Search superiority: Goal-directed attentional allocation creates more reliable incidental identity and location memory than explicit encoding in naturalistic virtual environments

Jason Helbing[a,1], Dejan Draschkow[a,b,*,1], Melissa L.-H. Võ[a]

[a] Scene Grammar Lab, Department of Psychology, Goethe University Frankfurt, Frankfurt am Main, Germany
[b] Department of Psychiatry, University of Oxford, Oxford, England, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

ABSTRACT

We use representations and expectations formed during life-long learning to support attentional allocation and perception. In comparison to traditional laboratory investigations, real-world memory formation is usually achieved without explicit instruction and on-the-fly as a by-product of natural interactions with our environment. Understanding this process and the quality of naturally formed representations is critical to understanding how memory is used to guide attention and perception. Utilizing immersive, navigable, and realistic virtual environments, we investigated incidentally generated memory representations by comparing them to memories for items which were explicitly memorized. Participants either searched for objects embedded in realistic indoor environments or explicitly memorized them for follow-up identity and location memory tests. We show for the first time that memory for the identity of naturalistic objects and their location in 3D space is higher after incidental encoding compared to explicit memorization, even though the subsequent memory tests came as a surprise to participants. Relating gaze behavior to memory performance revealed that encoding time was more predictive of subsequent memory when participants explicitly memorized an item, compared to incidentally encoding it. Our results suggest that the active nature of guiding attentional allocation during proactive behavior allows for behaviorally optimal formation and utilization of representations. This highlights the importance of investigating cognition under ecologically valid conditions and shows that understanding the most natural processes for encoding and maintaining information is critical for understanding adaptive behavior.

1. Introduction

In natural interactions with our environment we effortlessly cope with an overwhelming amount of sensory information, react to it, and are able to proactively adjust our behavior in anticipation of future events. What seems almost trivial to us, actually constitutes an amazing feat of our cognitive apparatus. In order to dynamically adjust our actions to the current constraints of our surroundings we need to incorporate relevant knowledge and expectations regarding the immediate behavioral context to guide attention and overcome its limits in both space and time (Nobre & Stokes, 2019). This shifts the responsibilities for efficient behavioral guidance to recent episodic as well as more long-term memory, that is, knowledge structures. In fact, contextual information acquired during life-long learning facilitates object recognition (Bar, 2004; Biederman, Mezzanotte, & Rabinowitz, 1982; Davenport & Potter, 2004), guides eye movements (Torralba, Oliva, Castelhano, & Henderson, 2006; Võ & Henderson, 2010; Võ & Wolfe, 2015; Wolfe, Võ, Evans, & Greene, 2011) and supports the formation of new memories (Draschkow, Wolfe, & Võ, 2014; Josephs, Draschkow, Wolfe, & Võ, 2016). Specifically, more recently established episodic long term memory (LTM) also guides perception (Patai, Doallo, & Nobre, 2012; Stokes, Atherton, Patai, & Nobre, 2012; Summerfield, Rao, Garside, & Nobre, 2011) and the allocation of attention (Aly & Turk-Browne, 2017; Brockmole & Henderson, 2006; Chun & Jiang, 1998, 1999, 2003; Draschkow & Võ, 2016, 2017; Fan & Turk-Browne, 2016; Hutchinson & Turk-Browne, 2012; Patai, Buckley, & Nobre, 2013; Võ & Wolfe, 2012).

The role of LTM in guiding proactive behavior makes it imperative to investigate and understand the nature of its representations. In this study, we focus on the nature of episodic LTM created incidentally

---

* Corresponding author at: Oxford Centre for Human Brain Activity, Department of Psychiatry, University of Oxford, Warneford Hospital, OX3 7JX Oxford, United Kingdom of Great Britain and Northern Ireland.
E-mail address: dejan.draschkow@psych.ox.ac.uk (D. Draschkow).
[1] Authors contributed equally.

during visual search or intentionally during explicit memorization. Previous findings regarding the scale and fidelity of episodic LTM representations have been mixed. On the one hand, studies have shown rather limited storage of memory details (Brainerd & Reyna, 2005; Rensink, O'Regan, & Clark, 1997; Simons et al., 1997; Wolfe, 1998). On the other hand, there is evidence for episodic LTM content to be massive in capacity (Brady, Konkle, Alvarez, & Oliva, 2008; Standing, 1973; Standing, Conezio, & Haber, 1970) and highly detailed (Brady et al., 2008; Brady, Konkle, & Alvarez, 2011; Konkle, Brady, Alvarez, & Oliva, 2010). What most investigations into episodic object memory have in common is that (1) memorization is explicitly instructed and (2) tasks are used which are often quite remote from natural behavior. However, to understand the nature of memory representations, it is crucial to investigate their quality and detail in the realm of active natural behavior (Foulsham, Walker, & Kingstone, 2011; Malcolm, Groen, & Baker, 2016; Tatler, 2014; Tatler, Hayhoe, Land, & Ballard, 2011). Insights from ecologically valid behavioral paradigms is needed in order to inform cognitive functions which have actual behavioral relevance (Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017). For instance, it is rare for us to make a concerted effort to remember our visual surroundings explicitly (e.g., the location of the plates when visiting the new flat of a friend). More often, we engage in goal-directed behavior, such as setting the table or looking for the salt, during which location and identity representations of the surrounding objects are generated "on the fly". There is strong support for reliable memory of incidentally encoded items (Castelhano & Henderson, 2005; Draschkow et al., 2014; Draschkow, Reinecke, Cunningham, & Võ, 2018; Draschkow & Võ, 2017; Hollingworth, 2012; Hout & Goldinger, 2010, 2012; Kit et al., 2014; Võ & Wolfe, 2012, 2013b; Williams, 2010; Williams, Henderson, & Zacks, 2005; Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011), but the importance of probing LTM content which was formed incidentally really becomes apparent when it is used to guide proactive behavior in naturalistic environments. Võ and Wolfe (2012) demonstrated that attention is most profoundly influenced by item memory established during search ("looking for"), compared to explicit memorization and free viewing in complex scenes ("looking at"). Incidental fixations on objects during search speed subsequent search for these objects in real (Draschkow & Võ, 2016) and virtual (Draschkow & Võ, 2017) environments. Task constraints in general modulate information extraction and storage from fixations in real environments (Tatler et al., 2013). Investigating the contents of incidentally formed LTM representations during natural behavior in realistic environments thus promises to elucidate the behaviorally optimal nature of memory representations.

Investigations of memory content after incidental encoding in naturalistic settings are rare. Most direct and informative are investigations comparing incidentally formed memories with explicitly formed ones. Strikingly, incidental identity memory for searched items can be as strong as memory for items which are explicitly memorized in artificial search displays and even more reliable if the search arrays constitute real-world scenes (Draschkow et al., 2014; Josephs et al., 2016). Further, identity memory for distractors (Castelhano & Henderson, 2005; Williams, 2010) does not differ between incidental and explicit encoding conditions. What remains unexplored is the representational content of incidentally formed identity and location memories for objects in realistic and navigable environments. Within immersive 360-degree environments, our study compared LTM fidelity for incidentally and explicitly encoded items. Virtual reality allows for realistic and unconstrained task settings, while maintaining a high degree of experimental control (Draschkow & Võ, 2017; Figueroa, Arellano, & Calinisan, 2018; Kit et al., 2014; Li, Aivar, Kit, Tong, & Hayhoe, 2016; Li, Aivar, Tong, & Hayhoe, 2018; Olk, Dinu, Zielinski, & Kopper, 2018). It further enables us, for the first time, to probe explicit location memory for items in a maximally natural and sensitive way – by asking participants to rebuild the environments they were previously exposed to. This explicit measure of location memory is not only extending the paradigm to yet another form of memory testing, but provides an aspect of an object's memory representation of high behavioral relevance - as location memory enables us to find this object much faster in the future (e.g., Võ & Wolfe, 2012, 2013b) and thus allows for more efficient interactions with our surroundings. Tracking participants' eye movements while they were searching or memorizing, furthermore, allows us to account for differences in memory performance that were due to different exposure durations on objects and to further relate gaze behavior to subsequent memory performance (Hollingworth & Henderson, 2002). Due to the even more realistic nature of this type of search in a VR setting, we expect to observe the search superiority effect found by Draschkow et al. (2014) and Josephs et al. (2016) in both of our memory measures; that is, we expect that search targets are remembered more often than intentionally memorized objects in the identity memory test and that they will be positioned in closer proximity to their original location in the location memory test.

2. Method

2.1. Participants

Twenty-one participants (mean age: 24.3 years, age range: 18–38 years, 14 women and 7 men, 17 right-handed and 4 left-handed) were recruited at Goethe University Frankfurt. They volunteered, gave informed consent, and were compensated with course credit. All participants were tested for normal or corrected-to-normal (contact lenses, no glasses) visual acuity (at least 20/25 vision) and had normal color vision as assessed by the Ishihara test. They were all fluent German speakers. The research protocol was approved by the local ethics commission of the Faculty of Psychology and Sport Sciences at Goethe University Frankfurt.

Sample size was planned in accordance with effect sizes and samples from Josephs et al. (2016). Simulation-based estimation of sample size was conducted by running 1000 linear mixed-models on simulated data of that study using the Mixedpower library (Kumle, Võ, & Draschkow, 2018). Twenty-one participants yielded > 90% power for the detection of the search vs. memorization comparison using the reported non-standardized effect sizes of Joseph et al.'s Experiment 4 (condition $\beta = 0.25$, gaze duration $\beta = 0.76$) and > 75% power when using a conservative correction (0.68 confidence interval) of these effects (Perugini, Gallucci, & Costantini, 2014).

2.2. Apparatus

Participants were equipped with an SMI Tethered Eye Tracking VR Head Mounted Display (HMD) based on HTC Vive, an HTC Vive Controller in their writing hand, and a set of Philips SHP8500 headphones. The HMD has a refresh rate of 90 Hz and a field of view of approximately $100°$ (horizontally) $\times 110°$ (vertically). It uses two 1080×1200 px resolution OLED screens and multiple sensors, including 37 infrared sensors, an accelerometer, and a gyroscope. The controller is wireless, has 24 infrared sensors, and allows for multiple input methods, three of which were used in the experiment: the trigger at the back of the controller (pulled with the index finger) and the menu button and trackpad at the front side (both pressed with the thumb). Location tracking to a fraction of a millimeter is achieved for both HMD and controller with the help of two base stations (Lighthouse tracking system) emitting 60 infrared pulses per second, which are detected by the sensors of the devices. The binocular eye tracker that is integrated in the HMD streams participants' eye movements at a sampling rate of 90 Hz (refresh rate of the HMD) with a spatial accuracy of approximately $0.4°$. The experiment was programmed and run in Unity (version 2017.3; Unity Technologies) using SteamVR (version 1.2.10; Valve Corporation) on a computer operated with Windows 10.

2.3. Stimuli

Ten virtual indoor scenes, that is, two of each of five room categories (living room, bedroom, bathroom, kitchen, and office), were used in the experiment. They all had the same size of approximately 380 × 350 × 260 cm (length × width × height). Wall coverings, flooring materials, and ceilings were chosen in accordance with the room category (e.g., tiles in the kitchen or inlay carpet in some bedrooms). There were 20 category-appropriate objects in each room: 5 global objects, that is, large static objects (e.g., couch, toilet, desk) and 15 local objects, which are smaller, usually movable, and often used when we interact with our environment (e.g., coffee mug, toothbrush, remote control). The differential roles of these two types of objects in visual search were recently examined by Boettcher, Draschkow, Dienhart, and Võ (2018). They found that many global objects can be considered anchors that provide spatial information of corresponding local objects. Hence, the local objects were the targets of visual search in this experiment. Global objects were pairs of different exemplars of the same set of objects in each room category (e.g., there was a black leather couch in one living room and a light fabric one in the other). All local objects were singletons, that is, there was no other exemplar of any object in the same or one of the other rooms. In addition to the 3D objects in the rooms, another 150 objects were used as lures in the recognition memory test. These were chosen in five groups of 30 objects each to match the five room categories (e.g., pot as a kitchen lure). Some of them were different exemplars of objects that were also used in the experimental environments, but those object pairs differed in appearance and were thus hardly confusable (e.g., baguette in the scene and a loaf of brown bread as a lure). Images of rooms and generic global and local objects are depicted in Fig. 1. Furthermore, there was a black display-like area (cue screen) on one of the walls that was used to present target words during the search and memorization tasks in a white font. One additional room was used for practice trials. It exclusively contained objects that would appear rather unusual in typical indoor environments (e.g., tree, rock, *T. rex*) in order to avoid any interference with rooms of the actual experiment. All 3D objects were free models taken from online repositories, namely, Archive 3D, CGTrader, Free3D, TurboSquid, and the Unity Asset Store.

2.4. Procedure

The experimental procedure consisted of two subsequent conditions: a search and a memorization task condition. Both conditions consisted of three phases: a phase in which the respective task was performed in the scene environments, an old-new recognition test to assess identity memory, and a location memory test. In order to counterbalance potential order effects, odd-numbered participants started with the search condition, even-numbered ones with the memorization condition. The assignment of rooms to the two conditions was counterbalanced across participants; there was always one room of each category in the search and the memorization condition. Subjects were led to believe that the two conditions were unrelated experiments. This was to avoid that participants who started with the memorization condition would anticipate the surprise memory tests after the search phase. In both conditions, the identity memory test always preceded the location memory test as switching their order would have exposed the participants to the rooms with the correct objects in the arrangement task before recognition testing of those same objects. Participants carried out all tasks of the experiment in the virtual environment. Trial procedures of all phases are displayed in Fig. 2 and a video with example trials of all of these is available at https://youtu.be/ZI2SkWYrCcI. There was a 10 min break between the two conditions. Before starting with the first task, participants were familiarized with the HMD, how to move around the lab space while wearing it, controller input methods, and the calibration procedure of the eye tracker. During the actual experiment, the eye tracker was re-calibrated before both the search and the memorization phase.

The *Search Phase* started with the presentation of a verbal target cue (in German) that appeared on the cue screen together with the respective room of that trial. Then, participants would search for the target. They were instructed to fixate on the target once they had found it and then pull the trigger. For each trial, there was a 10 s timeout limit. An error sound (1 s) would play if the target object was not looked at when the trigger was pulled or if there was a timeout. Upon pulling the trigger, the scene disappeared, leaving the participant in a completely dark room where they would turn back to the cue screen, which now displayed a fixation cue (the word *attention* in German, all caps). Once the eye tracker had detected they were looking at that cue, the next target word and the corresponding room appeared after 2 s. Participants were instructed to search as fast and accurately as possible. They were also informed about the fact that the target was always present in the scene and that they had 10 s to find it. They were not informed that there would later be memory tests on the scenes' objects. There were ten trials in each room, resulting in a total of 50 searches. The targets were randomly selected out of the 15 local objects for each room of each participant. Hence, there were five local objects per room for every participant that were not searched for. We will refer to these as distractors in the following. The order of trials was randomized irrespective of rooms (i.e., the trial order was not room by room). There were ten practice search trials in the aforementioned practice room before the actual trials.

The procedure of *Memorization Phase* was created to be as similar as
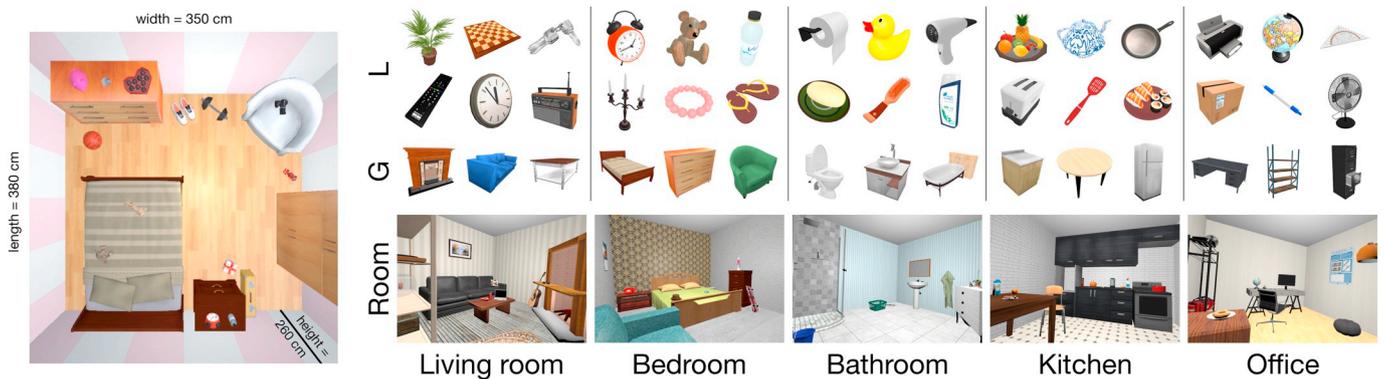
Fig. 1. A bird's-eye view of one of the bedrooms with dimensions (left) and one room of each category with global and local objects of that category (right). G = Global; L = Local.
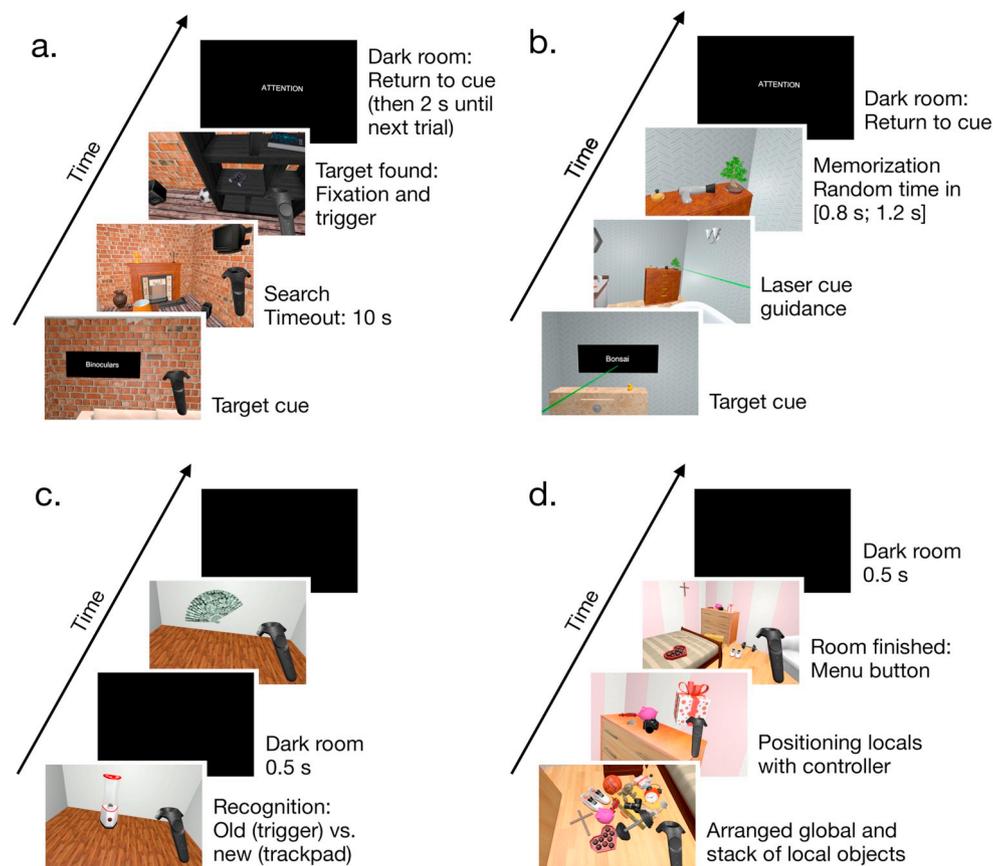
Fig. 2. Trial procedures of the visual search task (a), the memorization task in which a laser cue guided participants to the target object (b), the old-new recognition memory test that assessed identity memory of the objects in the scenes in which participants had searched for or memorized objects before (c), and the location memory test in which the local objects of those scenes had to be arranged from memory with the global objects given in their correct locations (d). Note that the depicted cues are in English for visualization purposes (German in the original experiment).

possible to the search trials, the major difference being that the necessity to search was eliminated. The verbal cue was presented on the cue screen like on a search trial. Then, a green laser-like line originating below the target word guided participants to the object they were supposed to memorize on this trial. Once the eye tracker had detected gaze on this object, the laser cue disappeared. The room then remained visible for a memorization time interval whose length was randomly drawn from a uniform distribution on [0.8 s; 1.2 s]. This interval was chosen based on previous pilot testing data to match the gaze durations on memorization targets with those of search targets which mostly succeeded (summed gaze on search targets: $M = 1.23$ s, $Q_1 = 0.69$ s, $Q_2 = 1.04$ s, $Q_3 = 1.60$ s; memorization targets: $M = 0.98$ s, $Q_1 = 0.66$ s, $Q_2 = 1.00$ s, $Q_3 = 1.25$ s). After the memorization time, the room went black and the next trial was initiated in the same way as in the search trial procedure. Participants were instructed to read the target cue, follow the laser, and memorize the appearance and location of the cued object as well as possible for later memory testing. Again, there were ten practice memorization trials (same objects as in the search condition) to familiarize participants with the laser cueing paradigm, though they were informed that they would not be tested on the practice objects. Just like in the search condition, there were trials on ten randomly selected local objects in each of five rooms (resulting in five distractors remaining in each room), yielding 50 trials the order of which was randomized.

In the *Identity Memory Test*, participants completed an old-new recognition task. On each trial, one object was presented in the center of

the room, statically hovering approximately 110 cm above the ground. Participants were instructed to indicate whether they had seen this object in one of the rooms in which they had previously performed either the search or the memorization task or whether it appeared novel to them. They could also move around the object to inspect it from different angles in their decision making. Once they had responded by either pulling the trigger (old object) or pressing trackpad (new object), the room turned completely dark and after 0.5 s became visible again with the next object in place. The wallpaper, floor, and ceiling textures of the room in which this test was performed were not used in any of the experimental rooms. All 15 local objects from each of the five scenes (ten targets and five distractors) of the respective condition as well as 15 lures of each room category (see Stimuli) were presented, yielding 150 trials (50 targets, 25 distractors, 75 lures). Thus, chance level was at 50%. The order of trials was randomized.

The *Location Memory Test* had participants revisit the five rooms of the respective condition in a random order. Local objects, however, were not in their original spots, but on a pile on the floor. Participants were instructed to move those objects back to their previous locations based on what they remembered from the search or memorization trials. To move an object, they would stick their controller into it and then pull the trigger to grab the object, move it with the controller (trigger pulled), and eventually let it go by releasing the trigger. With respect to physics, the objects behaved much like one would expect from real objects, that is, they would fall down, roll, or bounce off of other surfaces; however, their mesh remained static, meaning that, for

example, they could not bend or break. Objects whose original location was not on top of another object but in a hanging state (e.g., wall clock, painting) were not affected by gravity and would stick to vertical surfaces such as walls in a magnet-like fashion. Global objects were already in their original locations and could not be moved. This was because (a) they had not been targets before, (b) it is rather unusual to move them in everyday life, and (c) a recent VR scene building experiment by Draschkow and Võ (2017) shows that they are usually placed earlier in the building process and then serve as anchors to guide the arrangement of smaller objects that we typically interact with, which makes them appropriate location recall cues in our study. Once participants had arranged all objects in a room, they would use the menu button to move on to the next one (after 1 s of complete darkness).

2.5. Data analysis

The analysis of accuracy in the Identity Memory Test was performed only on trials that probed objects from the actual scenes (targets and distractors), meaning that lures were excluded. The mean correct rejection rate of lures was at 91.7% in the search condition and at 90.0% in the memorization condition. For the Location Memory Test, distances between the actual locations of the objects in the scenes and the locations where the participants placed those objects were computed. Distance was calculated using the Euclidian metric of the centers of the 3D objects' bounding boxes. Shorter distances were interpreted as better location memory performance. Raw eye movement data was sampled at 90 Hz during the search and memorization trials. To get a measure of how long local objects were looked at during those tasks, all samples that were part of groups of three or more consecutive samples with gaze on the same object were summed as a function of object and participant.

Generalized linear mixed-effects models (GLMMs) with a binomial distribution were used to analyze accuracy data from the Identity Memory Test. Linear mixed-effects models (LMMs) were used to analyze the distances obtained from the location memory test. These analyses were run using the lme4 package (version 1.1-17; Bates, Mächler, Bolker, & Walker, 2015) in the R statistical programming language (version 3.5.3; R Core Team, 2019) using RStudio (version 1.1.463; RStudio Team, 2016). We chose to use mixed-effects models as they allow for the simultaneous estimation of between-subject and between-stimulus variance and thus yield advantages over traditional F1/F2 analyses of variance (Baayen, Davidson, & Bates, 2008; Kliegl, Wei, Dambacher, Yan, & Zhou, 2011).

All GLMMs and LMMs were fitted with the maximum likelihood criterion. After inspecting the location memory distances' distribution, residuals, and power coefficient output of the Box–Cox procedure (Box & Cox, 1964), that was run with the MASS package (version 7.3-51.1; Venables & Ripley, 2002), its values were log-transformed in order to approximate a normal distribution more closely and meet LMM assumptions. Sum contrasts, where the grand mean of the dependent measure serves as the intercept and slope coefficients represent the difference between factor levels, were defined for the encoding task condition (search vs. memorization) and the objects' role on this task (target vs. distractor; object type). Gaze duration on objects was also included as a fixed effect (covariate) using scaled and centered values. We started with a maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013) that included random intercepts and slopes for participants and scenes. Since full models like these often fail to converge or lead to overparameterization (Bates, Kliegl, Vasishth, & Baayen, 2015), we used a principal component analysis (PCA) of the random effects variance-covariance estimates to identify overparameterization for each fitted model and removed random slopes that were not supported by the PCA and did not contribute significantly to the goodness of fit in a likelihood ratio (LR) test (Bates, Kliegl, et al.,

2015). The best-fitting accuracy GLMM's random effects structure had subject and scene intercepts as well as by-subject slopes for encoding task and object type. The random effects structure of the optimal LMM for predicting location memory distances had the same components plus a by-scene slope for object type. LR tests were also used to check if modelling gaze duration on objects as a multiplicative fixed covariate was supported by the data. Comparing the models with alternatives where the gaze covariate was additive showed that the multiplicative versions yielded a significantly better goodness of fit for both memory measures. Significant interactions of task and object type were broken down by defining difference contrasts with which the two critical comparisons (search vs. memorization for targets and for distractors) could be modelled. Details about all models and model comparisons can be found in the analysis script (see Data availability).

For the LMMs, we report regression coefficients β with the t statistic and apply a two-tailed criterion corresponding to a 5% error criterion for significance. In this case, the p-values for the continuous distance variable were calculated with Satterthwaite's degrees of freedom method using the lmerTest package (version 3.1-0; Kuznetsova, Brockhoff, & Christensen, 2017). For the GLMMs, where we report β with the z statistic and, again, use a two-tailed 5% error criterion for significance, the p-values for the binary accuracy variable are based on asymptotic Wald tests.

Figures in the Results section are partial effect plots whose adjusted dependent variable values were obtained with the remef package (version 1.0.6.9000; Hohenstein & Kliegl, 2019). In plots like these, the partial effects are computed from model parameters, which means that the estimated statistical effects are depicted; this allows for an unclouded, straightforward interpretation of the results. Plots of the untransformed data can be found in the Appendix. The ggplot2 package (version 3.1.0; Wickham, 2016) was used for data plotting. Standard errors for plots were computed using the Rmisc package (version 1.5; Hope, 2013).

2.6. Data availability

Experimental data and the script that was used for data analysis are available at the Open Science Framework at https://osf.io/qvyk5/.

3. Results

3.1. Identity memory

There were significant main effects of object type, β = 0.89, SE = 0.09, z = 9.83, p < .001, and gaze duration, β = 0.57, SE = 0.08, z = 6.77, p < .001, on identity recognition accuracy: Objects that were targets and objects that were looked at longer were remembered better. There was no significant main effect of encoding task, β = 0.11, SE = 0.09, z = 1.12, p = .26, but a significant interaction of task and object type, β = 0.32, SE = 0.08, z = 4.05, p < .001 (see Fig. 3a). Breaking this interaction down into planned comparisons revealed that this was because targets of visual search were remembered more accurately than targets that were intentionally memorized, β = 0.84, SE = 0.19, z = 4.48, p < .001. There was no significant effect of task for distractors, β = −0.42, SE = 0.29, z = −1.43, p = .15. Two other interactions, which involved gaze duration on objects, were significant: First, looking at objects longer led to stronger performance gains in the memorization condition as compared to search, β = −0.17, SE = 0.08, z = −2.06, p = .04 (see Fig. 3b). Second, accuracy of distractors improved more when they were looked at longer as compared to targets, β = −0.38, SE = 0.08, z = −4.51, p < .001 (see Fig. 3c).
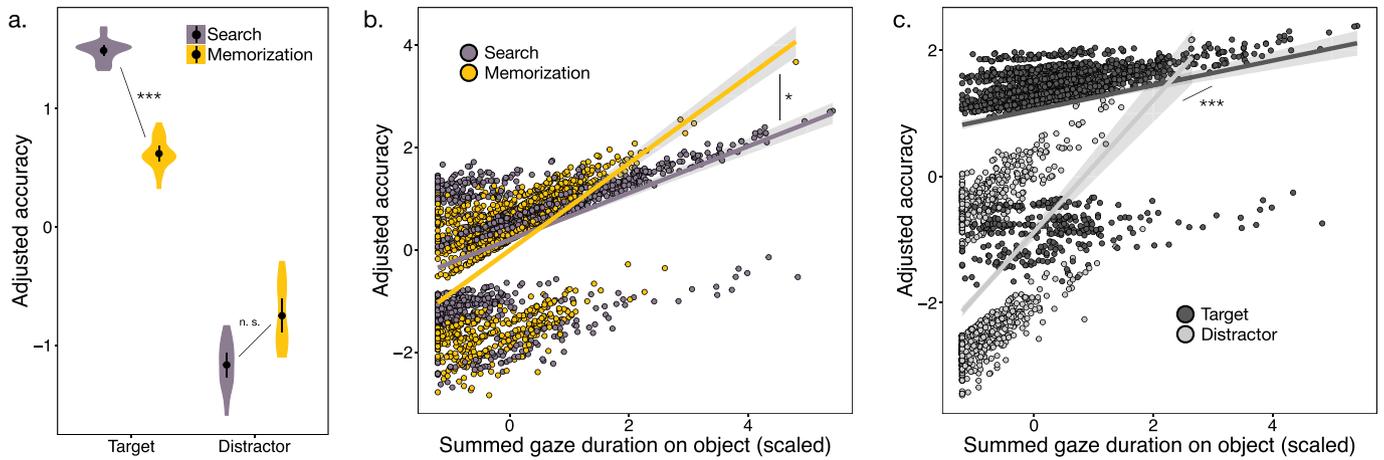
Fig. 3. Partial effect plots of the interactions of task (search vs. memorization) and object type (target vs. distractor; a), task and summed gaze duration on object (b), and object type and summed gaze duration on object (c) on identity memory accuracy. Note that higher values indicate better performance. Fig. A1a with untransformed values of (a) can be found in the Appendix. Error bars and shaded areas represent 95% confidence intervals. n. s. = not significant. *$p < .05$. ***$p < .001$.

3.2. Location memory

The results of the analyses of distances between the original locations of objects and where participants placed them in the location memory test yielded the very same pattern that was observed for identity memory accuracy: Significant main effects of object type, $\beta = -0.24$, $SE = 0.05$, $t = -4.68$, $p < .001$, and gaze duration, $\beta = -0.24$, $SE = 0.05$, $t = -4.68$, $p < .001$; no significant main effect of task, $\beta = -0.003$, $SE = 0.05$, $t = -0.06$, $p = .95$; interaction of task and object type, $\beta = -0.17$, $SE = 0.04$, $t = -4.05$, $p < .001$ (see Fig. 4a) due to search targets being placed closer to their original locations than memorization targets, $\beta = -0.34$, $SE = 0.09$, $t = -3.71$, $p < .001$, whereas, again, no such task effect was significant for distractors, $\beta = 0.32$, $SE = 0.17$, $t = 1.96$, $p = .051$; significant interaction between task and gaze duration, $\beta = 0.13$, $SE = 0.04$, $t = 3.05$, $p = .002$: Having looked at an object longer aided the correct placement more in the memorization than in the search condition (see Fig. 4b); and lastly, an interaction between object type and gaze duration, $\beta = 0.23$, $SE = 0.04$, $t = 5.41$, $p < .001$: With respect to correct arrangement, looking at an object longer was more beneficial for distractors than for targets (see Fig. 4c).

4. Discussion

Using a virtual reality eye tracking paradigm, we compared object memory representations formed during visual search with ones formed during intentional memorization in naturalistic indoor scenes. We found memory performance for both the identity and the location of search targets to be higher than that of explicitly memorized objects. Furthermore, dwell time on objects while performing the search or memorization task predicted identity and location memory, with longer dwell time leading to better performance. This relationship differed between the two tasks, as memory profited more profoundly from longer dwell times during explicit memorization than during search.

Superior identity memory performance for searched targets compared to memorized ones has been demonstrated before using naturalistic photographs presented on a screen (Draschkow et al., 2014; Josephs et al., 2016). We for the first time replicate these findings from highly constrained 2D paradigms showing that these results generalize to navigable and immersive environments. Further, the search superiority effect has thus far only been shown with free recall procedures, which offer a detailed and quantifiable method for accessing object memory (Bainbridge, Hall, & Baker, 2019). Here we replicate and extend these findings to recognition and location memory, suggesting that
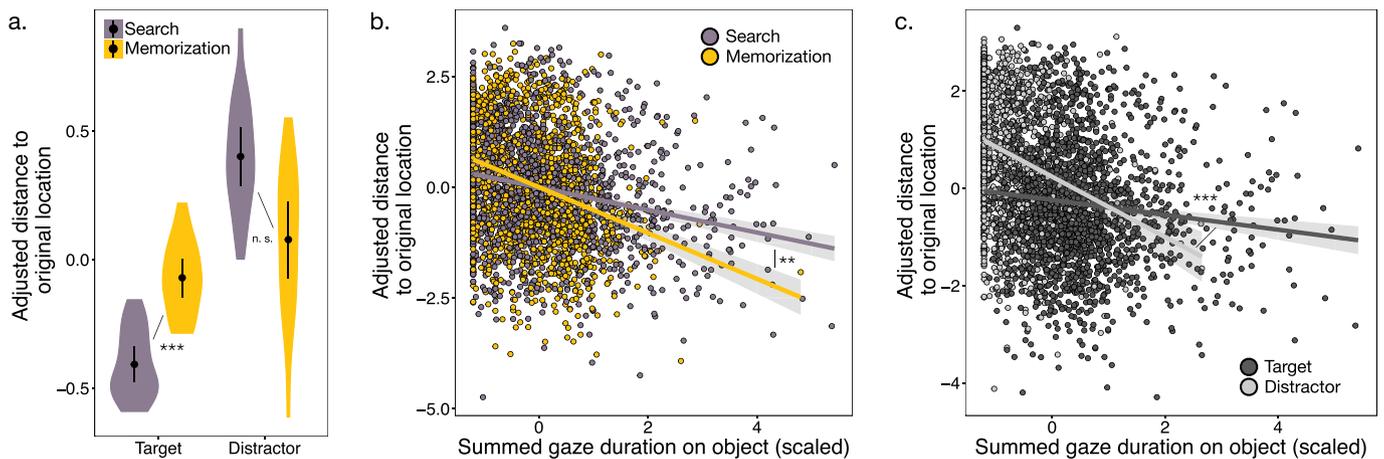
Fig. 4. Partial effect plots of the interactions of task (search vs. memorization) and object type (target vs. distractor; a), task and summed gaze duration on object (b), and object type and summed gaze duration on object (c) on distances between objects' original location and where participants placed them in the location memory test. Note that lower values indicate better performance. Fig. A1b with untransformed values of (a) can be found in the Appendix. Error bars and shaded areas represent 95% confidence intervals. n. s. = not significant. **$p < .01$. ***$p < .001$.

this effect is largely independent of the utilized memory test. Critically, by the nature of screen-based paradigms, previous studies were limited in their sensitivity of probing location memory of objects. Utilizing the full 3D space of an immersive environment, we can now show that explicit location memory was stronger after search compared to after memorization. Such reliable incidental encoding of item locations is supported by studies which investigate how previous exposure guides attentional allocation in tasks which involve locomotion (Draschkow & Võ, 2016, 2017; Li et al., 2016, 2018).

But what makes searching for an object such a potent process for creating a strong representation of the target? The simplest and most theoretically unimportant answer, albeit wrong, is that searched targets are looked at longer and thus receive more time for encoding. This, however, is not the case, as the search superiority effect persists if objects are viewed for a comparable length of time during memorization compared to search (Draschkow et al., 2014), or if differences in viewing time are accounted for by the statistical method, additionally to being largely equated in the procedure (Josephs et al., 2016, as well as this study). Understanding the intricate dynamics between environmental structure and encoding of visual information promises greater explanatory power for deciphering the search superiority effect.

With respect to the stage at which search superiority is established, our results provide evidence that the emergence of more reliable representations during search is established during the encoding and maintenance stage of visual information sampling. Considering that we generalize the same effect to two completely novel memory test procedures, it seems unlikely that search superiority is grounded in retrieval-based differences. Not only is this superiority apparent in now various explicit measures, but also when memories are used to guide attention (Võ & Wolfe, 2012). Further, by relating gaze behavior to memory performance, we show that visual information at encoding is used differentially in the two tasks. This supports the notion that the task-based differences in memory performance are supported by strategic differences in the extraction and retention of information from fixations (Tatler & Tatler, 2013). The fact that looking at objects longer led to stronger representations in the memorization compared to the search task is also further evidence that the strength of these object representations is established during encoding since Damiano and Walther (2019) found that longer dwell times during encoding rather than retrieval aid object recall performance.

Another critical determinant for effective incidental encoding and maintenance of information during search is *semantic* and structural (*syntax*) information in the environment (i.e., the rules that typically govern naturalistic settings, which we have come to call *scene grammar*; Biederman et al., 1982; Võ, Boettcher, & Draschkow, 2019; Võ & Wolfe, 2013a). When insufficient contextual information is provided (Draschkow et al., 2014) or participants are not given enough time to associate it with a given target (Josephs et al., 2016), identity memory performance between searched and memorized items does not differ, that is, the search superiority effect diminishes. Note, however, that representations after search are still comparable to explicitly generated ones. Thus, even with stimuli which are lacking semantic richness, extraction and retention of information from fixations is still extremely effective during search. Controlled manipulations of different scene components will be necessary to identify which of them are required for strong target memories or how the violation of specific aspects of scene grammar might differentially impact the formation of these memories. Further, to achieve a more mechanistic account of the effect, voluntary and non-voluntary dynamic attentional shifts can be compared, thus equating the scan paths, but manipulating the ecological relevance of search behavior. Virtual reality will play a critical role for these avenues of research.

With respect to the neural basis of strong encoding of targets of visual search, the way in which scene grammar guides attention could also play a key role: Aly and Turk-Browne (2016) demonstrate that goal-driven attention can promote the encoding of task-relevant visuospatial information in the hippocampus (for an overview see Aly & Turk-Browne, 2017). When

searching for a target in a naturalistic scene for the first time, the allocation of this goal-driven attention could be guided by knowledge about the structure of the scene (Hutchinson & Turk-Browne, 2012; Torralba et al., 2006; Wolfe, Võ, et al., 2011). This could explain why Draschkow et al. (2014) and Josephs et al. (2016) found that the strength of target memories diminishes so strongly once they are searched in arrays of isolated objects: The scene context might be necessary to activate top-down guidance of attention that helps finding the target which could eventually lead to the abovementioned reliable hippocampal encoding, resulting in the strong target representations observed in this study and earlier research. These representations could then be retrieved on a memory test or serve as strong episodic memory-guidance for attention in subsequent search trials for the same object (Chun & Turk-Browne, 2007; Summerfield, Lepsien, Gitelman, Mesulam, & Nobre, 2006; Wolfe, Alvarez, et al., 2011). Intentional memorization, which is typically linked to an interplay of hippocampal and prefrontal activity and not observed in this form during incidental encoding (Rugg, Fletcher, Frith, Frackowiak, & Dolan, 1997), might fail to elicit this specific goal-driven modulation of the hippocampus, since there is no need for semantic memory-guidance of attention without the need to search. Although this reasoning is merely speculative at this time, our results strongly promote future research looking into the neural mechanisms of incidental object encoding and its relation to scene context.

Knowing of the potential of visual search as a task that forms strong memory representations, another question that naturally arises is whether this effect can be applied to facilitate learning processes. A recent approach successfully used search as a learning tool for object-pseudoword-connections in VR environments (Voelker, Draschkow, & Võ, 2018), thus teaching vocabulary without explicit memorization. Our findings, together with the previous studies by Draschkow et al. (2014) and Josephs et al. (2016), support the idea that search forms strong memory representations across different kinds of environments and recall tasks and therefore encourages investigations into the effectiveness of such search-based mnemonic strategies. Josephs et al. (2016) point to a possible parallel between the search superiority effect and research from memory and educational research on the so-called testing effect. This refers to the finding that after a study period, a quiz on the studied material will produce better and more persistent learning than restudying the material for an equal amount of time (e.g., Abbott, 1909; Roediger & Karpicke, 2006a; for a review see Roediger & Karpicke, 2006b). Critically, this effect is modulated by the availability of prior knowledge with which to engage during the quiz and link newly studied material to (King, 1994). This phenomenon resembles the formation of memory for visual search targets in two ways: (1) A goal-oriented task (search/quiz) enhances memory strength more than additional study/memorization. (2) This is affected by whether the new information is well embedded in preexisting semantic knowledge structures.

From a methodological standpoint, our study successfully used a VR eye tracking paradigm to replicate and extend findings previously only observed in 2D screen-based paradigms. Thus, it also adds to the growing corpus of studies investigating visual search in more immersive virtual scenes (e.g., Draschkow & Võ, 2017; Figueroa et al., 2018; Kit et al., 2014; Li et al., 2016; Li et al., 2018; Olk et al., 2018). More specifically, the use of VR has enabled us to overcome many limitations of screen-based paradigms: We were able to investigate task behavior, probe memory, and track eye movements while participants moved freely in a multitude of realistic full-sized everyday scenes. At the same time, we had full control over the trial procedure and could instantaneously switch between different scenes, thus preserving the same level of experimental control given in a screen-based paradigm. Moreover, the interactive nature of the location memory test (grabbing and arranging objects while freely walking through the scene) enabled us to assess explicit memory for the precise location of objects with participants fully immersed into the scene that served as the contextual recall cue. We hope that our results encourage the attempts for replication of more classical findings in ecologically valid environments. Only then can we assert whether some of the "classic" effects demonstrated in more artificial settings can withstand scrutiny under

more real-life conditions and gain valuable insights from data that is simply not accessible with a static screen-based setup.

Our study demonstrates that searching for an object in a navigable and immersive environment creates a stronger memory representation of that target than intentionally memorizing it. We showed that this is (a) true for the identity and location of the object, and that (b) the link between looking at an object longer and having better memory of its identity and location is weaker for search targets than for intentionally memorized objects. This indicates that incidental extraction and retention of information from fixations is extremely effective during search and leads to behaviorally optimal representations which can be used to guide future behavior proactively. As searching is a ubiquitous sub-task of most natural interactions, it is critical to elucidate its contributions to long-term memory representations. Our results highlight the importance of investigating cognition under ecologically valid conditions and show that understanding the most natural processes for encoding and maintaining information is essential for understanding adaptive behavior.

Appendix
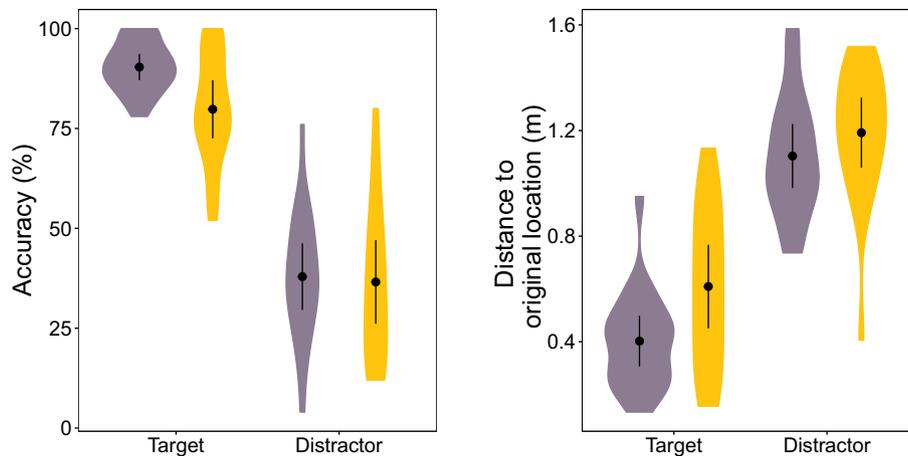
Fig. A1. Untransformed memory test data depicting the interaction of task (search vs. memorization) and object type (target vs. distractor) for identity (a) and location memory (b).

Author contributions

J.H., D.D., and M.L.-H.V. conceptualized the design of the study. J.H. programmed and conducted the experiment. D.D. and J.H. analyzed the data. J.H. and D.D. wrote the manuscript with edits from M.L.-H.V. All authors approved the final version of the manuscript.

Declaration competing interest

The authors declare no competing interests.

Acknowledgements

References

Abbott, E. A. (1909). On the analysis of the factor of recall in the learning process. *Psychological Monographs: General and Applied, 11*(1), 159–177. https://doi.org/10.1037/h0093018.

Aly, M., & Turk-Browne, N. B. (2016). Attention promotes episodic encoding by stabilizing hippocampal representations. *Proceedings of the National Academy of Sciences, 113*(4), E420–E429. https://doi.org/10.1073/pnas.1518931113.

Aly, M., & Turk-Browne, N. B. (2017). How hippocampal memory shapes, and is shaped by, attention. *The hippocampus from cells to systems: Structure, connectivity, and functional contributions to memory and flexible cognition* (pp. 369–403). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-50406-3_12.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Bainbridge, W. A., Hall, E. H., & Baker, C. I. (2019). Drawings of real-world scenes during free recall reveal detailed object and spatial information in memory. *Nature Communications, 10*(1), 1–13. https://doi.org/10.1038/s41467-018-07830-6.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617–629. https://doi.org/10.1038/nrn1476.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. Retrieved from http://arxiv.org/abs/1506.04967.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme 4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*(2), 143–177. https://doi.org/10.1016/0010-0285(82)90007-X.

Boettcher, S. E. P., Draschkow, D., Dienhart, E., & Võ, M. L.-H. (2018). Anchoring visual search in scenes: Assessing the role of anchor objects on eye-movements during visual search. *Journal of Vision, 18*(13), 1–13. https://doi.org/10.1167/18.13.11.

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

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11*(5), 4. https://doi.org/10.1167/11.5.4.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105.

Brainerd, C. J., & Reyna, V. F. (2005). In C. J. Brainerd, & V. F. Reyna (Eds.). *The science of false memory*Oxford University Presshttps://doi.org/10.1093/acprof:oso/9780195154054.001.0001.

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition, 13*(1), 99–108. https://doi.org/10.1080/13506280500165188.

Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition, 12*(6), 1017–1040. https://doi.org/10.1080/13506280444000634.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681.

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science, 10*(4), 360–365. https://doi.org/10.1111/1467-9280.00168.

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 29*(2), 224–234.

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory.

Current Opinion in Neurobiology, 17(2), 177–184. https://doi.org/10.1016/j.conb.2007.03.005.

Core Team, R. (2019). R: A language and environment for statistical computing. Vienna, Austria. Retrieved from https://www.r-project.org/.

Damiano, C., & Walther, D. B. (2019). Distinct roles of eye movements during memory encoding and retrieval. *Cognition, 184*(December 2018), 119–129. https://doi.org/10.1016/j.cognition.2018.12.014.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*(8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x.

Draschkow, D., Reinecke, S., Cunningham, C. A., & Võ, M. L.-H. (2018). The lower bounds of massive memory: Investigating memory for object details after incidental encoding. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/1747021818783722.

Draschkow, D., & Võ, M. L.-H. (2016). Of "what" and "where" in a natural search task: Active object handling supports object location memory beyond the object's identity. *Attention, Perception, & Psychophysics, 78*(6), 1574–1584. https://doi.org/10.3758/s13414-016-1111-x.

Draschkow, D., & Võ, M. L.-H. (2017). Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Scientific Reports, 7*, 1–12. https://doi.org/10.1038/s41598-017-16739-x.

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), 1–18. https://doi.org/10.1167/14.8.10.

Fan, J. E., & Turk-Browne, N. B. (2016). Incidental biasing of attention from visual long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(6), 970–977. https://doi.org/10.1037/xlm0000209.

Figueroa, J. C. M., Arellano, R. A. B., & Calinisan, J. M. E. (2018). A comparative study of virtual reality and 2D display methods in visual search in real scenes. In D. N. Cassenti (Ed.). *Advances in human factors in simulation and modeling* (pp. 366–377). Cham: Springer International Publishing.

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. https://doi.org/10.1016/j.visres.2011.07.002.

Hohenstein, S., & Kliegl, R. (2019). remef: Remove partial effects.

Hollingworth, A. (2012). Task specificity and the influence of memory on visual search: comment on Võ and Wolfe (2012). *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1596–1603. https://doi.org/10.1037/a0030237.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*(1), 113–136. https://doi.org/10.1037//0096-1523.28.1.113.

Hope, R. M. (2013). Rmisc: Ryan miscellaneous. Retrieved from https://cran.r-project.org/web/packages/Rmisc/Rmisc.pdf

Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception & Psychophysics, 72*(5), 1267–1282. https://doi.org/10.3758/APP.72.5.1267.

Hout, M. C., & Goldinger, S. D. (2012). Incidental learning speeds visual search by lowering response thresholds, not by improving efficiency: Evidence from eye movements. *Journal of Experimental Psychology. Human Perception and Performance, 38*(1), 90–112. https://doi.org/10.1037/a0023894.

Hutchinson, J. B., & Turk-Browne, N. B. (2012). Memory-guided attention: Control from multiple memory systems. *Trends in Cognitive Sciences, 16*(12), 576–579. https://doi.org/10.1016/j.tics.2012.10.003.

Josephs, E. L., Draschkow, D., Wolfe, J. M., & Võ, M. L.-H. (2016). Gist in time: Scene semantics and structure enhance recall of searched objects. *Acta Psychologica, 169*, 100–108. https://doi.org/10.1016/j.actpsy.2016.05.013.

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal, 31*(2), 338–368. https://doi.org/10.3102/00028312031002338.

Kit, D., Katz, L., Sullivan, B., Snyder, K., Ballard, D., & Hayhoe, M. (2014). Eye movements, visual search and scene memory, in an immersive virtual environment. *PLoS ONE, 9*(4), 1–11. https://doi.org/10.1371/journal.pone.0094362.

Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology, 1*, 1–12. https://doi.org/10.3389/fpsyg.2010.00238.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology. General, 139*(3), 558–578. https://doi.org/10.1037/a0019165.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron, 93*(3), 480–490. https://doi.org/10.1016/J.NEURON.2016.12.041.

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2018). *Mixedpower: A library for estimating simulation-based power for mixed models in R.* https://doi.org/10.5281/zenodo.1341047.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

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*(2016), 1–20. https://doi.org/10.1167/16.8.9.

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*(1), 1–11. https://doi.org/10.1038/s41598-018-22731-w.

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

Nobre, A. C., & Stokes, M. G. (2019). Premembering experience: A hierarchy of time-scales for proactive attention. *Neuron, 104*(1), 132–146. https://doi.org/10.1016/j.neuron.2019.08.030.

Olk, B., Dinu, A., Zielinski, D. J., & Kopper, R. (2018). Measuring visual search and distraction in immersive virtual reality. *Royal Society Open Science, 5*(5), 1–15. https://doi.org/10.1098/rsos.172331.

Patai, E. Z., Buckley, A., & Nobre, A. C. (2013). Is attention based on spatial contextual memory preferentially guided by low spatial frequency signals? *PLoS One, 8*(6), e65601. https://doi.org/10.1371/journal.pone.0065601.

Patai, E. Z., Doallo, S., & Nobre, A. C. (2012). Long-term memories bias sensitivity and target selection in complex scenes. *Journal of Cognitive Neuroscience, 24*(12), 2281–2291. https://doi.org/10.1162/jocn_a_00294.

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science, 9*(3), 319–332. https://doi.org/10.1177/1745691614528519.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368–373. https://doi.org/10.1111/j.1467-9280.1997.tb00427.x.

Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255.

Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210.

RStudio Team (2016). RStudio: Integrated development environment for R. Boston, MA. Retrieved from http://www.rstudio.com/.

Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1997). Brain regions supporting intentional and incidental memory: A PET study. *NeuroReport, 8*(5), 1283–1287. https://doi.org/10.1097/00001756-199703240-00045.

Simons, D. J., Levin, D. T., Lindberg, D. C., Grimes, J., Currie, C. B., Al., E., ... Ryan, L. (1997). Change blindness. *Trends in Cognitive Sciences, 1*(7), 261–267. https://doi.org/10.1016/S1364-6613(97)01080-2.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25*(2), 207–222. https://doi.org/10.1080/14640747308400340.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19*(2), 73–74. https://doi.org/10.3758/BF03337426.

Stokes, M. G., Atherton, K., Patai, E. Z., & Nobre, A. C. (2012). Long-term memory prepares neural activity for perception. *Proceedings of the National Academy of Sciences of the United States of America, 109*(6), E360–E367. https://doi.org/10.1073/pnas.1108555108.

Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., & Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron, 49*(6), 905–916. https://doi.org/10.1016/j.neuron.2006.01.021.

Summerfield, J. J., Rao, A., Garside, N., & Nobre, A. C. (2011). Biasing perception by spatial long-term memory. *Journal of Neuroscience, 31*(42), 14952–14960. https://doi.org/10.1523/jneurosci.5541-10.2011.

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. https://doi.org/10.1007/978-3-319-02868-2.

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. https://doi.org/10.1167/11.5.5.

Tatler, B. W., Hirose, Y., Finnegan, S. K., Pievilainen, R., Kirtley, C., & Kennedy, A. (2013). Priorities for selection and representation in natural tasks. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 368*(1628), 20130066. https://doi.org/10.1098/rstb.2013.0066.

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 (5).

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

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York, NY: Springer.

Võ, M. L.-H., Boettcher, S. E. P., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology.* https://doi.org/10.1016/j.copsyc.2019.03.009.

Võ, M. L.-H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision, 10*(3), 14.1–13. https://doi.org/10.1167/10.3.14.

Võ, M. L.-H., & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 23–41. https://doi.org/10.1037/a0024147.

Võ, M. L.-H., & Wolfe, J. M. (2013a). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychological Science, 24*(9), 1816–1823. https://doi.org/10.1177/0956797613476955.

Võ, M. L.-H., & Wolfe, J. M. (2013b). The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition, 126*(2), 198–212. https://doi.org/10.1016/j.cognition.2012.09.017.

Võ, 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. https://doi.org/10.1111/nyas.12667.

Voelker, L., Draschkow, D., & Võ, M. L.-H. (2018). Searching for meaning: Using pseudoword cues to investigate the formation of new object-word-connections in virtual reality. 41st European Conference on Visual Perception (ECVP) 2018 Trieste. *Perception, 48*(S1), 88.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* New York, NY: Springer.

Williams, C. C. (2010). Incidental and intentional visual memory: What memories are and are not affected by encoding tasks? *Visual Cognition, 18*(9), 1348–1367. https://doi.org/10.1080/13506285.2010.486280.

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

https://doi.org/10.3758/BF03193535.

Wolfe, J. M. (1998). Visual memory: What do you know about what you saw? *Current Biology, 8*(9), R303–R304. https://doi.org/10.1016/S0960-9822(98)70192-7.

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics, 73*(6), 1650–1671. https://doi.org/10.3758/s13414-011-0153-3.

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