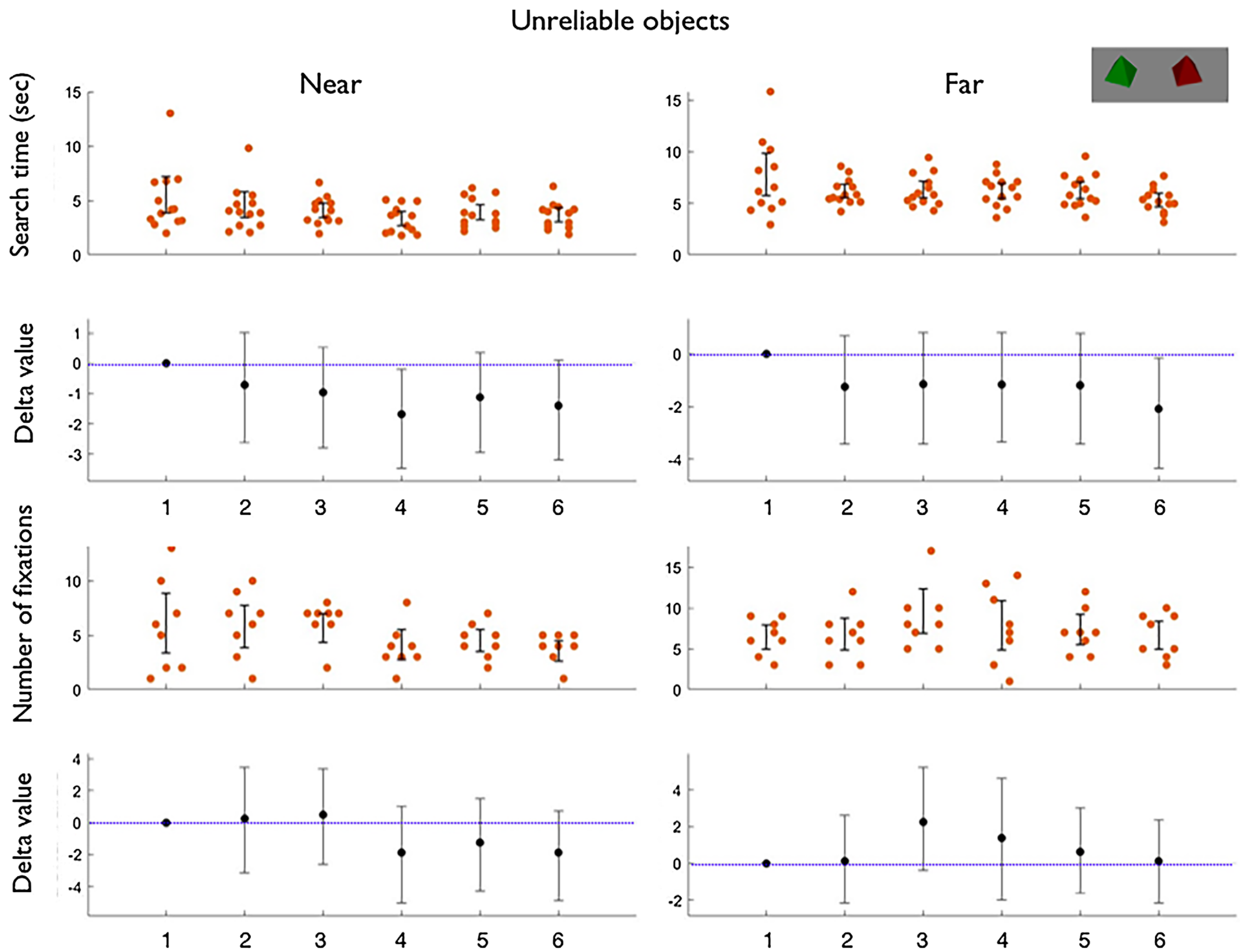


Figure 6. Search time and number of fixations for Unreliable objects during Phase 1. Raw data for search times (top) and number of fixations (bottom) is shown for random objects (red). Plots follow the same structure as those in Figure 5. The first time participants searched for one of the random objects is used as reference. It is important to remember that these objects were presented at a new location in each trial and never repeated location. Grouping is done by location, not by object. Measures are calculated from the moment the participant is in front of the door of the correct room. Search time is based on data from 13 participants and number of fixations is based on data from eight participants.

search episode proportion of errors was around 0.04 and 0.15 for reliable near and far objects respectively. Random objects, on the other hand, showed an average value around 0.5 for all objects. These differences suggest that, over repetitions, participants were able to take advantage of the fact that some objects did not change location to move directly to the correct room more often.

## Previously irrelevant objects: Incidental fixations and memory (Phase 2)

One of the purposes of this experiment was to analyze whether previous fixations on geometric objects (while they were irrelevant for the task) facilitated search for those objects later on. Did previous incidental fixations help? If participants had not acquired any information about the new objects during Phase 1, we would expect performance to be similar to that obtained the first time participants searched for any of the other objects at the beginning of the experiment. Therefore we compared the first searches of the different objects: *stable* objects (Phase 1) and *paired* and *always visible* objects (Phase 2). The results are shown in Figures 10 and 11, for reaction times and number of fixations respectively. LMMs were also used to estimate the size of the effects for each specific object (Tables S4 and S5).

*Always visible* objects showed significantly *faster* search times than *stable* objects the first time they were presented: the yellow cone was found 1 second faster than the red ball and the purple cone was found 3.6 seconds faster than the blue ball (Figure 10). For *paired* objects the pattern was different. Search times for the green cylinder were about 4.2 seconds

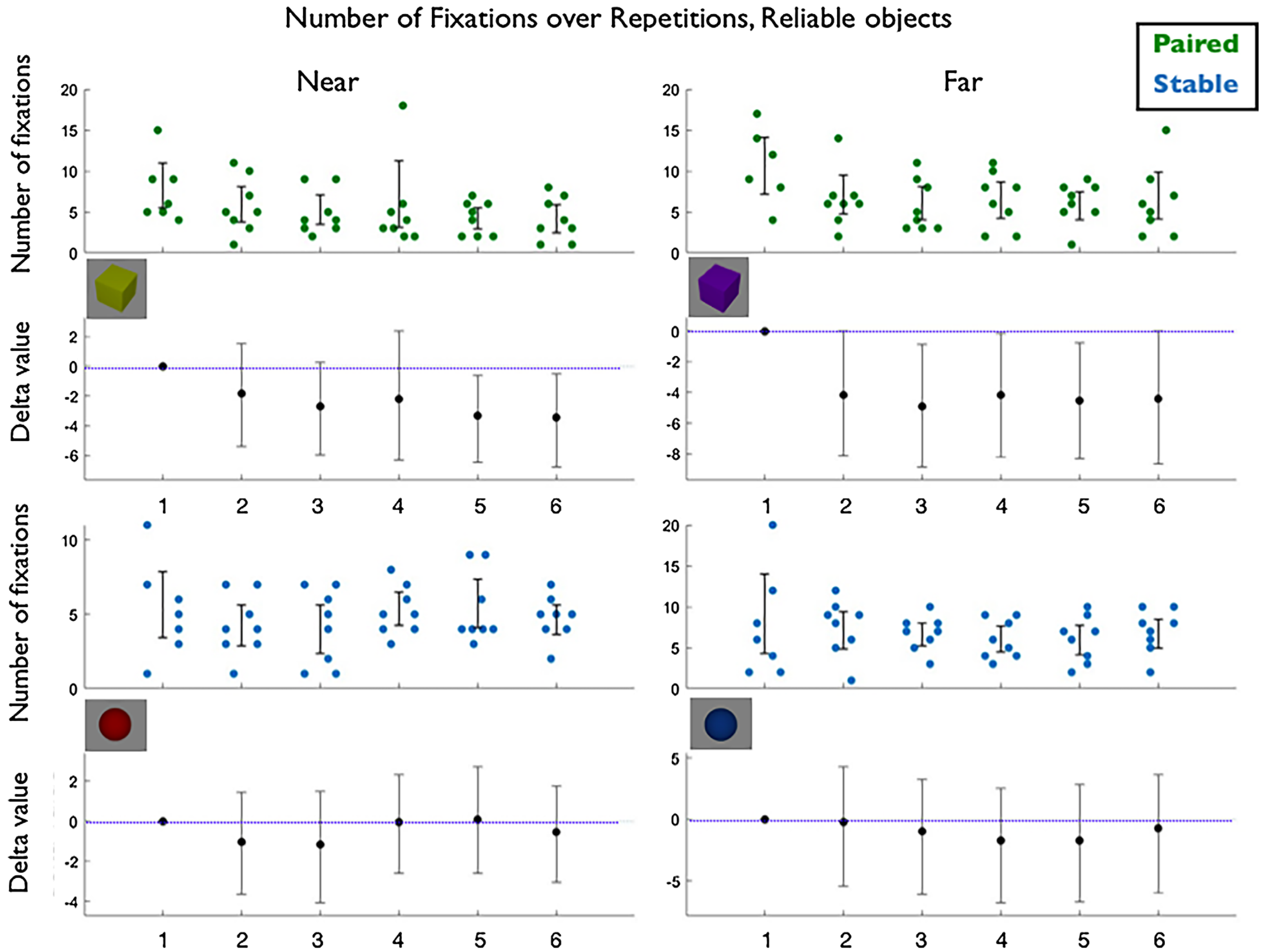


Figure 7. Number of fixations over repetitions for reliable objects during Phase 1. Raw data and average number of fixations are shown using a Cumming estimation plot for paired (green) and stable (blue) objects. Plots follow the same structure as those in Figures 5 and 6. The first time participants searched for each of the objects is used as reference. Measures are calculated from the moment the participant is in front of the door of the correct room. Data from eight participants.

faster than for the blue ball. However, finding the blue cylinder took *two seconds longer* than finding the red ball. Interestingly, previously irrelevant objects did not receive more fixations during the first search, except for the blue cylinder. This object differed significantly from the other objects and required 5.6 additional fixations to be found (see Figure 11). These results show that new, previously irrelevant, objects were found faster than objects in Phase 1 (see Movie 2 for an example, comparing new and old objects). The only exception was the blue cylinder, which took *significantly longer* than objects in Phase 1 to be found (Movie 3 shows an example).

We also analyzed the relation between *incidental fixations* during Phase 1 and search during Phase 2 (examples of incidental fixations can be seen in Movie 1). There was a high variability in the number of fixations that each participant made on each object (Figure 12), but there was a significant negative correlation ($R = -0.362$, $p = 0.042$) between incidental fixations and search fixations, indicating that those objects that received more fixations during Phase 1 required fewer fixations to be found during Phase 2 (Figure 13). The only object that practically did not receive any fixations during Phase 1 was the blue cylinder, which was partially hidden by nearby objects.

## Spatial learning and locomotion

There were differences in locomotion over trials and repetitions depending on the kind of target object. This was clear when comparing the first and the last search for the different objects. These results are shown in Figures 14, 15, 16 and 17. The four figures reflect some interesting patterns regarding body movements in the apartment.

First, searching repeatedly for the same objects during Phase 1 leads to both a shortening of trajectory length (see Figure 14) and an increase in movement

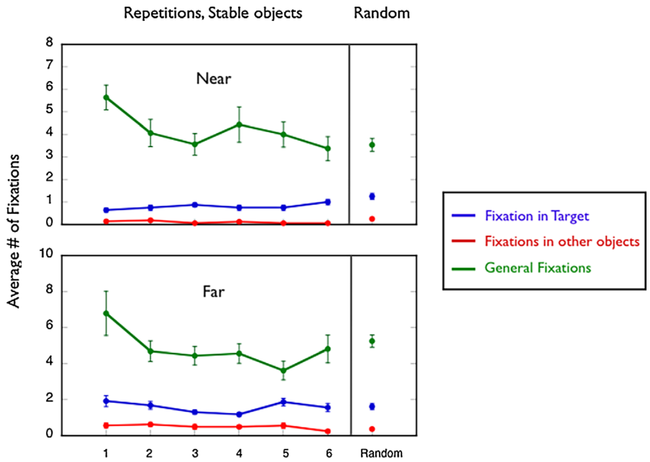


Figure 8. Distribution of fixations according to content. Average number of general fixations (green), fixations on the target object (blue) and fixations on other objects (red) are plotted for each repetition of reliable objects (left) and for random objects (right), separately depending on location. Error bars represent SEM. Data from eight participants.

velocity (see Figure 15). A LMM was used to obtain an estimation of the variation for both variables (see Tables S6 and S7). For all objects the length of the trajectory was significantly *shorter* and velocity was significantly *higher* after 6 repetitions (left part of Figure 17). The length of the trajectories decreased by about 15–25 cm, and average velocities increased by 0.05 to 0.20 m/sec. As expected, given the distribution of objects in the apartment, objects at the far location showed longer trajectories and also higher velocities. These results suggest that knowledge about object location led to an optimization of movement performance: trajectory lengths were adjusted to the task requirements (minimum distance needed to press remote button) and body movements increased velocity (see Movie 1).

Regarding Phase 2, it is very interesting that trajectories also show very direct paths to the new target objects in all cases, except for the blue cylinder (Figure 14). Considering that most objects received incidental fixations during Phase 1 this suggests that object location information was used to optimize movement control (see Movie 2). The only exception is the blue cylinder. Only in this case participants seem to walk straight, look around and then turn towards the object, as seen in the average trajectories shown in Figure 14 (see also Movie 3). The other kinematic variables show results that are also coherent with this interpretation. Average velocities show very similar profiles for all previously irrelevant objects (Figure 16), which are very similar to those obtained after six repetitions of the old objects and also do not vary much between the first and the second repetition. Results for length of trajectory and peak velocity also show little improvement between the two repetitions (right part of Figure 17). In the case of the blue cylinder, however, the average velocity profile shows a double peak on the first search (red line).[3] This might be indicating the need to stop or slow down once in the room until the target object was localized.

Because random objects changed location in each trial, it was not possible to analyze if repetition had an effect on the kinematics of these movements. For illustration purposes the average trajectories in each of the searches for these objects are presented in Figure 14 (upper right, blue lines). We also checked whether peak velocities showed variation over multiple presentations of these objects. Except very occasionally, peak velocities for random objects at both locations had very constant values.

## General discussion

The goal of this experiment was to evaluate the use of spatial memory and its effects at the motor level in a naturalistic visual search task. We summarize the main findings below.

### General and object-specific learning in the apartment

We found that experience led to a gradual reduction in the time needed to complete the task and a decrease in the number of errors (visits to the incorrect room). Over trials, participants chose the correct room more often and were faster in determining which room was correct. Once in the correct room search time decreased with experience. This resulted from the combination of two effects: learning how to move in the apartment, and learning target locations. Our results allowed us to estimate the size of both components. For reliable objects we estimated that search times decreased about three seconds after searching six times for a given object, while for objects that changed location we estimated an improvement of less than two seconds. This indicates that part of the reduction in search times is linked to the ability to remember the positions of the objects. These results replicate our previous findings (Li et al., 2016; Li et al., 2018).

### Changes in exploration strategies over trials

We found that general fixations decreased in number quickly over the first trials of the experiment, while fixations on the target object (one or two per trial) or on other objects did not vary much. This pattern

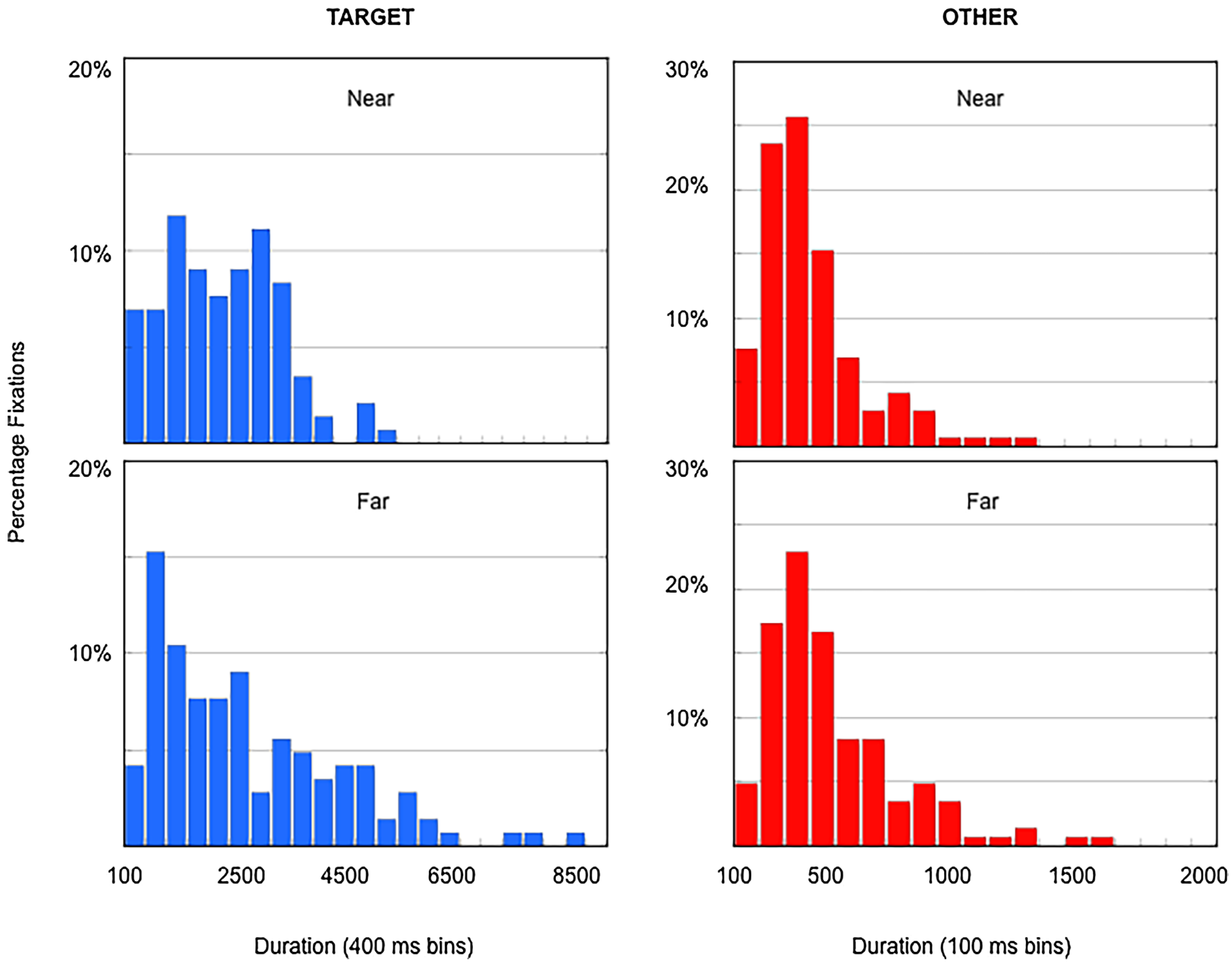


Figure 9. Distribution of fixation durations. Histograms represent the distribution of fixation durations, over trials and participants, for those fixations on the target object (blue) and those that occurred everywhere else (red), grouped depending on the location of the target (near/far). Attention should be given to the x axis and the different size bins used for both kinds of fixations. Data from eight participants.

suggests that improvements in visual search with more experience occur through a reduction of fixations on irrelevant objects or areas of the scene (as in Li et al., 2018). Once participants learn object locations, locomotion is directly guided towards the relevant object in each trial, as shown by the long duration of target fixations. Still, it is important to point out that in this experiment changes in search time and number of fixations occurred very gradually after the first block of trials, which suggests a slower learning rate than seen in our previous studies (Kit et al., 2014; Li et al., 2016; Li et al., 2018). This could be due to the controlled exposure to target objects in this experiment (only visible when they were targets). Experience in the apartment also affected locomotion. Our results showed that for all reliable objects the lengths of the trajectories significantly decreased and the velocity of the movements significantly increased from the first to the sixth repetition. This effect was modulated by object position in the room (near versus far). These results extend our previous findings and suggest that over trials participants learn to optimize their body movements to be able to complete the task in a faster and more effective way. The weight of the helmet and the characteristics of the experimental setup very likely contributed to participants trying to speed up their movements as well and might have reduced head movements, thus decreasing gaze fixations. This is relevant regarding the extensibility of our results to other situations.

## Role of incidental fixations

Our experiment also demonstrated that attention to previously irrelevant objects aided search. The previously irrelevant objects that received fixations during Phase 1 were found significantly faster when

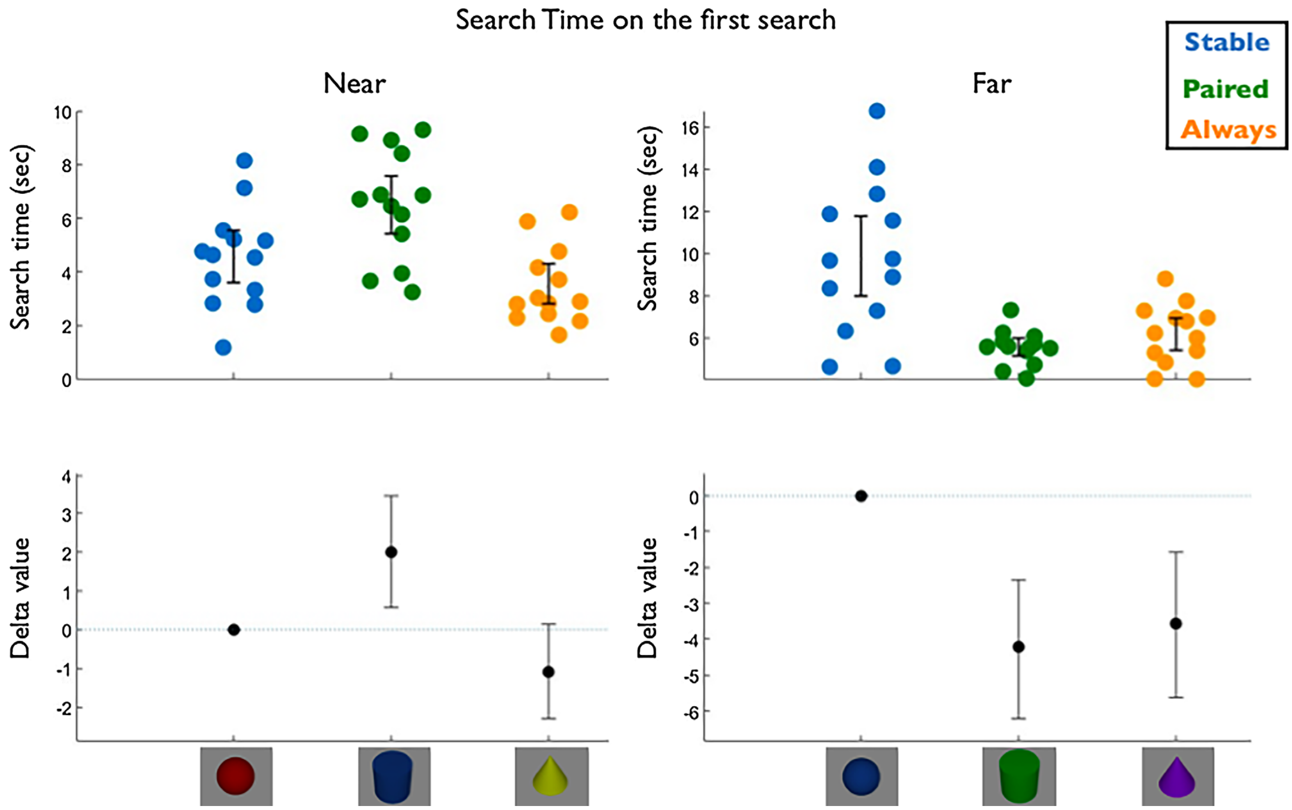


Figure 10. Search Time during Phase 2 for previously irrelevant objects. Raw data and averages are shown using a Cumming estimation plot for stable (blue), paired (green) and always visible (yellow) objects at near (left) and far (right) locations. On the lower axes mean differences are plotted as bootstrap sampling distributions. Dots depict mean differences and the vertical error bars indicate 95% confidence intervals. The first time participants searched for the red and blue ball (Phase 1) are used as reference. Data from 13 participants.

they appeared as targets in Phase 2. Performance for these new objects on the first search was very similar to that for old objects after six repetitions and significantly different from performance on the first presentation of each of the old objects. Moreover, when we related the number of fixations during search with the number of previous incidental fixations on the new objects, we found a significant negative correlation: those objects that received more fixations during Phase 1 required fewer fixations to be found during Phase 2. We interpret this result as showing that participants were able to remember the locations of the different geometrical objects, even though they were not instructed to do so and did not know that these objects would be tested.

The clear relation between incidental fixations and search facilitation that we found in this study helps clarify our previous inconclusive results (Li et al., 2016). In that study participants had to find contextual objects after 24 trials in which all target objects were geometric. Although contextual objects had received some incidental fixations, search efficiency did not improve, which suggested almost no effect of previous incidental fixations. However, in that study contextual objects varied in size and location. It is also likely that participants restricted their attention to geometric objects, since the number of incidental fixations on contextual objects varied a lot between participants and was very low overall (one to three incidental fixations per object were most common). By reducing the number of objects tested and making them similar to other target objects in the new experiment we managed to increase their attentional relevance. Therefore our new results do more clearly highlight the important role of incidental fixations in facilitating search.

## Memory effects on locomotion

Movement trajectories for previously irrelevant objects also provide information on how memory optimizes movement control. The first time they were presented as targets new objects showed very direct trajectories, and their velocity profiles were also very similar to those seen for the old objects after six search episodes. Moreover, these trajectories did not show much change between the first and the second search

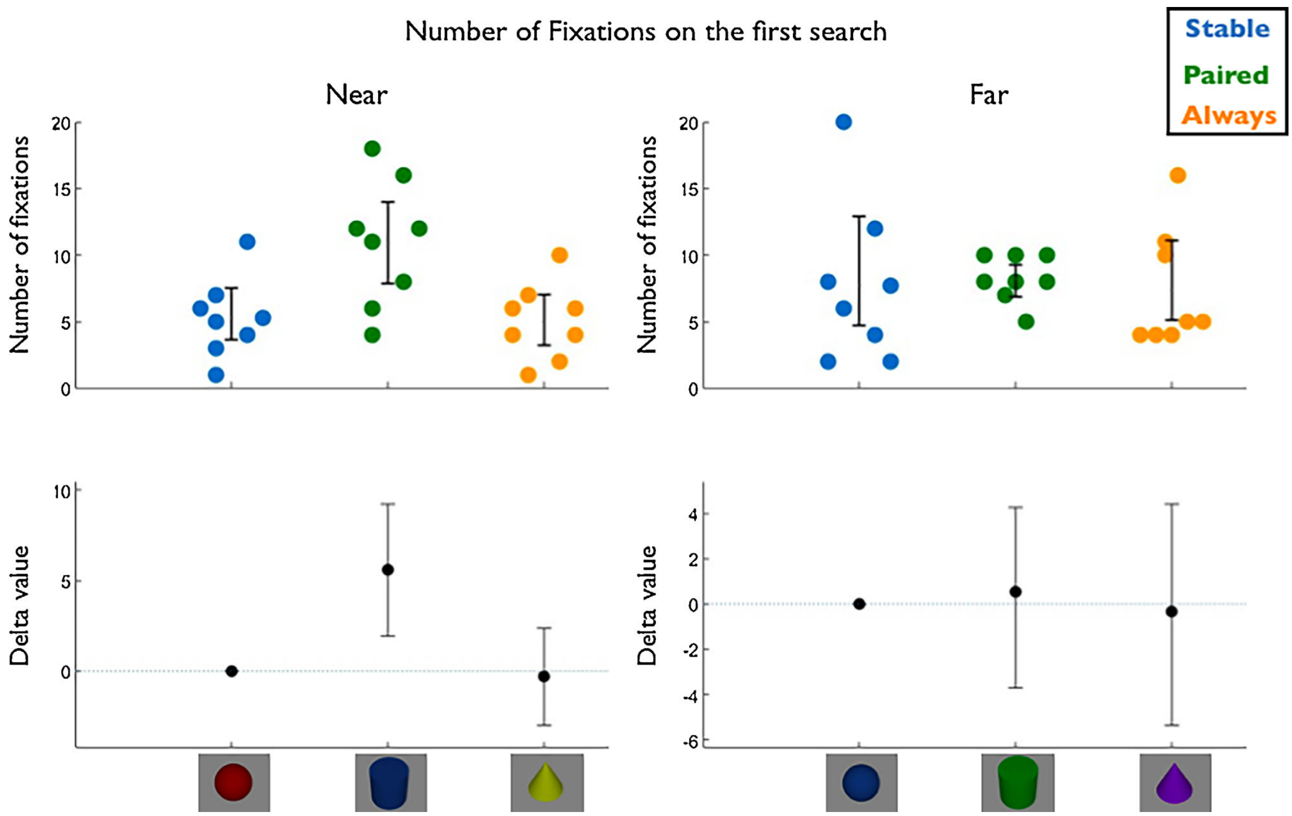


Figure 11. Number of fixations during Phase 2 for previously irrelevant objects. Raw data and averages are shown using a Cumming estimation plot for stable (blue), paired (green) and always visible (yellow) objects at near (left) and far (right) locations. Plots follow the same structure as those in Figure 10. Data from eight participants.

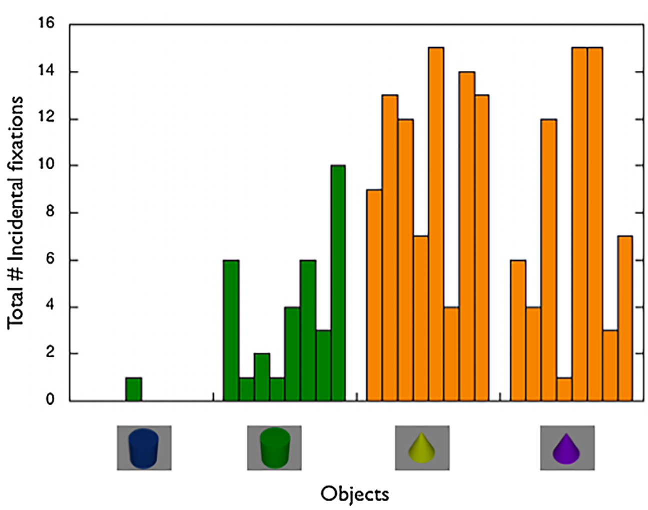


Figure 12. Total number of incidental fixations on each of the *paired* (green) and *always visible* (orange) objects during Phase 1. Each column shows the total number of fixations for object and participant, over all trials of Phase 1. Data from eight participants.

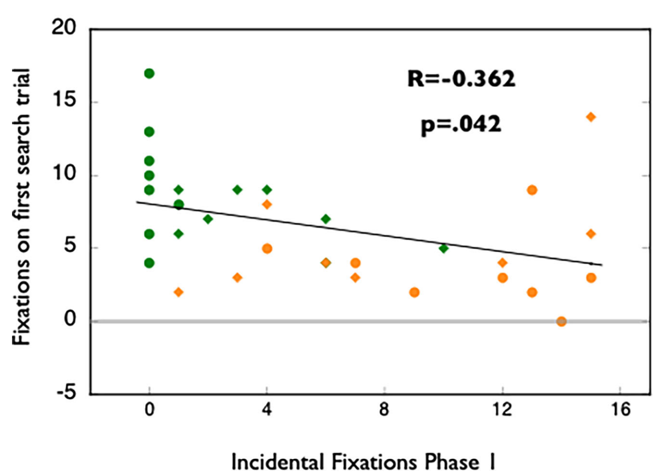


Figure 13. Number of search fixations to find each new object in Phase 2 as a function of the number of incidental fixations on that object during Phase 1. Each target object is represented with a different color and symbol (paired objects in green, always visible in orange). Incidental fixations during Phase 1 occurred while looking for other target objects. Number of search fixations during Phase 2 excludes fixations on the target object once found. Data from eight participants.

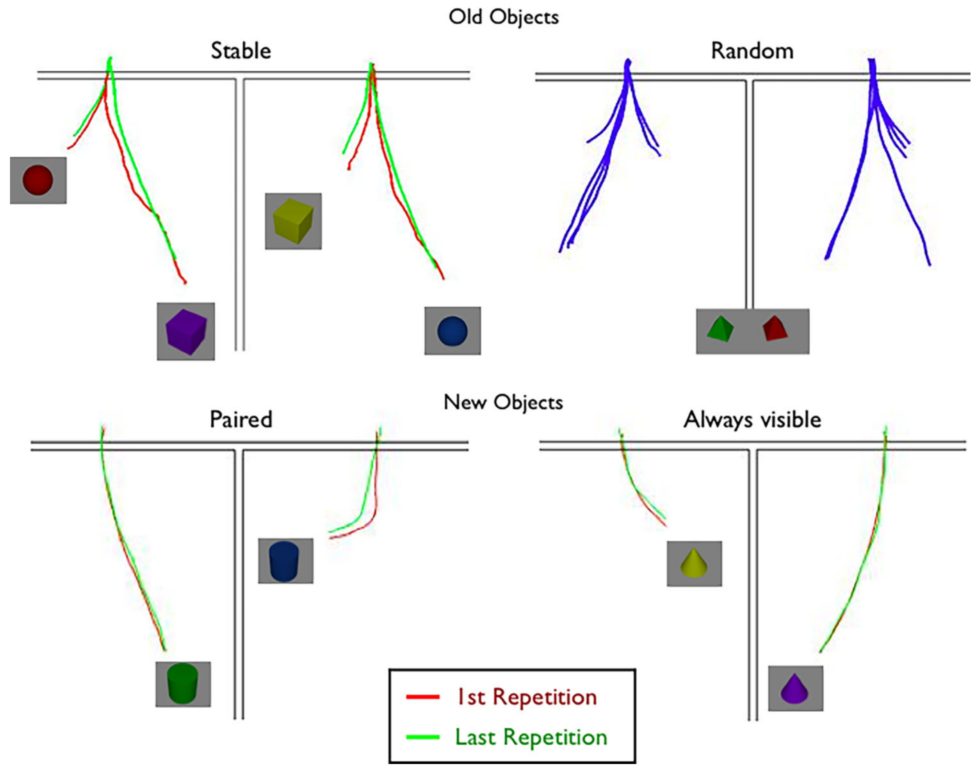


Figure 14. Average body trajectories when searching for each of the target objects. Top part shows averages for each of the objects presented during Phase 1 (stable and paired objects on the left, random on the right), and bottom part shows averages for the new objects presented in Phase 2 (paired on the left and always visible on the right). Black lines represent the general structure of the apartment (corridor and different rooms). Red and green lines represent first and last search for those objects in each phase of the experiment. Random objects are represented in blue since they changed location in each trial. Data from 13 participants.

during Phase 2. This is radically different from what we see for the six repetitions of the old objects, which reflect a gradual change in movement kinematics. Therefore we conclude that location information about the new target objects was acquired during the first phase of the experiment and used to effectively guide locomotion in the apartment during the second phase.

The only exception was the blue cylinder. The first time this object appeared as target during Phase 2, velocity profiles *differed* from those seen for the other new objects: they showed a double peak, which corresponds to a movement with two phases. Average trajectories suggest that in this case participants seem to walk straight to the center of the room, look around to find the object and then complete the movement towards it once it is found. This also led to a significantly higher number of fixations during search and the longest search times. It seems likely that this object was (accidentally) very inconspicuous.

## Visual search under natural conditions

What, then, can we conclude about how visual search proceeds in natural situations? First, it seems clear that search strategies in our virtual apartment differed considerably from those seen in monitor-based search tasks. The design of our task gave locomotion a central role, and this seems to have constrained visual search greatly. Our participants were able to quickly learn where the stable objects were, which led to fewer visits to the incorrect room and faster searches once in the correct room. The design of the apartment also provided strong semantic cues and multiple anchor objects (like tables or dressers), which might have also facilitated search (Boettcher, Draschkow, Dienhart & Võ, 2018; Helbing, Draschkow & Võ, 2022; Josephs, Draschkow, Wolfe & Võ, 2016; Pereira & Castelhano, 2019). Visual search per se proceeded very quickly, with very few visual fixations on the objects in the environment once object locations were known (between four and six total fixations on average, and only one or two of those were on the target object). That is, participants seem to have switched to a memory-based strategy, instead of repeatedly searching the environment. The cumbersome equipment used might have helped in this process as well. This is in contrast to what has been found on computer-based search tasks (Horowitz & Wolfe, 1998) or on more simple

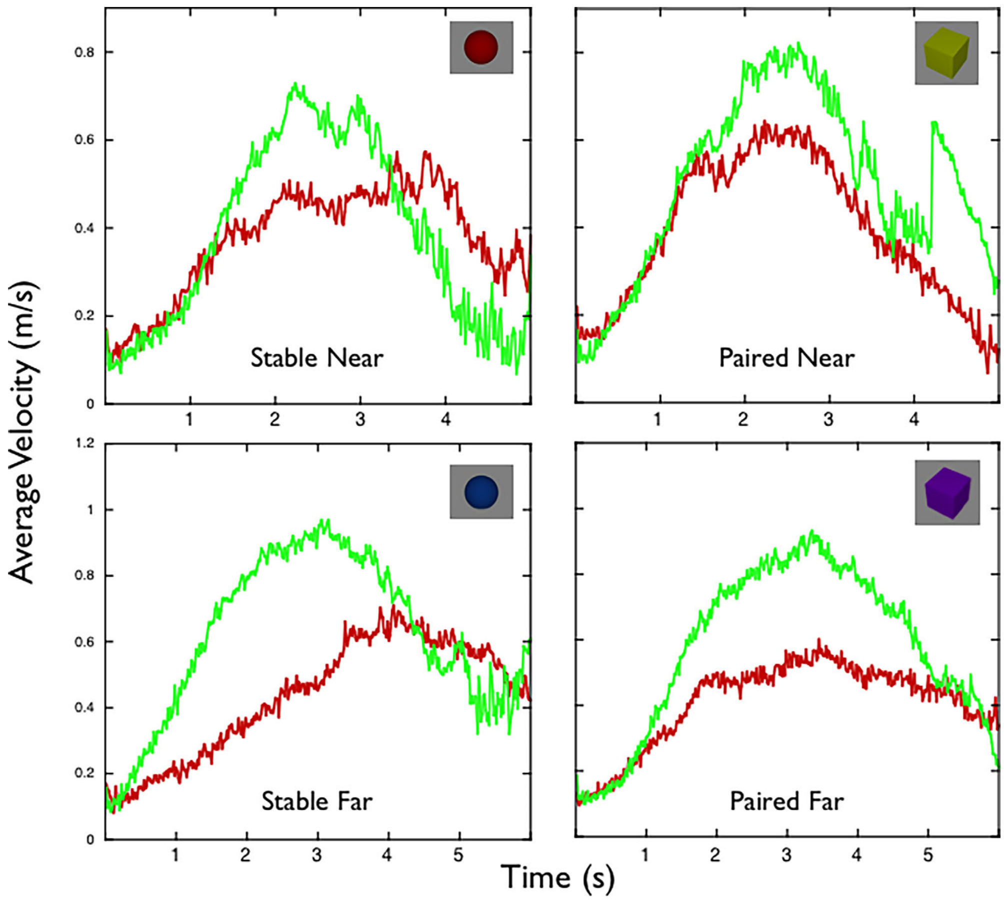


Figure 15. Average velocity profiles of the body movement for objects presented during Phase 1 (stable objects on the left, paired objects on the right). Red lines represent averages for the first search and green lines averages for the last (sixth) search, for each of the objects. Data from 13 participants.

manipulative tasks (Hayhoe, Bensinger & Ballard, 1998).

Natural behavior also seems to change the degree of attention given to other, possible irrelevant, objects in the environment. It is known that incidentally encoded information can lead to reliable memory, especially when natural scenes are used (Castelhano & Henderson, 2005; Hollingworth & Henderson, 2002; Olejarczyk, Luke & Henderson, 2014; Williams, Henderson & Zacks, 2005) or when performing natural actions (Tatler & Tatler, 2013). Recent results also suggest a bigger role of this kind of memory to optimize visual search in more ecologically valid settings (Draschkow, Wolfe & Vö, 2014; Guevara-Pinto et al., 2020; Helbing, Draschkow & Vö, 2020; Kristjánsson & Draschkow, 2021). Our experiment supports this claim. Incidental memory effects were manifest very clearly during Phase 2. This result suggests that participants can extract and retain detailed information about the properties of the objects in the environment, even those that are not relevant for the task. In our experiment this was facilitated by using objects from the same general category (geometric objects) and by placing them nearby other target objects, providing a local context for the search (Mack & Eckstein, 2011). It is also important to point out that our participants did not have explicit instructions requiring memorization of target objects or the virtual environment: knowledge about the different objects and their locations was acquired incidentally. This, paradoxically, seems to increase location memory (see Helbing et al., 2020).

If we combine our results regarding movement trajectories with those obtained for eye fixations an interesting picture emerges regarding how visual search evolves over time. Over the first trials of the experiment participants seem to learn the possible locations of objects in the apartment (relevant surfaces or *anchors*). From that point on, they concentrate their attention mostly on those areas, which speeds up search. With more experience participants were even able to locate the target object and fixate it from the door, even before entering the room. This mostly occurred for objects at reliable locations. The target object then was clearly used as the *heading point*, while other objects in the room were rarely fixated. This explains the unusually long fixations on the target, which

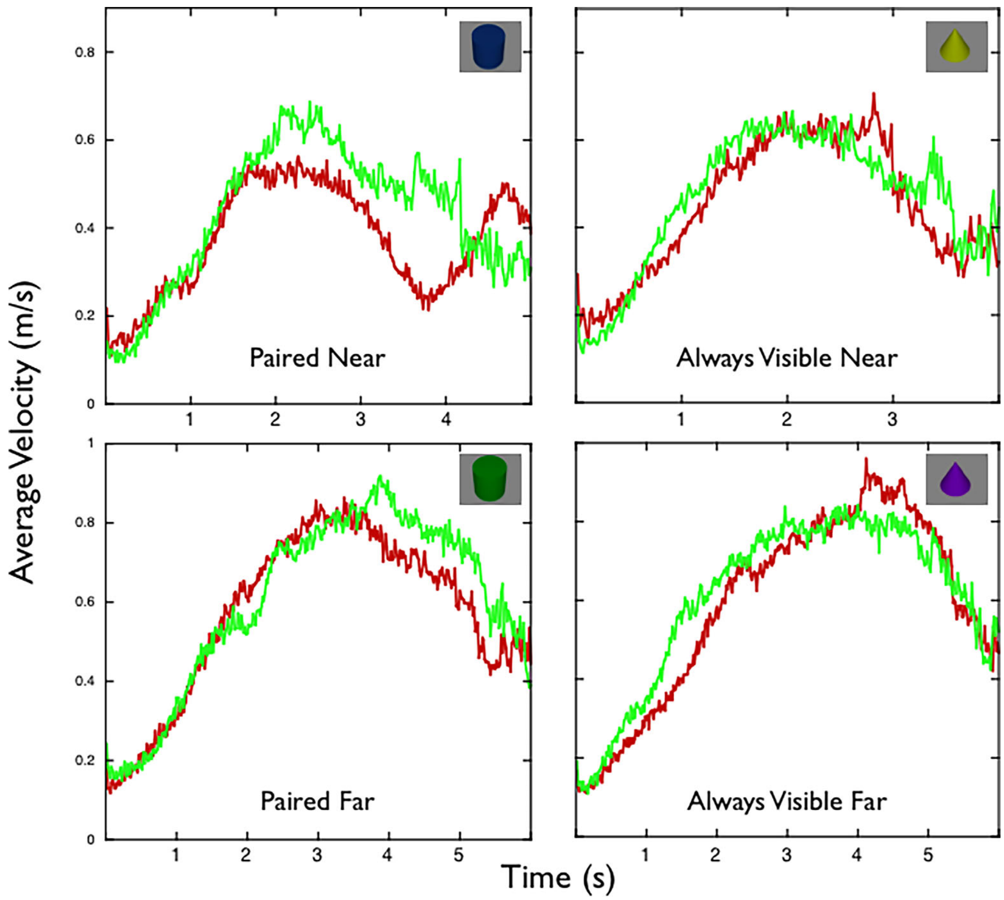


Figure 16. Average velocity profiles of the body movement for objects presented during Phase 2 (paired object on the left, always visible objects on the right). Red and green lines represent first and second search for each of the objects. Data from 13 participants.

were even longer when the object was placed at the far location, replicating previous results (Foulsham, Walker & Kingstone, 2011). This is also consistent with trajectories being very direct when object location was known, and average velocities increasing over trials. Fixations on other objects and the walls were more frequent during the first trials but decreased with experience. These fixations were also much shorter, with durations very similar to those reported by Enders et al. (2021). Our results also seem consistent with those obtained by Thom, Nobre, van Ede, and Draschkow (2023), who found a bias in heading direction towards memorized target locations. In our experiment we could not directly determine head orientation, but the video recordings suggest that participants often oriented themselves towards the target object, even before the door of the room opened, when its location was known.

## Attentional selection and memory in the control of actions

Overall, our results support the idea that task constraints differentially modulate visual search strategies, affecting how visual information is extracted and retained during the task (Hayhoe, 2009). When the task requires more body movements, vision seems to prioritize what is needed for motor control (Brenner & Smeets, 2022). In this context, fixations are restricted to those objects that are task-relevant, and there is tight spatial and temporal coupling with the actions (Hayhoe, 2017). For example, when making tea or sandwiches, gaze shows different patterns depending on whether visual information is needed to locate objects, direct an action, guide a process or check the situation (Land & Hayhoe, 2001). Our results are in line with these claims: participants' fixations were restricted to the target objects and their most likely locations and were tightly linked to the control of locomotion.

Spatial memory about the environment also seems to have an important role for optimizing movement control. To be able to move effectively, all effectors need to share a common coordinate system (Lappi, 2016). They also need to remain synchronized in time across multiple actions (Bekkering & Sailer, 2002). But latencies for head, hand, and body movements are longer than for eye movements, and the kinematic

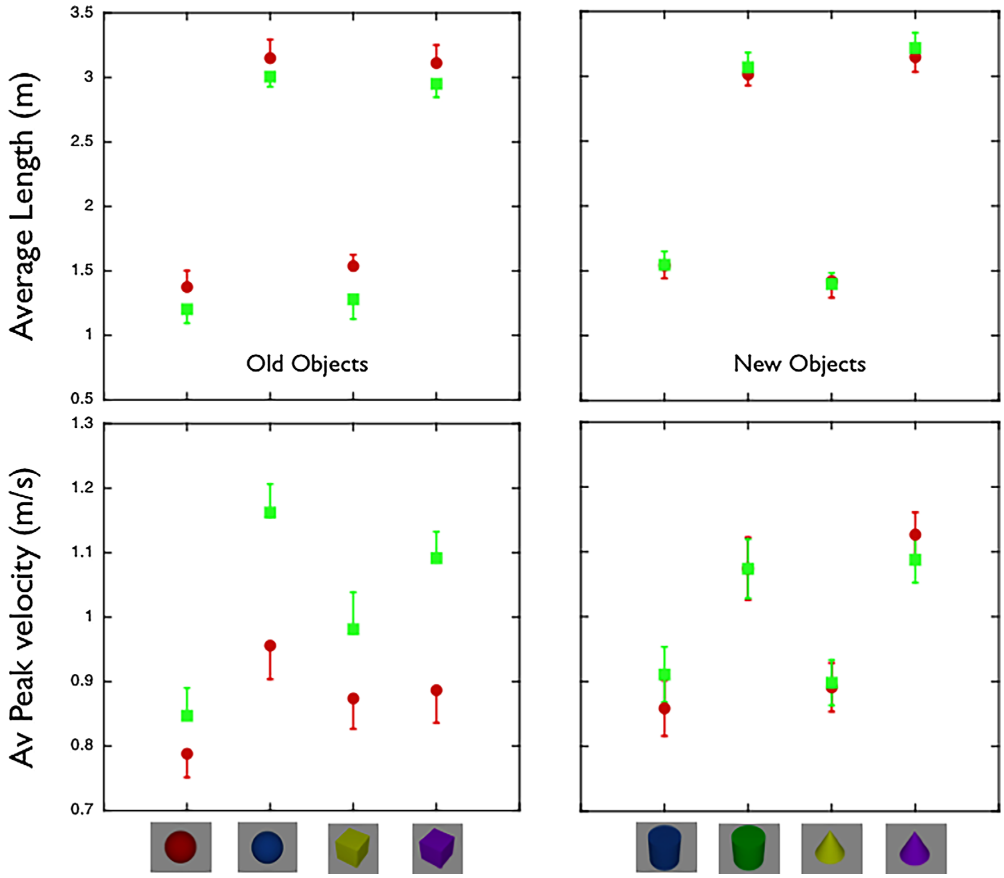


Figure 17. Variations in trajectory length (top) and peak velocity (bottom) depending on number of repetitions in each phase. Left side shows values for objects presented in Phase 1 and right side for objects presented in Phase 2. Values from the first search trial are shown in red, and those from the last search trial of that phase are shown in green. Error bars represent SEM. Data from 13 participants.

properties of the different effectors also differ. This is a problem for the motor system, which is often solved using gaze as an *anchor* for the control of hand (Foerster, 2019) or body movements (Lappi, 2016; Wann & Swapp, 2000). In this context, maintaining a memory representation of space facilitates motor coordination (Hayhoe, 2009). It allows optimal and accurate direction of gaze during complex actions, even to regions outside the field of view (Aivar, Hayhoe, Chizk & Mruczek, 2005; Land, Mennie & Rusted, 1999). The use of previously acquired knowledge also allows anticipation, which is more important for tasks with higher motor complexity (more effectors) or longer planning spans, like those implying locomotion (Bernardin et al., 2012; Higuchi, 2013; Patla, 1997). In those cases, gaze can anticipate future needs with time intervals on the order of seconds (Hayhoe, Shrivastavah, Mruczek & Pelz, 2003; Mennie, Hayhoe & Sullivan, 2007; Pelz & Canosa, 2001; Scrafton et al., 2017; Sullivan, Ludwig, Damen, Mayol-Cuevas & Gilchrist, 2021). Therefore precise location information about relevant objects, acquired in previous trials, provides a common framework that facilitates motor coordination. Our results support this claim: Once target locations were known, participants could use the target object as the heading point and this led to more direct trajectories and faster movements in the apartment.

While virtual environments involve many of the features of natural behavior, features such as the weight of the helmet and the more restricted field of view no doubt introduce different constraints. Peripheral retinal information, for example, probably plays a more important role than is often assumed in both visual search (David, Beitner & Võ, 2021) and the control of locomotion (Cao & Händel, 2019). This issue needs to be explored in more depth. Another consideration is the possible conflict between target colors and location reliability. As explained above, all target colors were used for *two* different objects, which means that color-location associations were not unique. For example, we presented a red ball at a reliable location, but also a red pyramid at unreliable locations. According to the BRAC framework (Frings et al., 2020), this could have interfered with feature binding and retrieval and affected action control. In fact, previous studies have shown a *mixing cost* when the object used to perform a complex action (grasping) has different feature combinations from trial to trial (Beyvers, Koch & Fiehler, 2022). Best task performance occurs when object features are unique and do not vary. This means that performance could have been partially impaired in our experiment, because one feature was repeated (i.e., color) whereas other object features (i.e., shape, location) changed. Spatial memory for target objects might improve when unique features are associated with each object.

To conclude, this study supports the idea that complex natural tasks show a trade-off between using currently available visual information and memory resources (Draschkow et al., 2021; Hayhoe, 2017). It also points to the relevance of *motor costs* in understanding the characteristics of visually guided action (Christie & Schrater, 2015; Hayhoe, 2017). When reliable location information is present in memory, motor commands can be based on this information, which reduces the need for sensory updates. Our results also agree with the claim that attentional selection processes are modulated by action intentions (Moskowitz et al., 2023; Park, Ahn & Zhang, 2021). This issue deserves further exploration, using more varied and complex tasks. Virtual Reality might be an excellent tool to test specific hypothesis regarding the effects of locomotion and different kinds of actions on attention, cognitive processes and memory.

*Keywords: eye fixations, visual search, visual memory, virtual reality, natural environments*

## Acknowledgments

Research supported by grant NIH EY05729, funded by the National Institutes of Health, and by grants PSI2013-43742 and PID2021-125162NB-I00, funded by MCIN/AEI/10.13039/501100011033 (Spain) and by “ERDF A way of making Europe” (EU). MPA supported by a UAM-Banco Santander Inter-University Cooperation Project and by a “José Castillejo” International Mobility Grant (Ministerio de Educación, Cultura y Deporte, Programa Nacional de Movilidad de Recursos Humanos del Plan Nacional de I+D+i 2008-2011).

Commercial relationships: none.
Corresponding author: M. Pilar Aivar.
Email: mariapilar.aivar@uam.es.
Address: Departamento de Psicología Básica, Facultad de Psicología, Universidad Autónoma de Madrid, Madrid 28049, Spain.

## Footnotes

[1]We are aware that this selection of participants is not ideal, but the limitations imposed by the system used (no eye corrections allowed) and the fact that motion sickness developed very easily in this setup made it really hard to obtain data from a more diverse group of participants.
[2]These delays probably contributed to motion sickness, especially once participants started moving more quickly inside the apartment.
[3]A check of the individual velocity profiles showed that this double peak did not really appear for all participants, only for a few.

## References

Aivar, M. P., Hayhoe, M. M., Chizk, C., & Mruczek, R. E. B. (2005). Spatial memory and saccadic targeting in a natural task. *Journal of Vision, 5*(3), 177–193.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*(1), 66–80.

Bekkering, H., & Sailer, U. (2002). Commentary: Coordination of eye and hand in time and space. In J. Hyönä, D. P. Munoz, W. Heide & R. Radach (Eds.), *Progress in Brain Research* (Vol. *140*, pp. 362–373). Elsevier.

Bernardin, D., Kadone, H., Bennequin, D., Sugar, T., Zaoui, M., & Berthoz, A. (2012). Gaze anticipation during human locomotion. *Experimental Brain Research, 223*, 65–78.

Beyvers, M. C., Koch, I., & Fiehler, K. (2022). Episodic binding and retrieval in sequences of discrete movements - evidence from grasping actions. *Journal of Cognition, 5*(1), 41.

Binsted, G., Chua, R., Helsen, W. F., & Elliott, D. (2001). Eye-hand coordination in goal-directed aiming. *Human Movement Science, 20*, 563–585.

Boettcher, S. E. P., Draschkow, D., Dienhart, E., & Vö, M. L. H. (2018). Anchoring visual search in scenes: Assessing the role of anchor objects in eye movements during visual search. *Journal of Vision, 18*(13), 11.

Brenner, E., & Smeets, J. B. J. (2022). Spatial vision for action. In O. Braddick (Ed.), *Oxford Research Encyclopedia of Psychology*. Oxford: Oxford University Press.

Brown, V. A. (2020). An introduction to linear mixed effects modeling in R. *PsyArXiv, April 11*.

Cao, L., & Händel, B. (2019). Walking enhances peripheral visual processing in humans. *PLoS Biology, 17*(10), e3000511.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*, 1484–1525.

Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition, 12*(6), 1017–1040.

Chrastil, E. R., & Warren, W. H. (2012). Active and passive contributions to spatial learning. *Psychonomic Bulletin and Review, 19*, 1–23.

Christie, S. T., & Schrater, P. (2015). Cognitive cost as dynamic allocation of energetic resources. *Frontiers in Neuroscience, 9*, 289.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4*(5), 170–177.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*, 28–71.

Conci, M., Zellin, M., & Müller, H. J. (2012). Whatever after next? Adaptive predictions based on short- and long-term memory in visual search. *Frontiers in Psychology, 3*, 409.

David, E. J., Beitner, J., & Võ, M. L. H. (2021). The importance of peripheral vision when searching 3D real-world scenes: A gaze-contingency study in virtual reality. *Journal of Vision, 21*(7), 3.

Draschkow, D., Kallmayer, M., & Nobre, A. C. (2021). When natural behavior engages working memory. *Current Biology, 31*, 869–874.

Draschkow, D., & Võ, M. L. H. (2017). Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Scientific Reports, 7*, 16471.

Draschkow, D., Wolfe, J. M., & Võ, M. L. H. (2014). Seek and you shall remember: Scene semantics interact with visual search to build better memories. *Journal of Vision, 14*(8), 10.

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision, 11*(5), 14.

Enders, L. R., Smith, R. J., Gordon, S. M., Ries, A. J., & Touryan, J. (2021). Gaze behavior during navigation and visual search of an open-world virtual environment. *Frontiers in Psychology, 12*, 681042.

Foerster, R. M. (2019). The function of "looking-at-nothing" for sequential sensorimotor tasks: Eye movements to remembered action-target locations. *Journal of eye movement research, 12*(2), https://doi.org/10.16910/jemr.16912.16912.16912.

Foulsham, T., Chapman, C., 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.

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research, 51*(17), 1920–1931.

Franchak, J. M., McGee, B., & Blanch, G. (2021). Adapting the coordination of eyes and head to differences in task and environment during fully-mobile visual exploration. *PLoS ONE, 16*(8), e0256463.

Frings, C., Hommel, B., Koch, I., Rothermund, K., Dignath, D., Giesen, C., . . . Philipp, A. (2020). Binding and retrieval in action control (BRAC). *Trends in Cognitive Sciences, 24*(5), 375–387.

Guevara-Pinto, J. D., Papesh, M. H., & Hout, M. C. (2020). The detail is in the difficulty: Challenging search facilitates rich incidental object encoding. *Memory and Cognition, 48*, 1214–1233.

Hardiess, G., Gillner, S., & Mallot, H. A. (2008). Head and eye movements and the role of memory limitations in a visual search paradigm. *Journal of Vision, 8*(1), 7.

Hayhoe, M. M. (2009). Visual memory in motor planning and action. In J. R. Brockmole (Ed.), *The Visual World in Memory* (pp. 117–139). East Sussex: Psychology Press.

Hayhoe, M. M. (2017). Vision and action. *Annual Review of Vision Science, 3*, 389–413.

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research, 38*(1), 125–137.

Hayhoe, M. M., Shrivastavah, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*(1), 49–63.

Helbing, J., Draschkow, D., & Võ, M. L. H. (2020). Search superiority: Goal-directed attentional allocation creates more reliable incidental identity and location memory than explicit encoding in naturalistic virtual environments. *Cognition, 196*, 1041–1047.

Helbing, J., Draschkow, D., & Võ, M. L. H. (2022). Auxiliary scene-context information provided by anchor objects guides attention and locomotion in natural search behavior. *Psychological Science, 33*(9), 1463–1476.

Higuchi, T. (2013). Visuomotor control of human adaptive locomotion: Understanding the anticipatory nature. *Frontiers in Psychology, 4*, 277.

Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond P values: Everyday analysis with estimation plots. *Nature Methods, 16*, 565–566.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology, HPP, 28*(1), 113–136.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature, 394*, 575–577.

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

Hulleman, J., & Olivers, C. N. (2017). The impending demise of the item in visual search. *Behavioral and Brain Sciences, 40*, 1–69.

Jiang, Y. V., Won, B., Swallow, K. M., & Mussack, D. M. (2014). Spatial reference frame of attention in a large outdoor environment. *Journal of Experimental Psychology, Human Perception and Performance, 40*(4), 1346–1357.

Josephs, E. L., Draschkow, D., Wolfe, J. M., & Vö, M. L. H. (2016). Gist in time: Scene semantics and structure enhance recall of searched objects. *Acta Psychologica, 169*, 100–108.

Kit, D., Katz, L., Sullivan, B., Snyder, K., Ballard, D. H., & Hayhoe, M. M. (2014). Eye movements, visual search and scene memory, in an immersive virtual environment. *PLoS ONE, 9*(4), e94362.

Kristjánsson, A., & Draschkow, D. (2021). Keeping it real: Looking beyond capacity limits in visual cognition. *Attention, Perception and Psychophysics, 83*(4), 1375–1390.

Kunar, M. A., Flusberg, S., & Wolfe, J. M. (2008). The role of memory and restricted context in repeated visual search. *Perception and Psychophysics, 70*(2), 314–328.

Lachaud, C., & Renaud, O. (2011). A tutorial for analyzing human reaction times: How to filter data, manage missing values, and choose a statistical model. *Applied Psycholinguistics, 32*(2), 389–416.

Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*, 3559–3565.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*, 1311–1328.

Land, M. F., & Tatler, B. W. (2009). *Looking and Acting: Vision and eye movements in natural behavior.* Oxford: Oxford University Press.

Lappi, O. (2016). Eye movements in the wild: Oculomotor control, gaze behavior and frames of reference. *Neuroscience and Biobehavioral Reviews, 69*, 49–68.

Li, A., Chen, Z., Wolfe, J. M., & Olivers, C. N. L. (2023). How do people find pairs? *Journal of Experimental Psychology: General, 152*(8), 2190–2204.

Li, C. L., Aivar, M. P., Kit, D. M., Tong, M. H., & Hayhoe, M. M. (2016). Memory and visual search in naturalistic 2D and 3D environments. *Journal of Vision, 16*(8), 9.

Li, C. L., Aivar, M. P., Tong, M. H., & Hayhoe, M. M. (2018). Memory shapes visual search strategies in large-scale environments. *Scientific Reports, 8*(4324).

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

Mennie, N., Hayhoe, M., & Sullivan, B. (2007). Look-ahead fixations: anticipatory eye movements in natural tasks. *Experimental Brain Research, 179*, 427–442.

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Languaje, 112*(104092), 1–22.

Moskowitz, J. B., Berger, S. A., Castelhano, M. S., Gallivan, J. P., & Flanagan, J. R. (2022). The influence of movement-related costs when searching to act and acting to search. [Electronic Version]. *bioRxiv* from https://doi.org/10.1101/2022.03.01.482521.

Moskowitz, J. B., Fooken, J., Castelhano, M. S., Gallivan, J. P., & Flanagan, J. R. (2023). Visual search for reach targets in actionable space is influenced by movement costs imposed by obstacles. *Journal of Vision, 23*(6), 4.

Nachtnebel, S. J., Cambronero-Delgadillo, A. J., Helmers, L., Ischebeck, A., & Höfter, M. (2023). The impact of different distractions on outdoor visual search and object memory. *Scientific Reports, 13*(1), 16700.

Olejarczyk, J. H., Luke, S. G., & Henderson, J. M. (2014). Incidental memory for parts of scenes from eye movements. *Visual Cognition, 22*(7), 975–995.

Over, E. A. B., Hooge, I. T. C., Vlaskamp, B. N. S., & Erkelens, C. J. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Research, 47*(17), 2272–2280.

Park, H., Ahn, S., & Zhang, W. (2021). Visual search under physical effort is faster but more vulnerable to distractor interference. *Cognitive Research: Principles and Implications, 6*, 17.

Patla, A. E. (1997). Understanding the roles of vision in the control of human locomotion. *Gait and Posture, 5*, 54–69.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research, 41*, 3587–3596.

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

Pereira, E. J., & Castelhano, M. S. (2019). Attentional capture is contingent on scene region: Using surface guidance framework to explore attentional mechanisms during search. *Psychonomic Bulletin and Review, 26*, 1273–1281.

Sailer, U., Eggert, T., Ditterich, J., & Straube, A. (2000). Spatial and temporal aspects of eye-hand coordination across different tasks. *Experimental Brain Research, 134*, 163–173.

Scrafton, S., Stainer, M. J., & Tatler, B. W. (2017). Coordinating vision and action in natural behavior: Differences in spatiotemporal coupling in everyday tasks. *Canadian Journal of Experimental Psychology, 71*(2), 133–145.

Seltman, H. J. (2012). *Experimental Design and Analysis*. Pittsburgh: Carnegie Mellon University.

Solman, G. J. F., & Kingstone, A. (2014). Balancing energetic and cognitive resources: Memory use during search depends on the orienting effector. *Cognition, 132*, 443–454.

Solman, G. J. F., & Smilek, D. (2012). Memory benefits during visual search depend on difficulty. *Journal of Cognitive Psychology, 24*(6), 689–702.

Spotorno, S., Dragusin, I., Kirtley, C., & Tatler, B. W. (2019). Cognitive and perceptual influences on eye movements and object memory in real environments. *Journal of Vision, 19*(10), 147c.

Sullivan, B., Ludwig, C. J. H., Damen, D., Mayol-Cuevas, W., & Gilchrist, I. D. (2021). Look-ahead fixations during visuomotor behavior: Evidence from assembling a camping tent. *Journal of Vision, 21*(13), 1–17.

Tatler, B. W. (2014). Eye movements from laboratory to life. In M. Horsley, M. Eliot, B. A. Knight & R. Reilly (Eds.), *Current trends in eye tracking research.* (pp. 17–35). Cham: Springer International Publishing.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*(5), 5.

Tatler, B. W., & Land, M. F. (2016). Everyday visual attention. In J. Fawcett, E. Risko & A. Kingstone (Eds.), *The Handbook of Attention* (pp. 391–422). Cambridge: MIT Press.

Tatler, B. W., & Tatler, S. L. (2013). The influence of instructions on object memory in a real-world setting. *Journal of Vision, 13*(2), 1–13.

Thom, J. L., Nobre, A. C., van Ede, F., & Draschkow, D. (2023). Heading direction tracks internally directed selective attention in visual working memory. *Journal of Cognitive Neuroscience, 35*(5), 856–868.

Vö, M. L. H., & Wolfe, J. M. (2015). The role of memory for visual search in scenes. *Annals of the New York Academy of Sciences, 1339*(1), 72–81.

Wann, J. P., & Swapp, D. K. (2000). Why you should look where you are going. *Nature Neuroscience, 3*, 647–648.

Williams, C. C., Henderson, J. M., & Zacks, R. T. (2005). Incidental visual memory for targets and distractors in visual search. *Perception and Psychophysics, 67*(5), 816–827.

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

Zhang, Z., & Carlisle, N. B. (2023). Explicit attentional goals unlock implicit spatial statistical learning. *Journal of Experimental Psychology: General, 152*(8), 2125–2137.

Zinchenko, A., Conci, M., Müller, H. J., & Geyer, T. (2018). Predictive visual search: Role of environmental regularities in the learning of context cues. *Attention, Perception and Psychophysics, 80*, 1096–1109.