To search or to like: Mapping fixations to differentiate two forms of incidental scene memory

Kyoung Whan Choe

Department of Psychology, The University of Chicago, Chicago, IL, USA

Omid Kardan

Department of Psychology, The University of Chicago, Chicago, IL, USA

Hiroki P. Kotabe

SKK Graduate School of Business, Seoul, Korea

John M. Henderson

Center for Mind and Brain and Department of Psychology, University of California, Davis, CA, USA

Marc G. Berman

Department of Psychology, The University of Chicago, Chicago, IL, USA

We employed eye-tracking to investigate how performing different tasks on scenes (e.g., intentionally memorizing them, searching for an object, evaluating aesthetic preference) can affect eye movements during encoding and subsequent scene memory. We found that scene memorability decreased after visual search (one incidental encoding task) compared to intentional memorization, and that preference evaluation (another incidental encoding task) produced better memory, similar to the incidental memory boost previously observed for words and faces. By analyzing fixation maps, we found that although fixation map similarity could explain how eye movements during visual search impairs incidental scene memory, it could not explain the incidental memory boost from aesthetic preference evaluation, implying that implicit mechanisms were at play. We conclude that not all incidental encoding tasks should be taken to be similar, as different mechanisms (e.g., explicit or implicit) lead to memory enhancements or decrements for different incidental encoding tasks.

Introduction

Memory for scenes, objects, and faces can be formed without trying, a phenomenon called *incidental memory* (Coin & Tiberghien, 1997; Craik, 2002; Craik & Lockhart, 1972; Hollingworth & Henderson, 2002). For example, after searching for objects in scenes, people retain visual information of those objects (Castelhano & Henderson, 2005; Williams, Henderson, & Zacks, 2005) and scenes (Wolfe, Horowitz, & Michod, 2007) without explicit instruction to memorize them. Often, intentional encoding creates stronger memory compared to incidental encoding (e.g., Beck, Levin, & Angelone, 2007; Tatler & Tatler, 2013); however, previous research has demonstrated that incidental memory can be better than intentional memory. People had better memory for searched objects within naturalistic scene contexts than for the objects that they were explicitly asked to memorize (Draschkow, Wolfe, & Võ, 2014; Josephs, Draschkow, Wolfe, & Vo, 2016). Evaluating the pleasantness or likability of faces without explicit instruction to memorize them has also resulted in better memory than intentionally memorizing them (Bernstein, Beig, Siegenthaler, & Grady, 2002; Grady, Bernstein, Beig, & Siegenthaler, 2002; Smith & Winograd, 1978; Warrington & Ackroyd, 1975). Together, these studies established task effects (e.g., visual search vs. intentional memorization) on visual memory (for a review, see Coin & Tiberghien, 1997; Võ & Wolfe, 2015), but two important questions remain. The first is determining what the mechanisms are that make incidental memory more (or less) successful than intentional memory. The second is whether all forms of incidental memory formation (e.g., visual search vs. aesthetic evaluation) depend on the same mechanisms.

Utilizing eye-tracking, explicit mechanisms have been proposed to explain task effects on incidental visual memory. An increased number of fixations was

related to greater subsequent recognition for natural scenes (Loftus, 1972), objects (Pertzov, Avidan, & Zohary, 2009; Tatler & Tatler, 2013), and faces (Bloom & Mudd, 1991), suggesting that tasks requiring more elaborate inspection (e.g., when judging likability of faces vs. their gender) can lead to enhanced encoding (Anderson & Lynne, 1979; Winograd, 1981). In addition, where viewers look and attend to during inspection may affect visual memory. Different viewing tasks have been shown to influence the extent and type of visual information attended to during viewing (Castelhano, Mack, & Henderson, 2009; Henderson, 2003; Kardan, Henderson, Yourganov, & Berman, 2016; Rothkopf, Ballard, & Hayhoe, 2007; Triesch, Ballard, Hayhoe, & Sullivan, 2003), opening the possibility that some portion of the task effects on incidental memory can be attributed to the deployment of visual attention (Hollingworth, 2012; Olejarczyk, Luke, & Henderson, 2014; Tatler & Tatler, 2013).

Previous research has also suggested diverse implicit mechanisms that may facilitate incidental encoding. One proposal, the depth of processing theory (Craik, 2002; Craik & Lockhart, 1972), posited that more semantic/cognitive operations are a form of "deeper processing/encoding," thereby creating stronger memory traces. However, determining which operations are deep is neither simple nor agreed upon (Baddeley, 1978) and, worse, can be circular (i.e., tasks that lead to better memory are assumed to involve deeper encoding). Moreover, for scenes, making semantic (i.e., living/nonliving) decisions failed to produce better memory than intentional memorization (Grady, McIntosh, Rajah, & Craik, 1998), evidence that is counter to the depth of processing theory.

Recent research proposes more mechanistic explanations. Task-relevant objects in a scene were found to be better remembered than the task-irrelevant objects in the same scene even when controlling for viewing duration (Castelhano & Henderson, 2005). These results suggest that viewing tasks may affect the extraction (Võ & Wolfe, 2012) and/or the retention (Maxcey-Richard & Hollingworth, 2013) of visual information within fixations. In doing so, top-down goals may have prioritized task-relevant information over task-irrelevant information (Draschkow & Võ, 2016; Tatler et al., 2013; Tatler & Tatler, 2013). In addition, viewing tasks (e.g., visual search) may have contributed to better integration of bottom-up visual information and contextual scene semantics, leading to stronger memory representations than intentional encoding (Draschkow et al., 2014; Josephs et al. 2016; Võ & Wolfe, 2015). Self-related processing of information (Rogers, Kuiper, & Kirker, 1977) and, surprisingly, merely making a voluntary choice (Murty, DuBrow, & Davachi, 2015) have been shown to enhance memory. As encoding and retrieval processes

are interdependent (Kolers, 1973; Morris, Bransford, & Franks, 1977; Roediger & Guynn, 1996; Tulving & Thomson, 1973), retrieval activities should be examined to fully understand memory performance. These studies suggest that there are many factors that may increase visual memory from incidental encoding.

By reanalyzing previous eye-tracking data (Luke, Smith, Schmidt, & Henderson, 2014; Nuthmann & Henderson, 2010; Pajak & Nuthmann, 2013), the current study aims to investigate factors that affect intentional and incidental scene memory, and importantly, aims to differentiate contributions of explicit and implicit mechanisms. We compared an intentional memorization task and two different incidental encoding tasks: (a) a visual search task (Castelhano & Henderson, 2005; Draschkow et al., 2014; Hollingworth, 2012; Josephs et al., 2016; Olejarczyk et al., 2014; Võ & Wolfe, 2012; Williams et al., 2005) and (b) an aesthetic preference evaluation task (Bernstein, Beig, Siegenthaler, & Grady, 2002; Coin & Tiberghien, 1997; Craik, 2002; Craik & Lockhart, 1972; Grady et al., 2002; Smith & Winograd, 1978; Warrington & Ackroyd, 1975), both of which were known to modulate incidental memory. Similarly, Wolfe et al. (2007) used the intentional memorization, visual search, and preference evaluation tasks to study scene memory, although they did not compare those tasks directly.

To study possible mechanisms for these different incidental and intentional encoding task effects on scene memory, we systematically compared eye movements during natural scene encoding, eye movements during retrieval, and subsequent scene memory between the intentional memorization, visual search, and aesthetic preference evaluation tasks. Specifically, we examined the number of fixations and fixation maps (Henderson, 2003; Pomplun, Ritter, & Velichkovsky, 1996; Wooding, 2002) obtained while participants were performing the viewing tasks. Comparing fixation maps within and across tasks could be useful to interrogate the spatial manifestation of overt attention for different tasks and to investigate the role of visual attention in the formation of incidental and intentional memory. To perform this investigation, we examined the fixation maps of each scene viewed by a group of participants who were then tested for subsequent scene memory, which we call G1 hereafter. We compared those participant's fixation maps to the averaged fixation maps from a different group of participants who were told to memorize the presented scenes but not later tested on them (i.e., no subsequent scene memory test); we call this group G2M. By establishing such template fixation maps for scene encoding, we were able to quantify the deployment of visual attention on a trial-by-trial basis, allowing us to implement trial-level generalized linear mixed-effect models (GLMMs) that

can isolate effects of viewing tasks, fixation counts, and visual attention, respectively. This procedure mimics analyses from recent eye-tracking studies, which used linear mixed-effect model analyses to study task effects on fixation durations (Einhäuser & Nuthmann, 2016; Nuthmann, 2017) and object memory (Draschkow & Võ, 2016; Josephs et al., 2016; Tatler et al., 2013). We also examined the relationship between scene memory and scene aesthetic preference by collecting scene preference ratings from a group of participants that were recruited online, called G3, because scene aesthetics are known to be significantly associated with scene memorability (Isola, Xiao, Parikh, Torralba, & Oliva, 2014).

We found that scene memorability *decreased* after visual search (one incidental memory task) compared to intentional memorization and that the scene preference task (another incidental memory task) produced the greatest memory (outperforming even intentional memorization). Although the number of fixations was significantly associated with scene memory, it could not explain the changes in scene memorability across tasks. By analyzing fixation maps, we found that the allocation of overt attention could fully explain why visual search impairs incidental scene memory, but it could not explain why the preference task boosts scene memory. Therefore, we conclude that not all incidental-encoding tasks are similar as different mechanisms lead to memory enhancements or decrements for different incidental encoding tasks.

Methods

Datasets

The current study is based on a previously collected eye-tracking dataset (Luke et al., 2014; Nuthmann & Henderson, 2010; Pajak & Nuthmann, 2013), which has been used in prior publications (Einhäuser & Nuthmann, 2016; Kardan, Berman, Yourganov, Schmidt, & Henderson, 2015; Kardan et al., 2016; Luke et al., 2014; Nuthmann, 2017; Nuthmann & Henderson, 2010). A subset of the dataset (i.e., 36 participants who were tested for scene memory—G1) was analyzed in Nuthmann & Henderson (2010), and the full data (i.e., 72 participants—both G1 and G2) were analyzed in Einhäuser & Nuthmann (2016), Luke et al. (2014), Nuthmann (2017), Kardan et al. (2015), Kardan et al. (2016), and Pajak & Nuthmann (2013). A new dataset (G3) was collected in this study to obtain preference ratings for scenes so that we can examine the relationship between scene memory and scene preference. Because this study is a reanalysis of a previously collected dataset, the sample sizes were determined for

different analysis purposes. Here we report all experimental conditions, measures, data exclusions, and data losses. All of our analysis code can be downloaded from the Center for Open Science (https://osf.io/2k6zs/).

Participants

Two groups of 36 undergraduate students (G1 and G2) from the University of Edinburgh participated in the lab experiment. Both G1 and G2 participants performed the first phase (Encoding phase) of the experiment in which viewing task was manipulated, and shortly after the Encoding phase, only G1 participants engaged in the second phase testing scene recognition (Test phase). All 72 participants had 20/20 corrected or uncorrected vision, were naive to the purposes of the experiment, and provided informed consent as administered by the Institutional Review Board of the University of Edinburgh.

In addition, 60 adults (G3) participated in the online experiment via the online labor market Amazon Mechanical Turk (AMT), in which workers complete human intelligence tasks (HITs) for requestors. Worker qualifications included location in the United States, a HIT approval rate greater than or equal to 98%, and number of HITs approved greater than or equal to 1,000. The participants were recruited using TurkPrime (Litman, Robinson, & Abberbock, 2017), and participation was allowed only between 10 a.m. to 6 p.m. CST. G3 participants were also naive to the purpose of the experiment, and provided informed consent as administered by the Institutional Review Board of the University of Chicago. Sample size was set based on our goal of more than 50 valid ratings per scene, which has yielded highly reliable ratings in our previous research (Ibarra et al., 2017; Kotabe, Kardan, & Berman, 2016, 2017). All experiments adhered to the Declaration of Helsinki guidelines.

Stimuli

In the Encoding phase and in the online experiment, 135 full-color (32 bit) 800 × 600 pixel photographs of real-world, indoor and outdoor scenes were used. Each scene contained one predefined search object. Search targets were chosen such that they occurred only once in the scene, did not appear at the center of the scene, were not occluded, and were large enough to be easily recognized. The resulting mean object width and height was 2.82° (*SD* = 0.84°) and 2.77° (*SD* = 0.88°), respectively. The scene images are available from the author J. M. H. upon request. In the Test phase, 66 of those scenes were presented in an identical form ("old" stimuli in the recognition task), and the other 66 scenes

were presented in a horizontally mirrored form ("altered" stimuli in the recognition task), and the remaining three scenes were not presented. In addition, 22 new scenes ("new" stimuli in the recognition task) were presented in the Test phase. In this paper, we present the results of the 132 scenes that were used in both the Encoding and Test phases.

Lab experiment apparatus and procedures

Participants sat 90 cm away from a 21-in. CRT monitor and placed their head on a chin and forehead rest. The scenes were displayed full screen in their native resolution and subtended $25.8° \times 19.4°$ in visual angle. Eye movements were recorded from the right eye, although viewing was binocular, via an SR Research (Ottawa, Ontario, Canada) Eyelink 1000 eye tracker with a sampling rate of 1000 Hz. The experiment was controlled with SR Research Experiment Builder software. The eye tracker was calibrated using the built-in nine point calibration routine. The calibration was not accepted until the average error was less than $0.49°$ and the maximum error was less than $0.99°$. Participants were recalibrated at the start of each block.

In the Encoding phase, the 135 scenes were split into three blocks of 45, and the scene split was the same across participants. During each block, participants were instructed to perform one of three tasks on the scenes presented for 8 s: (a) memorize the scene for a subsequent old/new recognition test, (b) search for an object, or (c) make an aesthetic preference judgment. All scenes were presented for 8 s. The task assignment and order of each block were determined by a dual-Latin square design and counterbalanced across participants. Participants viewed each scene for 8 s and performed the viewing tasks assigned to each block while their eye movements were recorded. Within each task block, the scenes were presented in a random order. The participants completed all three blocks.

During the scene memorization task block, participants were asked to intentionally memorize the scenes for 8 s but were not required to make any responses during encoding. During the search task block, an object name was presented (Ariel font, vertical height of $1.62°$) for 800 ms followed by a central fixation cross for 200 ms and the scene for 8 s. If the participant located the search target, they responded by pressing a trigger on a controller (Microsoft Sidewinder). The scene remained on the screen until the 8 s was over. The search task started with three practice trials. The behavioral results showed that in 88.3% of the search trials, the target was found by the participant, as confirmed by their fixation coordinates when they pressed the response key to indicate that the object was found. During the aesthetic preference task block, participants were instructed to examine each scene and then rate how much they liked it. After each scene, a response screen appeared asking the participant to respond on a four-point like/dislike affective rating scale (1 = *dislike*, 4 = *like*). Responses were made via four buttons on the controller. However, those ratings were lost due to a coding error, and therefore could not be used in the preference analysis. So, we conducted a separate preference task, which will be described below. The preference task started with three practice trials.

After a short break, G1 engaged in the Test phase, a scene recognition task. Before the task, participants were informed that their memory would be tested for all the scenes they had previously encountered, not just the scenes they had been instructed to remember in the memorization block. In each trial, a scene was shown for 3 s, and participants were asked to identify whether the scene was "old" (encountered in the Encoding phase during *any* block, not just the memorization block, and presented in an identical form), "altered" (encountered in the Encoding phase but presented in a horizontally mirrored form), or "new." A total of 154 scenes, consisting of seven categories of 22 scenes—old and memory, old and search, old and preference, altered and memory, altered and search, altered and preference, and new—were used in the recognition task.

Eye movement analyses

Raw eye movement data from the Encoding phase were preprocessed using Eyelink Data Viewer (SR Research, Toronto, Canada). Saccades were defined with a $50°$/s velocity threshold using a nine-sample saccade detection model. Fixations were excluded from analysis if they were preceded by or co-occurred with blinks, were the first or last fixation in a trial, or had durations less than 50 ms or longer than 1200 ms. For each scene and each participant, discrete fixations (blue crosses in Figure 2a) and fixation durations during the 8 s of scene viewing were identified. The *fixation count* is the number of discrete fixations, regardless of their duration, that landed on the scenes. Excluding outlier trials (outside ± 2 *SD* range), in which fixation counts were ≤ 14 (218 trials) or ≥ 34 (52 trials), did not change the results.

The fixation map analyses were performed using custom MATLAB (The MathWorks, Natick, MA) scripts. First, an individual fixation map (Henderson, 2003; Pomplun et al., 1996; Wooding, 2002) of a participant viewing a scene (the upper row in Figure 2d) was constructed by convolving a Gaussian kernel over its duration-weighted fixation locations. The full width at half maximum of the Gaussian kernel was set to $2°$ (i.e., $\sigma = 0.85°$) to simulate the central foveal
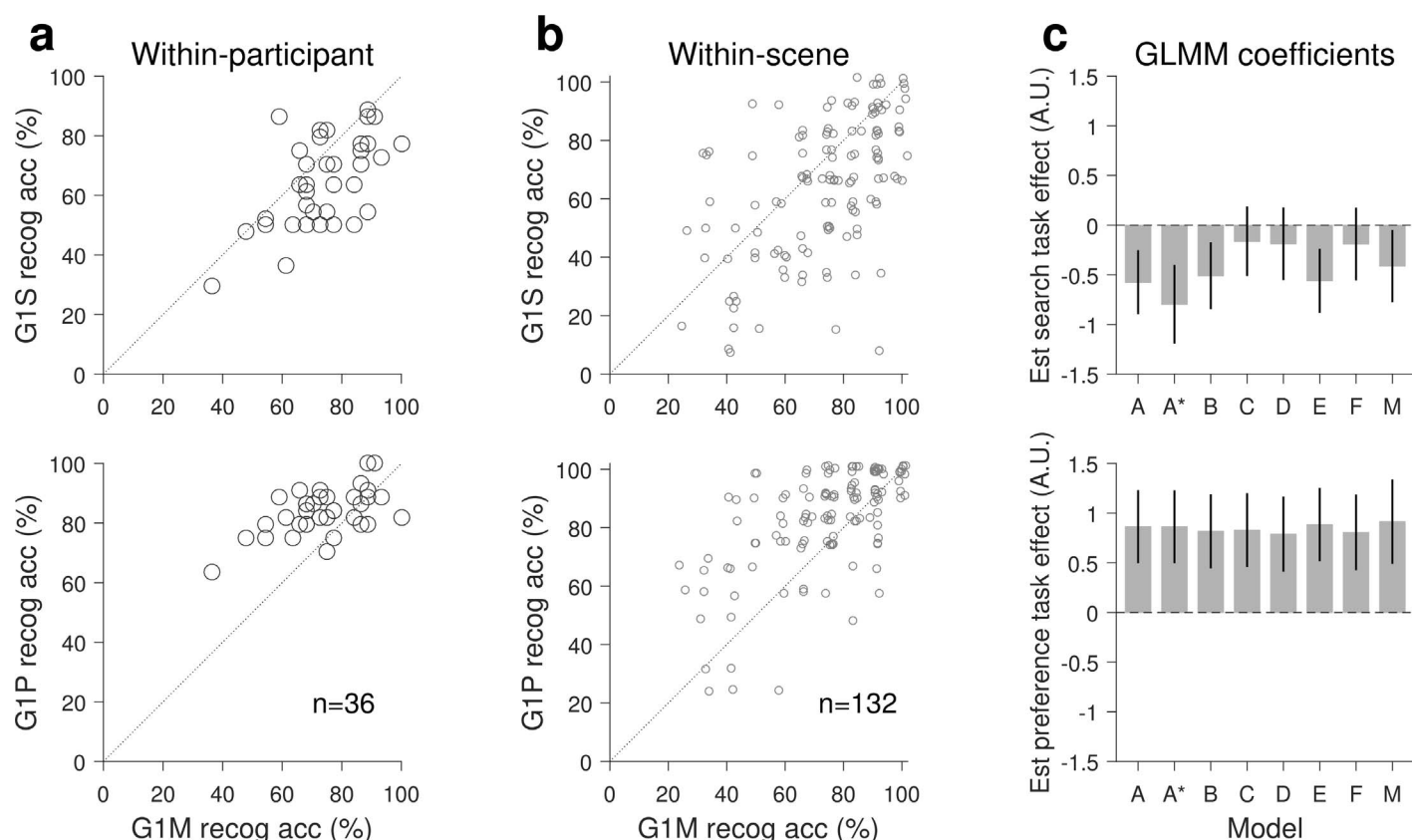
Figure 1. The effects of viewing tasks on scene memory. (a) Within-participant comparisons of the average recognition accuracy of each participant in the Test phase. G1S, G1M, and G1P are the visual search, intentional memorization, and preference evaluation tasks, respectively, in which G1 participants performed and viewed the scenes. Circles indicate each participant, and the dotted lines are identity lines. (b) Within-scene comparisons of the average recognition accuracy of each scene in the Test phase. Circles indicate each scene. To discern overlapping data points, the points were jittered by adding a small random noise. (c) Estimated task effects (i.e., beta coefficients) of the search task (upper) and preference task (lower) on scene memory. Seven different trial-level, generalized linear mixed-effects (GLMM) analyses were performed (Models A–M; see Table 1 for the model description). The error bars indicate 95% CI of the estimates.

vision (Rayner, 1998) and to take into account the measurement errors of video-based eye trackers (Choe, Blake, & Lee, 2016; Drewes, Zhu, Hu, & Hu, 2014; Wyatt, 2010). Then a template fixation map for memorizing a scene was constructed by averaging the individual fixation maps of G2 participants who saw that scene during the intentional memorization (G2M). Before averaging, each fixation map was normalized so that the sum of all pixel intensities of a map equaled one. Finally, the individual fixation maps of G1 participants who saw a scene during the memorization (G1 Memory; G1M), search (G1 Search; G1S), and preference (G1 Preference; G1P) task blocks, respectively, were compared to the template fixation map for encoding that scene. The *fixation map similarity* (to the template map) of an individual fixation map is the Fisher *z*-transformed correlation coefficient between the vectorised individual map and the vectorised template map.

To quantify the ability of an ideal observer to predict whether a scene would be correctly recognized in the subsequent test based on the fixation counts and the fixation map similarity, we performed receiver operating characteristic (ROC) curve analyses (Swets, 1973). ROC curves were constructed by plotting the rate of correctly recognized trials that were predicted to be correct (i.e., Hit) against the rate of incorrect trials that were predicted to be correct (i.e., False alarm) at various threshold settings. The area under curve (AUC), which is equivalent to the probability of the correct prediction for a given pair of correct and incorrect trials, of the ROC curves was calculated. Whether the observed AUC values were statistically significant was tested using a permutation test, in which the null distribution was constructed by randomly shuffling the correctness label of each trial, generating ROC curve based on the new labels, calculating AUC, and repeating the procedure for 10,000 times.
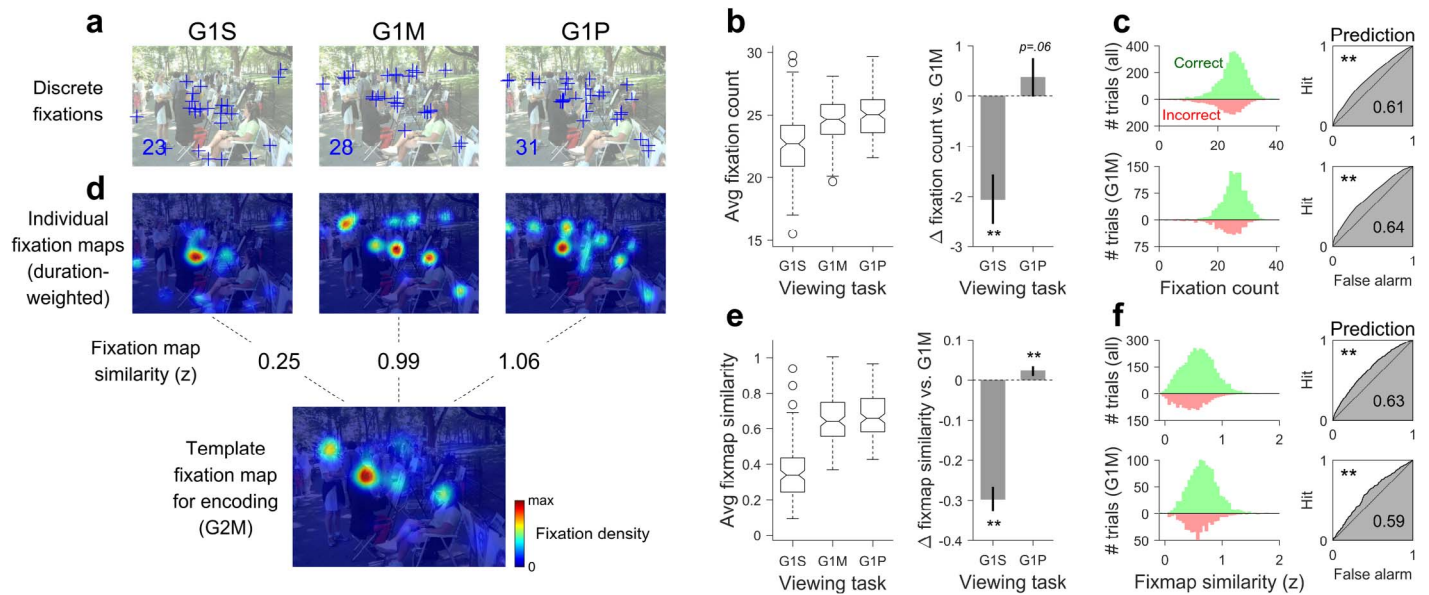
Figure 2. Illustration of the fixation map analyses. (a) Discrete fixation locations of a scene from three different participants, who were performing the visual search (G1S), intentional memorization (G1M), and preference evaluation (G1P) tasks, respectively. G1S participants were asked to locate a *roller skate* around the center. Each blue cross represents a fixation, and the numbers in the left bottom indicate the *fixation counts*. (b) The effects of viewing tasks on fixation count. Left: The fixation counts for each scene were averaged across participants within task. Circles indicate the scenes with outlier values. Right: The differences in fixation count between G1S versus G1M and G1P versus G1M. The error bars indicate 95% CI of the means. ** $p < 0.001$ (paired t test). (c) The effects of fixation count on scene memory. Top row: All (i.e., G1S, G1M, and G1P) trials. Bottom row: G1M trials only. Left panels: Contrasts of the histograms of fixation count of each trial between the correct recognition trials (green) and the incorrect recognition trials (red). Right panels: Receiver operating characteristic curves. The horizontal and vertical axes specify the rate of false alarms (i.e., the incorrect trials that were wrongly classified as correct) and hits (i.e., the correct trials that were correctly classified), respectively. ** $p < 0.001$ (permutation test). (d) Fixation map similarity analysis. Top row, individual fixation maps overlapped on the original image. The individual fixation maps were constructed by convolving a Gaussian kernel over its duration-weighted fixation locations (see Methods). The color corresponds to the fixation density, thus the red areas indicate dense fixations toward those regions across participants. The *fixation map similarity* is the Fisher *z*-transformed correlation coefficient between the individual map and the template map (bottom). The template map was constructed by averaging the individual fixation maps of G2 participants who saw that scene during the intentional memorization (i.e., G2M). (e) The effects of viewing tasks on fixation map similarity. (f) The effects of fixation map similarity on scene memory.

To investigate the factors affecting scene memory, we ran the trial-level Generalized Linear Mixed Effects Model (GLMM) analyses using the Statistics and Machine Learning Toolbox in MATLAB. We opted for the GLMM because it allows controlling for different scenes and participants simultaneously (Baayen, Davidson, & Bates, 2008). In each model, the dependent variable was the correctness of single trials in the scene recognition task (1: correct, 0: incorrect), and the binomial distribution was used to specify the dependent variable. Individual analyses contained different effects, which are described in the Results section. When running GLMMs with a binomial distribution, the assumption of normally distributed residuals is no longer required, and the residual diagnostic plots often used for transforming variables in the standard, linear mixed-effect models are not applicable. So, we used the *z*-scored variables for simplicity and better interpretability (Schielzeth, 2010).

As for random effects, the by-participant random intercepts and slopes for the viewing tasks and the by-scene random intercepts were included, and the random effects covariance matrices were reported, following the guidelines of Barr, Levy, Scheepers, and Tily (2013) and of Matuschek, Kliegl, Vasishth, Baayen, and Bates (2017). In fitting the models, the binomial distribution was specified for the dependent variable, and Laplace approximations were used to approximate the likelihood (Bolker et al., 2009). All models converged to a stable solution.

Online experiment procedures

To investigate the relationship between scene memorability and aesthetic preference, we conducted an online version of the aesthetic preference task on AMT and collected the scene ratings from a larger pool of

participants, G3. First, participants were taken to a web page containing a consent form and showing only high-level information about the purpose of the study. While participants viewed this page, all 135 scenes were preloaded to prevent loading delay during the task. Once preloading was complete, participants could agree to the consent form and proceed to the experiment. Next, participants were taken to an instruction page where they were told that they would be presented a series of 135 indoor and outdoor scenes and that we wanted them to rate how much they disliked or liked each scene using an affective rating scale with four options: "Dislike," "Somewhat Dislike," "Somewhat Like," and "Like." A screenshot of the rating scale was shown to help them prepare. Then, participants started the preference task. Each scene was presented one by one on a plain white background along with text asking "How much do you dislike or like this scene?" and the four-point like/dislike rating scale. The scene images were dynamically resized to be presented in the maximal size, accommodating for different monitor resolutions of each participant, while still allowing the text and the rating scale to be seen on the same page without scrolling down. Participants interacted with the rating scale by clicking on radio buttons that were attached to each option, and they could provide a preference rating at any time. Immediately after submitting a rating, the next scene was loaded automatically until all the scenes were rated. The order of scene presentation was randomized for each participant. After the task, participants were asked for demographic information and feedback on the study. Finally, participants were given a random code that they had to submit on AMT to verify their completion of the experiment.

Data and code availability

The eye-tracking dataset is available from the author J.M.H. upon request. The fixation map analyses were performed using custom MATLAB scripts, which are available at https://osf.io/2k6zs/. All statistical tests were performed using MATLAB R2015b. The *fitglme* function was used to perform GLMM analyses, the *corrcoef* function was used to calculate the effect size (95% CI) of Pearson's correlations, and the *polyfit* and *polyconf* functions were used to calculate the 95% confidence band of a regression. The effect size of *t* tests is reported with Hedges' *g* (Hedges, 1981), the standardized difference between the means of the two groups, and its 95% CI were calculated using the MES toolbox v1.4 (Hentschke & Stüttgen, 2011), which can be obtained from https://www.mathworks.com/matlabcentral/fileexchange/32398. The intraclass correlation was calculated using the *ICC* function, which

can be obtained from https://www.mathworks.com/matlabcentral/fileexchange/22099.

Results

Task effects on scene memory

We first examined whether there were differential effects among the incidental encoding tasks (i.e., visual search and aesthetic preference evaluation) and the intentional encoding (i.e., intentional memorization) task on scene memory. In the Encoding phase, G1 participants viewed each scene for 8 s and were asked to perform one of three tasks on the scene: (a) Try to memorize it (G1M), (b) search for an object in the scene (G1S), or (c) evaluate aesthetic preference for the scene (G1P). Each participant performed all three tasks, viewing 45 scenes in each task (see Methods for details). Shortly after the Encoding phase, their memory for all of the scenes was assessed using a recognition task. We used the recognition accuracy from G1M as baseline intentional memory and compared it with the subsequent incidental memory from G1S and G1P on a within-participant (Figure 1a) and a within-scene basis (Figure 1b). Compared to the memorization task condition, participants' overall recognition performance was poor in the search task condition (the upper panel in Figure 1a): paired $t(35) = 4.28$, $p < 0.001$, Hedges' $g = 0.64$ [0.34, 1.01]; and better in the preference task condition (the lower panel in Figure 1a): paired $t(35) = 5.22$, $p < 0.001$, Hedges' $g = 0.88$ [0.54, 1.29]. Also, we found that a scene was less recognized (i.e., less memorable) if it was seen in the search task than if it was seen in the memorization task (the upper panel in Figure 1b): paired $t(131) = 5.17$, $p < 0.001$, Hedges' $g = 0.42$ [0.27, 0.59]. Similar to the incidental memory boost observed in words and faces, we found that rating a scene's aesthetic value during encoding significantly increased its memorability compared to the intentional memorization (the lower panel in Figure 1b): paired $t(131) = 6.75$, $p < 0.001$, Hedges' $g = 0.53$ [0.37, 0.71].

To estimate and quantify task effects on scene memory, we conducted a trial-level GLMM analysis, which allows controlling for different scenes and participants simultaneously. As expected, our simplest trial-level GLMM analysis (Model A in Table 1) established those task effects by showing significant regression coefficients for the variables associated with viewing tasks (Figure 1c). In this analysis, the dependent variable was the correctness of all 4,752 trials in the recognition test (1: correct, 0: incorrect). The fixed effects were the viewing task type (G1S, G1M, and G1P), scene orientation (i.e., whether a

Model	Fixed effects[a]	DF	Adj. R^2	BIC[b]
A	Viewing task[c] $\times$ scene orientation[d]	4746	27.14%	4730.2
A*	Viewing task $\times$ scene orientation + task engagement duration[e]	4745	27.17%	4735.4
B	Viewing task $\times$ (scene orientation + fixation count[f])	4743	29.96%	4667.8
C	Viewing task $\times$ (scene orientation + fixation map similarity[g])	4743	28.90%	4704.0
D	Viewing task $\times$ (scene orientation + fixation count + fixation map similarity)	4740	31.10%	4660.5
E	Viewing task $\times$ (scene orientation + scene preference[h])	4743	27.34%	4742.8
F	Viewing task $\times$ (scene orientation + fixation count + fixation map similarity + scene preference)	4737	31.18%	4675.7
M	Viewing task $\times$ scene orientation $\times$ encoding-recognition fixation map similarity[i]	4740	33.74%	4670.1

Table 1. Summary of trial-level generalized linear mixed-effects model (GLMM) analyses. *Notes*: Dependent variable: the correctness of all 4,752 trials in the recognition task (1: correct, 0: incorrect). All models included the by-participant random intercepts and slopes for the viewing tasks and the by-scene random intercepts, i.e., (1 + Viewing Task | Participant) + (1 | Scene) in the Wilkinson-Rogers (Wilkinson & Rogers, 1973) notation. In fitting the models, the binomial distribution was specified for the dependent variable, and Laplace approximations were used to approximate the likelihood. All models converged to a stable solution. [a] The fixed effects were specified in the Wilkinson-Rogers notation. [b] The Bayesian information criterion (Schwarz, 1978). [c] The types of viewing task— intentional memorization (G1M), visual search (G1S), and preference evaluation (G1P) tasks—were entered as nominal effects. [d] Whether the scene was presented in an identical form or in a horizontally mirrored form was coded (0: identical, 1: mirrored). [e] The normalized task engagement durations. During the memorization and preference tasks, we assumed that a participant was fully engaged in scene viewing for 8 s during a trial, so we entered 1 for the G1M and G1P trials. During the search task, we assumed that a participant was engaged in during search, so we entered the search duration / 8 s (range: 0–1) for the G1S trials. [f] The z-scored fixation counts during 8 s of viewing the scene (Figure 2c) were entered. [g] The z-scored similarity values between the individual fixation map and template map (Figure 2f) were entered. [h] The z-scored scene preference ratings obtained from G3 were entered. [i] The z-scored similarity values between the encoding and recognition fixation maps of each individual (Figure 6a) were entered.

scene was horizontally flipped in the recognition test), and their interaction. Also, the by-participant random intercepts and slopes for the tasks and the by-scene random intercepts were entered into the model. This model ($df = 4746$) explained 27.14% of the variance (adjusted R^2), and the full results are presented in Supplementary Table 1. Model A confirmed a significant negative effect of the search task ($\beta = -0.58$ [$-0.90, -0.25$], $p < 0.001$) and a significant positive effect of the preference task ($\beta = 0.86$ [0.49, 1.23], $p < 0.001$) on scene memory compared to intentional memorization. Also, the model showed that presenting scenes in a flipped form during test (i.e., the scene orientation effect) significantly decreased their recognition ($\beta = -1.04$ [$-1.44, -0.65$], $p < 0.001$). The interaction between task type and scene orientation was not significant, $F(2, 4746) = 1.61$, $p = 0.20$.

In the following sections, we examined the number of fixations and the fixation maps obtained while participants were performing the viewing tasks to investigate possible attentional mechanisms for these different incidental and intentional encoding task effects on scene memory.

Examination of fixation count

Previously, it was shown that an increased number of fixations was related to greater subsequent recognition for natural scenes (Loftus, 1972), objects (Pertzov et al., 2009; Tatler & Tatler, 2013) and faces (Bloom & Mudd,

1991). Therefore, we examined whether *fixation count* (i.e., the number of discrete fixations that landed on the scene during viewing; Figure 2a) could explain the task effects on scene memory. To do so, we first checked the relationships between the viewing task and fixation counts (Figure 2b) and between fixation counts and scene memory (Figure 2c). Then we conducted a trial-level GLMM analysis (Model B in Table 1) to examine whether fixation counts affected the estimation of the search task and the preference task effect coefficients (the vertical axes in Figure 1c).

Regarding the relationship between the tasks and fixation counts, the participants made a smaller number of fixations in the search task than in the memorization task (Figure 2b): paired $t(131) = 8.25$, $p < 0.001$, Hedges' $g = 0.91$ [0.68, 1.17], compared on a scene-by-scene basis, and made a comparable number of fixations in the preference task: paired $t(131) = 1.89$, $p = 0.06$, Hedges' $g = 0.22$ [$-0.00, 0.46$]. Regarding the relationship between fixation counts and scene memory, correct recognition trials were associated with significantly greater fixation counts during encoding than incorrect recognition trials (left panels in Figure 2c): ANOVA correct (green) versus incorrect (red) trials, $F(1, 4750) = 168.6$ for all trials; $F(1, 1582) = 93.9$ for only intentional memorization (G1M) trials, both ps < 0.001, consistent with previous research (Bloom & Mudd, 1991; Loftus, 1972). Also, receiver operating characteristic (ROC) analyses (see Methods) confirmed that whether a scene would be correctly recognized later could be predicted using fixation counts at levels
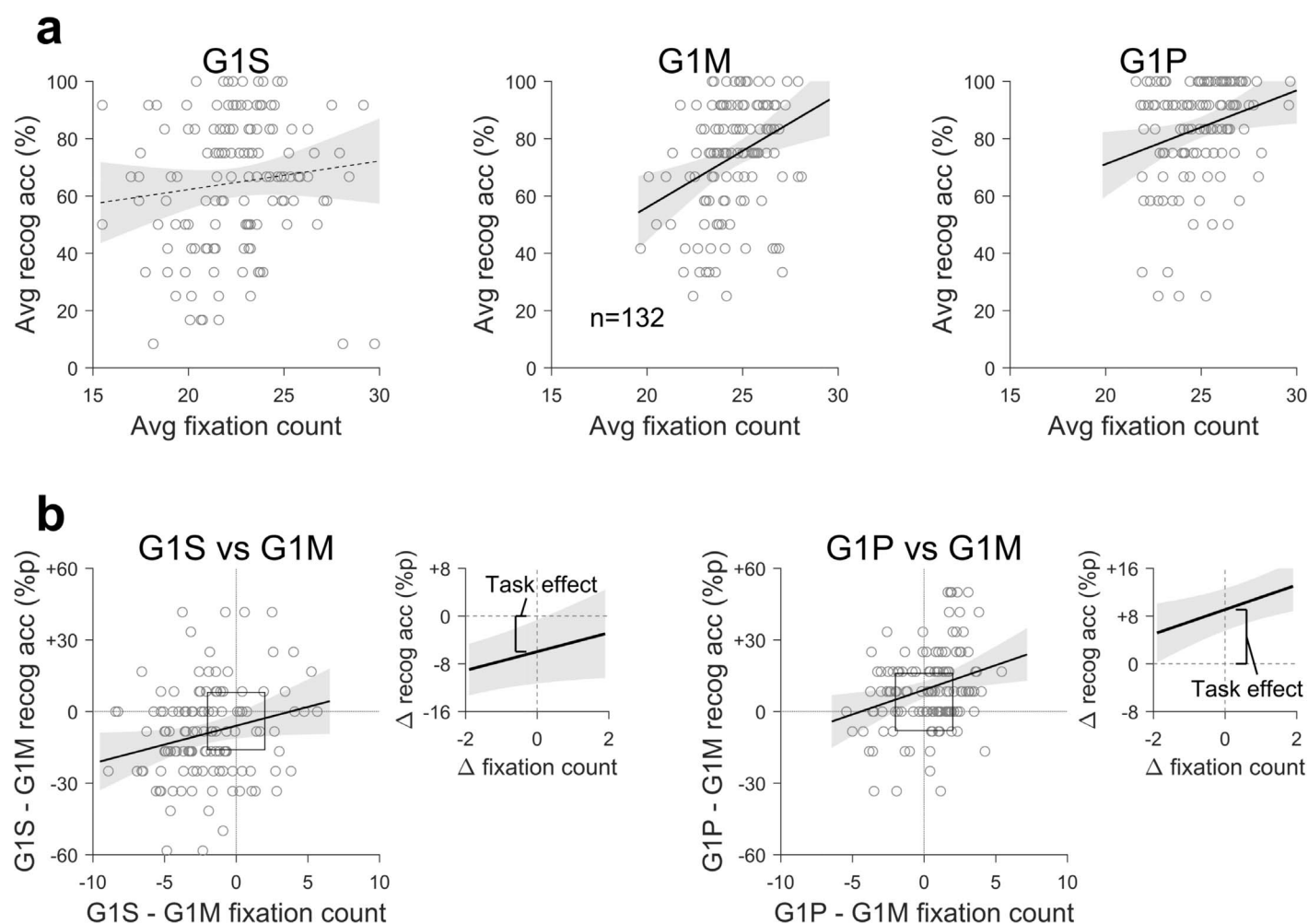
Figure 3. Results for the fixation count. (a) Relationships between the fixation count and recognition accuracy within each task. Each circle represents a scene. The solid lines are the significant regression lines (*p*s < 0.001 and 0.004 for the G1M and G1P panels, respectively), the dashed line is the nonsignificant (*p* = 0.19) regression line, and the gray shades represent the 95% confidence bands. (b) Comparisons of the differences in fixation count and recognition accuracy between tasks. The solid lines are the significant regression lines (*p*s = 0.01 and 0.001 for the G1S vs. G1M and G1P vs. G1M panels, respectively). The areas marked with a black square were zoomed and presented at the right side.

significantly above chance (right panels in Figure 2c): area under curve (AUC) = 0.61 for all trials and 0.64 for G1M trials, both *p*s < 0.001 from permutation testing.

To examine the extent to which fixation counts could explain the task effects on scene memory, we conducted a trial-level GLMM analysis (Model B in Table 1), in which fixation count and its interactions with task type were added to Model A as fixed effects. Model B (*df* = 4743) explained 29.96% of the variance, and the full results are presented in Supplementary Table 2. Model B showed a significant positive effect of fixation counts (β = 0.75 [0.57, 0.94], *p* < 0.001) and significant interactions between the tasks and fixation counts (βs = −0.49 [−0.71, −0.28] and −0.45 [−0.69, −0.21] for the search and preference tasks, respectively; both *p*s < 0.001). However, the effects of search and preference

tasks on scene memory both remained significant in Model B (Figure 1c): βs = −0.51 [−0.85, −0.17] and 0.82 [0.44, 1.19], *p* = 0.003 and < 0.001, respectively.

These GLMM results can be illustrated by correlating fixation count and scene memory (Figure 3). The correlation between fixation count and recognition accuracy was strongest in the memorization task: Pearson's *r*(130) = 0.34 [0.17, 0.48], *p* < 0.001 (the G1M panel in Figure 3a); and decreased in the search and preference tasks: Pearson's *r*(130) = 0.12 [−0.06, 0.28] and 0.25 [0.08, 0.40], *p*s = 0.19 and 0.004, respectively (the G1S and G1P panels), consistent with a significant positive effect of fixation counts and significant negative task interactions in Model B. Also, the memorability changes across tasks in each scene were significantly correlated with the fixation count changes from memorization (Figure 3b): Pearson's *r*(130) = 0.22

[0.05, 0.38] and 0.28 [0.11, 0.43], $ps = 0.01$ and 0.001, for the search and preference tasks, respectively. But, when fixation counts were controlled for across tasks, we still observed the memorability change intercepts that were significantly different from 0 (the zoomed panels in Figure 3b), i.e., there were still significant task effects. Together, these results show that fixation counts were significantly associated with scene memory and interacted with viewing tasks (e.g., a stronger effect in the memorization task), but could not explain away the variance due to the main effect of task-type.

Examination of fixation map similarity

Next, we examined the relationship between the viewing tasks, scene memory, and individual fixation maps. By measuring *where* viewers look in scenes (i.e., not just fixation counts, but locations and durations of those fixations), eye-tracking enables the examination of the allocation of overt attention under different viewing tasks (Castelhano et al., 2009; Henderson, 2003; Kardan et al., 2016), which can be illustrated with fixation maps (Henderson, 2003; Wooding, 2002). In relating fixation maps and scene memory, we assumed that scene regions that are highly fixated during intentional memorization are memory-relevant and support scene encoding. This assumption is based on the fact that ocular targeting is guided by both bottom-up information and top-down goals (Henderson, 2007, 2011, 2013). The average of the individual G2M (i.e., G2 participants intentionally memorizing a scene) fixation maps provided information for where the memory-relevant regions were in a scene (e.g., the bottom panel in Figure 2d), so we used the averaged map as a template for encoding that scene. We hypothesized the more overt attention allocated to those memory-relevant regions during scene encoding, the better scene memory would be. This hypothesis is based on the supposition that visual attention is necessary for visual memory (Castelhano & Henderson, 2005; Draschkow et al., 2014; Hollingworth, 2012; Hollingworth & Henderson, 2002; Olejarczyk et al., 2014; Tatler & Tatler, 2013; Williams et al., 2005; Wolfe et al., 2007).

The extent to which the memory-relevant regions were attended by a participant during encoding a scene was quantified by calculating the *fixation map similarity* (i.e., the Fisher *z*-transformed correlation coefficients; Figure 2d) between the participant's fixation map and the template fixation map for encoding that scene. Regardless of the viewing tasks, the averaged G2M fixation map was used as the template for comparison. The search and memorization tasks often produced qualitatively different fixation maps. For example, when asked to locate a *roller skate,* participants looked at objects in the scene that are related to roller skates such as people's feet and shoes (G1S panel in Figure 2a), yielding a different pattern of fixations compared to attempting to memorize the scene. As such, the similarity of G1S fixation maps to the template fixation map was significantly lower than that of G1M fixation maps (Figure 2e): paired $t(131) = 19.2$, $p < 0.001$ across scenes, Hedges' $g = 2.13$ [1.82, 2.50], compared on a scene-by-scene basis. In contrast, and somewhat paradoxically, the preference task resulted in significantly higher fixation map similarity, although small (see the right panel of Figure 2e), than the memorization task itself: paired $t(131) = 3.48$, $p < 0.001$ across scenes, Hedges' $g = 0.18$ [0.08, 0.29].

We then examined the relationship between fixation map similarity and scene memory. Consistent with our hypothesis, correct recognition trials were associated with significantly higher fixation map similarity than were incorrect recognition trials (left panels in Figure 2f): ANOVA correct (green) versus incorrect (red) trials, $F(1, 4750) = 204.6$ for all trials; and $F(1, 1582) = 29.9$ for G1M trials, both $ps < 0.001$. ROC analyses also showed that whether a scene would be correctly recognized later could be predicted using fixation map similarity significantly above chance (right panels in Figure 2f): area under curve (AUC) = 0.63 for all trials and 0.59 for G1M trials, both $ps < 0.001$ from permutation testing. These results suggest that the more one attends to memory-relevant regions, the better encoded that scene will be for subsequent test.

Having established the effect of fixation map similarity on scene memory, we examined the extent to which fixation map similarity can explain the task effects on scene memory. We conducted two GLMM analyses (Models C and D in Table 1). In Model C, fixation map similarity and its interactions with task type were added to Model A as fixed effects. Model C ($df = 4743$) explained 28.90% of the variance, and the full results are presented in Supplementary Table 3. Model C showed a significant positive effect of fixation map similarity on scene memory ($\beta = 0.29$ [0.13, 0.46], $p < 0.001$) and nonsignificant interactions between the search task and fixation map similarity ($\beta = 0.15$ [−0.08, 0.38], $p = 0.19$) and between the preference task and fixation map similarity ($\beta = 0.08$ [−0.16, 0.33], $p = 0.51$). Interestingly, Model C showed a nonsignificant effect of the search task ($\beta = -0.16$ [−0.51, 0.19], $p = 0.36$; Figure 1c) but still a significant effect of the preference task ($\beta = 0.83$ [0.46, 1.20], $p < 0.001$). Similar results were obtained in Model D, in which fixation count, fixation map similarity, and their interactions with task were added to Model A as fixed effects. Model D ($df = 4740$) explained 31.10% of the variance, and the full results are presented in Supplementary Table 4. Consistent with Model C, Model D showed a significant effect of the preference task ($\beta = 0.79$ [0.41,
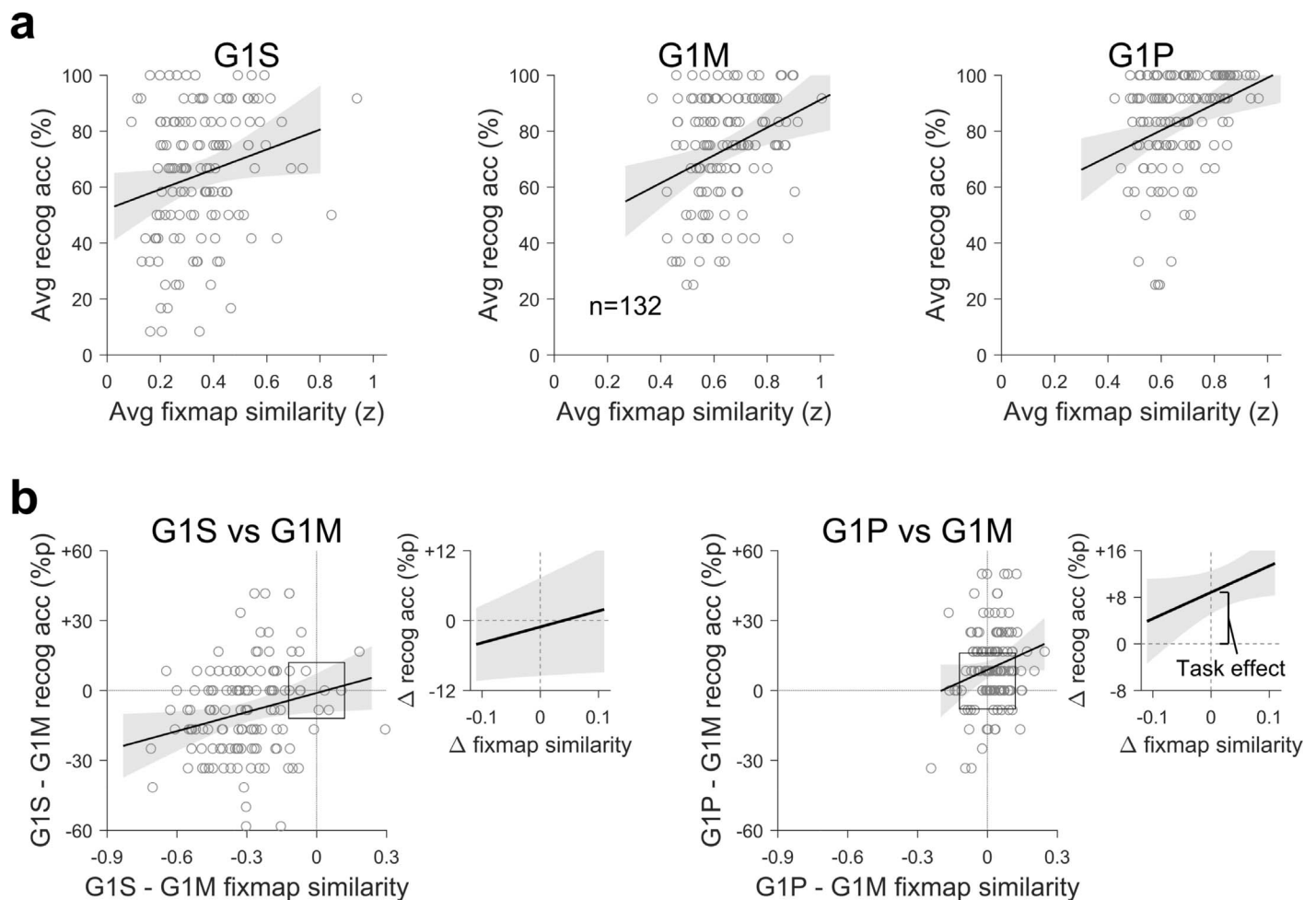
a

b

Figure 4. Results for the fixation map similarity. (a) Relationships between the fixation map similarity and recognition accuracy within each task. Each circle represents a scene. The solid lines are the significant regression lines (*p* = 0.01 for the G1S panel, and *ps* < 0.001 for the G1M and G1P panels), and the gray shades represent the 95% confidence bands. (b) Comparisons of the differences in fixation map similarity and recognition accuracy between tasks. The solid lines are the significant regression lines (*ps* = 0.006 and 0.02 for the G1S vs. G1M and G1P vs. G1M panels, respectively). The areas marked with a black square were zoomed and presented at the right side.

1.17], $p < 0.001$) and a nonsignificant effect of the search task ($\beta = -0.19$ [−0.55, 0.18], $p = 0.32$) while replicating significant effects of fixation count ($\beta = 0.72$ [0.54, 0.91], $p < 0.001$) and fixation map similarity ($\beta = 0.22$ [0.06, 0.39], $p = 0.008$).

These GLMM results can be illustrated by correlating the fixation map similarity and scene memory (Figure 4). The correlation values between the fixation map similarity and recognition accuracy were 0.23 [0.06, 0.38] ($p = 0.01$) for the search task, 0.32 [0.16, 0.47] ($p < 0.001$) for the memorization task, and 0.34 [0.18, 0.48] ($p < 0.001$) for the preference task (Figure 4a). Also, the memorability changes across tasks in each scene were significantly correlated with the differences in fixation map similarity across tasks (Figure 4b): Pearson's $r(130) = 0.24$ [0.07, 0.39] and 0.20 [0.03, 0.36], $ps = 0.006$ and 0.02, for the search and preference tasks, respectively. When fixation map similarity was controlled for across tasks, we still observed a significantly positive intercept in the preference task but not in the search task (the zoomed panels in Figure 4b). Together, these results show that the similarity between individual fixation maps during inspection and the template fixation map for encoding (i.e., the averaged G2M fixation map) could explain why the search task decreased scene memory but not why the preference task boosted scene memory.

Examination of the visual search task

During the visual search task, participants scanned for the cued objects in the scenes for 3.47 s on average (*SD* = 2.26 s across 1,584 trials) but viewed the scenes for 8 s total because those scenes remained on the screen even after participants' responses. So, partici-
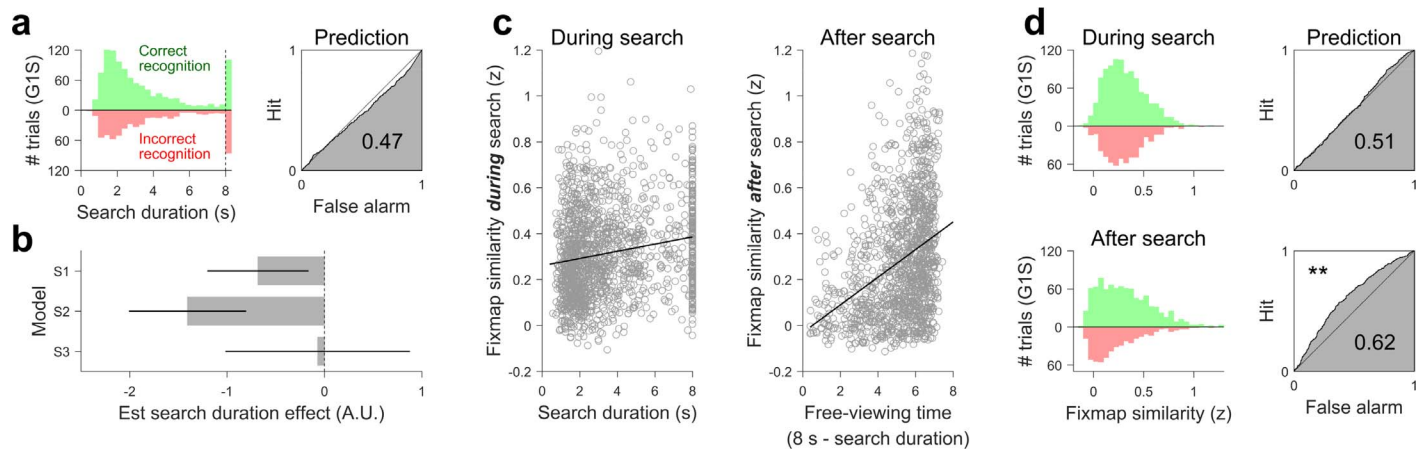
Figure 5. The effects of search duration and the fixation map similarity *during* and *after* search on scene memory in the visual search task (G1S). (a) The effect of search duration on scene memory. Left panel: Contrasts of the histograms of search duration of each G1S trial between the correct recognition trials (green) and the incorrect recognition trials (red). Dashed vertical line indicates the response window of 8 s, and the trials without response during 8 s were plotted on the right side of that line. Right panel: ROC curve. The horizontal and vertical axes specify the rate of false alarms (i.e., the incorrect trials that were wrongly classified as correct) and hits (i.e., the correct trials that were correctly classified), respectively. (b) Estimated effect (i.e., beta coefficients) of search duration on scene memory. Three different trial-level GLMM analyses were performed (Models S1–3; see Table 2 for the model description). The error bars indicate 95% confidence interval of the estimates. (c) Relationships between the search duration and fixation map similarity. Each circle indicates a trial. Fixation map similarity *during* (or *after*) search was calculated with the individual fixation maps that were generated using the fixations made *during* (or *after*) search. The *x* axis of the After search panel was reversed (i.e., 8 s: search duration) and named to Free-viewing time to indicate that there was no clear instruction to participants. The solid lines are the significant regression lines (*p*s < 0.001 for both). The 95% confidence bands were omitted for clarity. (d) The effects of fixation map similarity during and after search on scene memory. Left panels: Contrasts of the histograms of fixation map similarity between the correct recognition trials (green) and the incorrect recognition trials (red). Right panels: ROC curves. ** *p* < 0.001 (permutation test).

pants spent a large and variable portion of the time without a clear instruction during the search task, which might have contributed to the poor memory performance and the large variance in recognition accuracy in the search task, assuming that the amount of time spent on the encoding tasks determines the resulting scene memory. Therefore, we tested two hypotheses that are derived from this assumption.

The first hypothesis is that the short and variable search duration can explain the poor and variable scene memory from G1S. To test this, we conducted a trial-level GLMM analysis (Model A* in Table 1), in which the task engagement duration (i.e., the search duration in the search task and the scene presentation duration of 8 s for the memorization and preference tasks) was added to Model A as fixed effects. Model A* (*df* = 4745) explained 27.17% of the variance, and the full results are presented in Supplementary Table 5. However, Model A* showed significant effects of the search and preference tasks on scene memory (Figure 1c): βs = −0.80 [−1.19, −0.40] and 0.86 [0.49, 1.23], respectively, both *p*s < 0.001; and a nonsignificant effect of visual search duration (β = −0.42 [−0.86, 0.02], *p* = 0.06), rejecting the first hypothesis.

The second hypothesis is that the search duration is positively associated with scene memory within the

search task. We found that correct recognition trials (green in Figure 5a) were associated with significantly shorter visual search duration than incorrect (red) recognition trials: Search RT = 3.36 ± 2.17 s and 3.67 ± 2.41 s for correct and incorrect, respectively; ANOVA correct versus incorrect trials within G1S, $F(1, 1582) = 7.13$, $p = 0.008$. Although the ROC analysis based on the search duration (the right panel in Figure 5a) failed to predict whether a scene would be correctly recognized later at levels significantly above chance (AUC = 0.47 for G1S trials, $p = 0.07$ from permutation testing), a more sensitive GLMM analysis (Model S1 in Table 2; see Supplementary Table 6 for the full results) confirmed a significant negative effect of search duration on scene memory (Figure 5b): β = −0.68 [−1.20, −0.16], $p = 0.01$. Also, the effect of search duration remained significantly negative in Model S2 (β = −1.41 [−2.01, −0.80], $p < 0.001$; see Supplementary Table 7 for the full results), in which the fixation counts during 8 s were added to Model S1. Note that due to high correlation values between the search duration and fixation counts before and after search, Pearson's rs = 0.962 [0.958, 0.966] and −0.66 [−0.69, −0.63], respectively, both ps < 0.001, we opted to include only the overall fixation counts in Model S2, which were still significantly correlated with the search duration, $r =$

Model	Fixed effects	DF	Adj. R^2	BIC
S1	Scene orientation + search duration[a]	1581	34.19%	1883.3
S2	Scene orientation + search duration + fixation count[b]	1580	37.32%	1891.5
S3	Scene orientation + search duration + fixation count + fixation map similarity *during* search[c] + fixation map similarity *after* search[d]	1344	40.19%	1673.7
S4	Scene orientation + search duration + fixation count + fixation map similarity *during* search + fixation map similarity *after* search + scene orientation × encoding-recognition fixation map similarity[e]	1342	61.40%	1695.6

Table 2. Summary of trial-level GLMM analyses for the visual search task. *Notes*: Dependent variable: the correctness in the recognition task (1: correct, 0: incorrect). For Models S1 and S2, all 1,584 G1S trials were entered. For Model S3, the 1,350 G1S trials with at least one fixation in both during *and* after search were entered. All models included the by-participant random intercepts and slopes for all fixed effects and the by-scene random intercepts. For example, the random effect specification of Model S2 is (1 + Scene orientation + Search duration + Fixation count | Participant) + (1 | Scene) in the Wilkinson-Rogers notation. In fitting the models, the binomial distribution was specified for the dependent variable, and Laplace approximations were used to approximate the likelihood. All models converged to a stable solution. [a] The search duration was normalized to the range [0, 1] by dividing 8 s, i.e., the scene presentation duration. [b] The *z*-scored fixation counts during 8 s of viewing the scene were entered. [c] The *z*-scored similarity values between the template map and the individual fixation maps that were generated using the fixations made *during* search were entered. [d] The *z*-scored similarity values between the template map and the individual fixation maps that were generated using the fixations made *after* search were entered. Note that the number of trials entered into Model S3 is smaller the analysis was limited to the trials that have both the fixation map similarity values during and after search. [e] The *z*-scored similarity values between the encoding and recognition fixation maps of each individual (Figure 6a) were entered.

0.38 [0.34, 0.42], p < 0.001. This negative relationship between the search duration and scene memory rejects the second hypothesis and challenges the original assumption that the longer one is engaged in the encoding task, the better the encoding becomes, at least if the task is a visual search task.

To examine how participants were processing the scenes during and after search, we examined the fixation map similarity *during* and *after* search and their relationship to search duration. The fixation map similarity during (or after) search was calculated in the same manner (Figure 2d) but using the individual fixation maps that were generated from the fixations during (or after) search. Regarding the three different fixation map similarity values, we found that the similarity values during and after search were weakly correlated with each other (Pearson's r(1348) = 0.06 [0.00, 0.11], p = 0.04) and that the similarity values from the full 8 s data mainly reflected the similarity after search (Pearson's r = 0.83 [0.82, 0.85], p < 0.001). The correlation between the similarity from the 8 s and that during search was 0.47 [0.43, 0.51], p < 0.001. Regarding the relationships between the fixation map similarity and search duration (Figure 5c), we found that the fixation map similarity during and after search both increased with viewing time (Pearson's rs = 0.18 [0.13, 0.23] and 0.35 [0.30, 0.39], respectively, both ps < 0.001), which was either the search duration or free-viewing duration; but the association was stronger in the after-search, free-viewing condition: ANCOVA F(1, 2930) = 92.54, p < 0.001). Regarding the relationship between fixation map similarity and scene memory, the fixation map similarity after search, but not that during search, could predict whether a scene

would be correctly recognized later (Figure 5d): AUCs = 0.51 and 0.62, ps = 0.37 and < 0.001 from permutation testing, for during and after search, respectively. A GLMM analysis (Model S3 in Table 2; see Supplementary Table 8 for the full results) confirmed a significant positive effect of fixation map similarity after search on scene memory (β = 0.35 [0.16, 0.55], p < 0.001) and a nonsignificant effect of fixation map similarity during search (βs = −0.02 [−0.19, 0.14], p = 0.80). Moreover, the effect of search duration became nonsignificant in Model S3 (Figure 5b): βs = −0.07 [−1.01, 0.88], p = 0.89 after adding fixation map similarity during and after search as fixed effects. Together, these results suggest that successful scene encoding happened not during the visual search task but rather after visual search when participants began fixating on memory-relevant regions, providing an explanation to the puzzling negative relationship between search duration and ensuing scene memory.

Comparison of the fixation maps during encoding and retrieval

Previous memory research suggests that the degree of match between encoding and retrieval conditions affects memory performance (Roediger & Guynn, 1996; Rugg, Johnson, Park, & Uncapher, 2008). For example, color improved the recognition of natural scenes only when the colored scenes were used at both encoding and retrieval (Spence, Wong, Rusan, & Rastegar, 2006). In addition, the match between encoding and retrieval was shown to have a larger
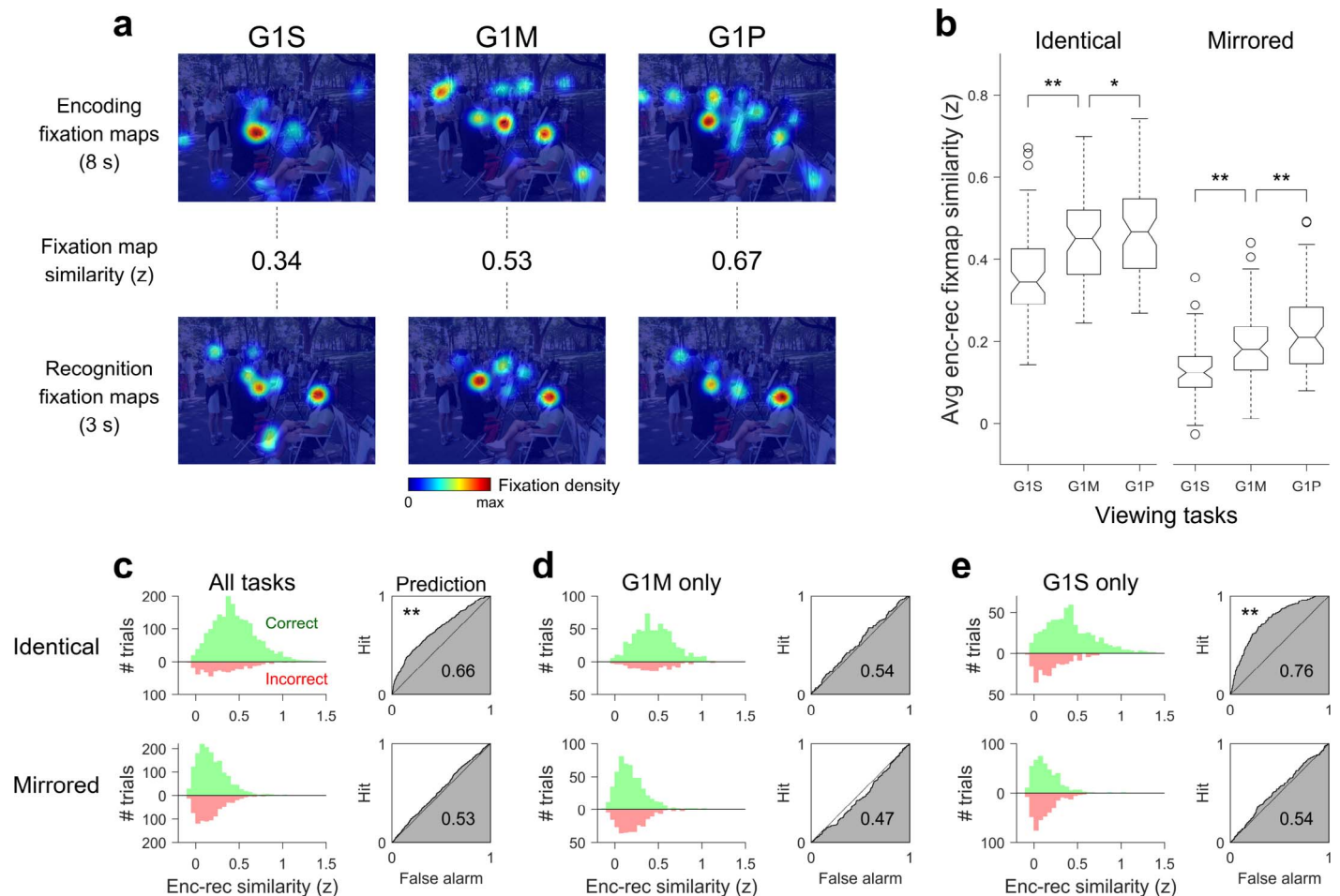
Figure 6. The effects of encoding-recognition fixation map similarity on scene memory. (a) The encoding-recognition fixation map similarity analysis. Top row, the same individual fixation maps during encoding as in Figure 2d. Bottom row, the individual fixation maps during the recognition test from the same individuals. The *encoding-recognition fixation map similarity* is the Fisher *z*-transformed correlation coefficient between the encoding fixation maps (top) and recognition fixation maps (bottom), which are from the same scene. (b) The effects of viewing tasks and scene orientation (i.e., identical or horizontally-mirrored) on encoding-recognition fixation map similarity. The similarity values for each scene were averaged across participants within task. Circles indicate the scenes with outlier values. * $p < 0.01$, ** $p < 0.001$ (paired *t* test). (c) The relationship between encoding-recognition fixation map similarity and scene memory in the all (i.e., G1S, G1M, and G1P) trials. Top row: The trials in which the identical scenes were used in the recognition test. Bottom row: The trials in which the horizontally mirrored scenes were used in the recognition test. Left panels: Contrasts of the histograms of encoding-recognition fixation map similarity of each trial between the correct recognition trials (green) and the incorrect recognition trials (red). Right panels: ROC curves. ** $p < 0.001$ (permutation test). (d) The relationship between encoding-recognition fixation map similarity and scene memory in the G1M trials. (e) The relationship between encoding-recognition fixation map similarity on scene memory in the G1S trials.

influence on prospective memory performance than characteristics of encoding and retrieval, such as encoding instructions and retrieval targets (Hannon & Daneman, 2007). Therefore, we examined whether fixation map similarity between encoding and recognition (Figure 6a) was associated with scene memory, hypothesizing that the higher the encoding-recognition fixation map similarity, the greater the likelihood of successful recognition. We also sought to determine whether encoding-recognition fixation map similarity could explain the observed task effects. In doing so, we

separately analyzed the trials in which the identical scenes were presented in the recognition test and those in which the horizontally mirrored scenes were presented (see Methods) because scene orientation affects the eye movement patterns.

Encoding-recognition fixation map similarity for the identical scenes was 0.36 ± 0.11 ($M \pm SD$ across 66 scenes), 0.45 ± 0.10, and 0.47 ± 0.12 for the search, memorization, and preference tasks, respectively (Figure 6b). As expected, encoding-recognition fixation map similarity for the horizontally-mirrored scenes

(0.13 ± 0.07, 0.19 ± 0.09, and 0.22 ± 0.10, for the search, memorization, and preference tasks, respectively), was lower when the horizontally mirrored scenes (indicated with "Mirrored" in Figure 6b through e) were presented in the recognition test. When comparing between the tasks, the encoding-recognition similarity was lower in the search task than in the memorization task (Figure 6b): paired $t(65) = 5.39$ and 5.54, both $ps < 0.001$, Hedges' $g = 0.79$ [0.47, 1.14] and 0.73 [0.47, 1.02], for the identical and mirrored scenes, respectively, compared on a scene-by-scene basis, and higher in the preference task: paired $t(65) = 2.69$ and 4.80, $p = 0.009$ and < 0.001, Hedges' $g = 0.25$ [0.08, 0.44] and 0.38 [0.23, 0.38].

Next, we asked whether encoding-recognition fixation map similarity could explain the variance in recognition accuracy across all trials regardless of the tasks. We found that correct recognition trials were associated with significantly greater encoding-recognition fixation map similarity than incorrect recognition trials (left panels in Figure 6c): ANOVA correct (green) versus incorrect (red) trials, $F(1, 2374) = 106.37$ and $p < 0.001$ for the identical trials, $F(1, 2374) = 8.15$ and $p = 0.004$ for the mirrored trials. Also, ROC analyses confirmed that whether a scene would be correctly recognized later could be predicted using encoding-recognition fixation map similarity at levels significantly above chance in the identical trials (right panels in Figure 6c): Area under curve (AUC) = 0.66 and $p < 0.001$ for the identical trials; and AUC = 0.53 and $p = 0.06$ for the mirrored trials, seemingly supporting the importance of encoding-retrieval similarity in explaining subsequent scene memory.

We then asked whether encoding-recognition fixation map similarity could explain the variance in recognition accuracy within each viewing task, hypothesizing that it would do so if it determines scene memory. In the memorization and preference tasks, the encoding-recognition fixation map similarity and scene memory were not significantly associated; encoding-recognition similarity was not significantly different between correct and incorrect trials in the memorization task (left panels in Figure 6d): ANOVA correct versus incorrect trials, $F(1,790) = 2.71$ and 1.16, $ps = 0.10$ and 0.28, for the G1M identical and mirrored trials, respectively; and the preference task, $F(1, 790) = 0.13$ and 1.21, $ps = 0.72$ and 0.27, for the G1P identical and mirrored trials, respectively. Accordingly, encoding-recognition similarity failed to predict whether a scene would be correctly recognized later in the memorization task (right panels in Figure 6d; AUC = 0.54 and 0.47, $ps = 0.08$ and 0.95, for the G1M identical and mirrored trials, respectively) and in the preference task (AUC = 0.50 and 0.52, $ps = 0.46$ and 0.23, for the G1P identical and mirrored trials, respectively). In addition, we failed to find a significant relationship

between encoding-recognition similarity and scene memory in the G1S mirrored trials (bottom panels in Figure 6e): ANOVA correct versus incorrect trials, $F(1, 790) = 2.74$, $p = 0.10$; AUC = 0.54, one-tailed, uncorrected $p = 0.04$. However, within the G1S identical trials, encoding-recognition similarity and scene memory were significantly associated (top panels in Figure 6e): ANOVA correct versus incorrect trials, $F(1,790) = 117.9$, $p < 0.001$; AUC = 0.76, $p < 0.001$. Moreover, adding encoding-recognition similarity and its interaction with scene orientation to Model S3 (Model S4 in Table 2; Supplementary Table 9 for the full results) could explain 20% of additional variance in scene memory from G1S. These results suggest that encoding-retrieval similarity could only explain significant variance in the visual search task and could not explain the memory boost exhibited by the preference task.

To examine the extent to which encoding-recognition fixation map similarity could explain the task effects on scene memory, we conducted a trial-level GLMM analysis (Model M in Table 1), in which encoding-recognition similarity and its interactions with task type and scene orientation were added to Model A as fixed effects. Model M ($df = 4740$) explained 33.74% of the variance, and the full results are presented in Supplementary Table 10. Model M showed a nonsignificant main effect of encoding-recognition similarity ($\beta = 0.12$ [−0.11, 0.36], $p = 0.30$), consistent with Figure 6d; and significant interactions between the search task and encoding-recognition similarity ($\beta = 1.06$ [0.71, 1.40], $p < 0.001$); and between the search task, encoding-recognition similarity, and scene orientation ($\beta = -0.67$ [−1.18, −0.15], $p = 0.01$), consistent with Figure 6e, which showed a significant relationship within the G1S identical trials but not within the G1S mirrored trials. However, the effects of search and preference tasks on scene memory both remained significant in Model M (Figure 1c): $\beta s = -0.41$ [−0.78, −0.05] and 0.91 [0.49, 1.34], $p = 0.03$ and < 0.001, respectively, suggesting that encoding-recognition fixation map similarity could not explain the memorability changes across tasks.

Examination of scene aesthetic preference

In searching for factors contributing to the memorability boost by the aesthetic preference task, we examined the influence of scene aesthetics, which has previously been found to be significantly (although negatively) associated with scene memorability (Isola et al., 2014). To do so, we collected scene preference ratings in a separate online study (see Methods) from a larger pool of participants, G3. As we aggregated individual raters' preferences to estimate the overall

Predictor	β coefficient	SE(β)	t(4737)	p
Intercept (G1M)	1.83 [1.44, 2.21]	0.19	9.37	< 0.001
Task type: G1S	−0.19 [−0.56, 0.18]	0.19	−1.02	0.31
Task type: G1P	0.81 [0.43, 1.19]	0.19	4.16	< 0.001
Scene orientation	−1.10 [−1.48, −0.71]	0.20	−5.58	< 0.001
Fixation count	0.72 [0.53, 0.91]	0.10	7.52	< 0.001
Fixation map similarity	0.22 [0.06, 0.39]	0.08	2.63	0.008
Scene preference	0.13 [−0.06, 0.32]	0.10	1.32	0.19
G1S × scene orientation	0.10 [−0.26, 0.45]	0.18	0.54	0.59
G1P × scene orientation	−0.28 [−0.69, 0.14]	0.21	−1.31	0.19
G1S × fixation count	−0.53 [−0.75, −0.31]	0.11	−4.73	< 0.001
G1P × fixation count	−0.49 [−0.74, −0.25]	0.13	−3.91	< 0.001
G1S × fixation map similarity	0.14 [−0.10, 0.38]	0.12	1.16	0.25
G1P × fixation map similarity	0.09 [−0.16, 0.34]	0.13	0.67	0.50
G1S × scene preference	0.15 [−0.03, 0.32]	0.09	1.64	0.10
G1P × scene preference	0.13 [−0.06, 0.33]	0.10	1.34	0.18

Table 3. Results of GLMM analysis: Model F (the full model). *Notes*: Dependent variable: the correctness of all 4,752 trials (1: correct recognition, 0: incorrect recognition). The by-participant random intercepts and slopes for the viewing tasks and the by-scene random intercepts, i.e., (1 + Viewing Task | Participant) + (1 | Scene) in the Wilkinson-Rogers notation, were included in the model. In fitting the models, the binomial distribution was specified for the dependent variable, and Laplace approximations were used to approximate the likelihood. *df* = 4737. R² (adjusted) = 31.18%. 95% CIs are stated beside the βs.

preference of each scene, we checked interrater reliability using Shrout and Fleiss's Case 2 (i.e., a two-way random effects model) intraclass correlation formula for average measures (Shrout & Fleiss, 1979), in which scene and rater are both modeled as random effects. The interrater reliability was 0.96 [0.94, 0.97], *p* < 0.001, which falls in the conventionally "excellent" range (Cicchetti, 1994). The mean aesthetic preference rating of the scenes was 2.57 (*SD* = 0.59; note that the scene preference was rated on 4-point scale, 1 = *dislike*, and 4 = *like*).

To examine the relationship between scene memory and preference ratings, we conducted a GLMM analysis (Model E in Table 1), in which scene preference and its interactions with task type were added to Model A as fixed effects. Model E (*df* = 4743) explained 27.34% of the variance, and the full results are presented in Supplementary Table 11. Model E showed a nonsignificant effect of scene preference on scene memory (β = 0.11 [−0.09, 0.30], *p* = 0.28) and a significant interaction between the search task and scene preference (β = 0.20 [0.03, 0.37], *p* = 0.02), suggesting that higher preferred scenes were more remembered in the search task relative to the memorization task. However, the effects of search and preference tasks on scene memory both remained significant in Model E (Figure 1c): βs = −0.56 [−0.88, −0.24] and 0.88 [0.51, 1.25], respectively; both *ps* < 0.001), suggesting that the scene preference ratings could not explain the memorability changes across tasks.

Finally, the full model (Model F, *df* = 4737, R² = 31.18%; Table 3), in which fixation counts, fixation map similarity, and scene preference, and their interactions with task type were added to Model A as fixed effects, summarizes our findings in a comprehensive manner (see Supplementary Table 12 for the random effects covariance parameters). Model F confirmed significant positive effects of fixation count (β = 0.72 [0.53, 0.91], *p* < 0.001) and fixation map similarity (β = 0.22 [0.06, 0.39], *p* = 0.008) on scene memory and a nonsignificant effect of scene preference rating (β = 0.13 [−0.06, 0.32], *p* = 0.19). Importantly, Model F confirmed a nonsignificant effect of the search task (β = −0.19 [−0.56, 0.18], *p* = 0.31; Figure 1c) and a significant positive effect of the preference task (β = 0.81 [0.43, 1.19], *p* < 0.001).

Discussion

Here, we used eye-tracking and trial-level GLMM analyses to systematically compare how performing different viewing tasks on scenes affects their subsequent memorability. Incidental scene memory decreased after visual search and increased after aesthetic preference evaluation compared to intentional memorization (Figure 1). The three viewing tasks also resulted in different eye-movement patterns and different deployments of overt visual attention. Relating fixation counts to scene memory, we found that fixation counts were significantly associated with overall scene memory (Figure 2c), consistent with previous research (Bloom & Mudd, 1991; Loftus, 1972; Pertzov et al., 2009; Tatler & Tatler, 2013), but could not fully explain

the changes in scene memorability due to tasks (Figure 3). Relating attention allocation to scene memory, we found that fixation map similarity was significantly associated with overall scene memory (Figure 2f) and, importantly, explained why visual search decreases scene memory (Figure 4). However, it was not significantly associated with the memorability boost in aesthetic preference evaluation. Additionally, we found that the search durations (Figure 5), encoding-recognition fixation map similarity (Figure 6), and scene preference ratings, respectively, could not explain the task effects on scene memorability. Together, these results demonstrate the importance of examining where individuals fixate in a scene and not just how many fixations individuals make.

We investigated if the locations where viewers fixated during encoding influenced scene memory by analyzing fixation maps (Figure 2d through f), which provided rich spatial information. When intentionally memorizing a scene, participants allocated overt attention in a consistent manner, as evidenced by the high level of similarity (0.65 ± 0.25, mean ± SD across trials) between the individual G1M fixation maps and the template fixation map for encoding (i.e., the averaged G2M fixation map). Ocular targeting is guided by both bottom-up information and top-down goals (Ballard & Hayhoe, 2009; Henderson, 2007, 2011, 2013), so regions that are highly fixated upon are likely memory-relevant and support scene encoding. We consistently found significant positive associations between the deployment of visual attention and scene memory (Figure 2f), suggesting that the more a viewer's attention is diverted away from the memory-relevant regions, the less effective the encoding (both intentional and incidental) of those scenes becomes. Importantly, fixation map similarity was able to explain the decrease in scene memorability in the search task (Models C, D, and F in Figure 1c), in which participants attended more to the search-relevant-and-memory-irrelevant regions (e.g., shoes in the G1S panel in Figure 2a). By providing a mechanistic link between fixation maps during encoding and subsequent scene memory, our results extend research for how visual attention aids scene memory (Castelhano & Henderson, 2005; Draschkow et al., 2014; Hollingworth, 2012; Hollingworth & Henderson, 2002; Olejarczyk et al., 2014; Tatler & Tatler, 2013; Williams et al., 2005; Wolfe et al., 2007).

In the visual search task, participants spent a large and variable portion of the free-viewing time looking at the scenes without a clear instruction, as those scenes remained on the screen after their responses (i.e., after the search object had been found). Although not intended, this provided an interesting opportunity to examine the relationship between search duration, i.e., the time spent on the encoding task, and resulting scene

memory. We found a significantly negative effect of search duration (or a significantly positive effect of free-viewing duration; unfortunately we could dissociate one from the other in the current study) on scene memory (Figure 5b). To dig deeper into this result, we investigated how participants were processing the scenes during and after search by examining fixation map similarity during and after search and their relationships to scene memory. Interestingly, we found that fixation map similarity during search failed to predict whether a scene would be correctly recognized later (Figure 5d). Based on previous research suggesting that viewing tasks may affect the extraction (Võ & Wolfe, 2012) and/or the retention (Maxcey-Richard & Hollingworth, 2013) of visual information and that top-down goals may have prioritized task-relevant information over task-irrelevant information (Draschkow & Võ, 2016; Tatler et al., 2013; Tatler & Tatler, 2013), we speculate that the visual search task prioritized search-related operations during search, which could increase memory for those specific objects, but doing so reduces the extraction and retention of overall scene visual information that is irrelevant to search but critical for incidental encoding of the scene in totality. When the visual search task is completed (i.e., the search object is found), however, the priority of encoding-related operations is normalized and incidental scene encoding resumes during free-viewing (i.e., the viewing time remaining after the search object has been found). As such, the fixation map similarity after search increased with free-viewing time (Figure 5c) and was positively associated with incidental scene memory. Moreover, fixation map similarity during and after search fully explained the negative effect of search duration on scene memory, strengthening the relationship between fixation map similarity and scene memory. These results suggest that our search task condition was actually a combination of the search task and free-viewing, and that scene memory was incidentally formed while participants were free-viewing the scenes without a clear instruction, providing a cautionary remark on designing incidental encoding tasks. Also, the precise relationship between the time spent on visual search and resulting incidental scene memory should be examined in a new study that is not confounded by free-viewing after the search object was found.

Inspired by the encoding-retrieval match principle (Roediger & Guynn, 1996; Rugg et al., 2008), we investigated if fixation map similarity between encoding and recognition influenced scene memory. First, we found that greater encoding-recognition fixation map similarity in the search task was related to better scene memory (Figure 6e). We suggest that encoding-recognition fixation map similarity in the search task reflects whether the searched scenes

successfully reinstated the context of visual search during the memory test because, if successful, participants will also remember the searched objects and look at them during recognition. In this context, the searched objects are highly informative retrieval cues, which can easily distinguish the searched scenes from memory (Nairne, 2002) and overcome weak encoding from the search task and/or a limited capacity retrieval process (Beck, Peterson, & Angelone, 2007; Beck & van Lamsweerde, 2011). Therefore, if encoding-recognition similarity is high for a searched scene, its recognition will be more likely to be successful. Also, an additional 20% of variance in scene memory from G1S was explained by including the effect of encoding-recognition similarity (Model S4), suggesting that whether the search context was successfully reinstated substantially contributed to incidental scene memory from visual search. At least in visual search, our eye-tracking results dovetail with recent functional neuroimaging studies (Danker, Tompary, & Davachi, 2017; Ritchey, Wing, LaBar, & Cabeza, 2013; Wing, Ritchey, & Cabeza, 2015), which suggest that successful reinstatement of encoding-related neural activity can help successful retrieval. However, the relationships between encoding-recognition similarity and scene memory were nonsignificant in the memorization and preference tasks (see Figure 6d) although the overall encoding-recognition similarity was higher in those tasks than in the search task. This is consistent with recent research suggesting that the absolute level of encoding-retrieval match alone cannot predict memory performance and that the diagnostic value of retrieval cues is important (Goh & Lu, 2012; Nairne, 2002; Poirier et al., 2012). Our results could be explained if the retrieval cues (i.e., fixated regions during the memory test) in the memorization and preference task conditions were salient regions, which are more likely to attract fixations regardless, thus resulting in the high level of encoding-recognition similarity, but those cues were less helpful in recognizing the specific scenes that were encoded during the tasks. Moreover, we found that encoding-recognition similarity could not explain the observed task effects in scene memory (Model M), suggesting that the effects of viewing tasks are mainly limited to encoding.

When evaluating scene preference, participants had better memory for those scenes, even though they were not explicitly asked to memorize them. Evaluating the pleasantness or likability of a face has been shown to enhance incidental memory compared to intentional memorization (Bernstein et al., 2002; Grady et al., 2002; Smith & Winograd, 1978; Warrington & Ackroyd, 1975), but few studies compared the effect of preference evaluation to intentional memorization of natural scenes. One such

study (Grady et al., 1998) compared intentional picture memory to incidental memory from making semantic (e.g. living/nonliving) or nonsemantic (e.g., size of picture) judgments on pictures and reported that intentional memorization produced the greatest memory. Importantly, however, they did not test incidental memory in a preference judgment task. Another study (Wolfe et al., 2007), which reported poor incidental scene memory after visual search, also used a preference judgment task (i.e., rating the likeability) as a viewing task for scenes and textures and reported a slight increase in resulting memory after preference judgments compared to intentional memorization. Wolfe et al. (2007) used an old/new recognition task, and their participants showed the high recognition performance of d′ around 3 in both conditions, perhaps limiting the memory boosting effect of preference evaluation. Our study establishes the memory enhancing effect of preference judgments on natural scenes, extending the incidental memory boost to a new stimulus, while simultaneously demonstrating memory decrements for another form of incidental encoding, i.e., visual search.

It is important to try to understand how preference evaluation increased incidental scene memory. The memory boost due to the preference task could not be explained by fixation counts or the spatial manifestation of visual attention (the preference task effect panel in Figure 1c), suggesting that memory-facilitating processes operated at an implicit level during preference evaluation. Based on previous research, we speculate three possible implicit mechanisms: (a) Evaluating scene preference may have involved more scene semantic processing, which can facilitate incidental scene encoding by improving the consolidation of bottom-up visual information (Draschkow et al., 2014; Josephs et al., 2016; Võ & Wolfe, 2015); (b) evaluating preference may have increased self-related processing, resulting in more semantic elaboration of the scenes and helping to organize of the information being processed (Rogers et al., 1977; Symons & Johnson, 1997); and lastly, (c) preference judgments may have activated the brain's reward circuitry and thus enhanced memory (Biederman & Vessel, 2006; Yue, Vessel, & Biederman, 2007), but this explanation is complicated given that scene preference ratings could not explain the memory boost from the preference task. However, the very act of making a preference rating, independent of the actual preference rating generated, could boost memory. In line with this idea is a study that showed that making a voluntary choice (independent of any rating) enhances memory via the striatum (Murty et al., 2015), a region involved in decision making and subjective evaluation (Bartra, McGuire, & Kable, 2013), and thus could enhance memory just

by making an evaluation of a stimulus. Regardless, further research to test these possible mechanisms in order to better understand how scene memory was enhanced by the act of preference evaluation could lead to many interesting applications in psychology, education, and marketing research.

Conclusions

In summary, our results show that the deployment of overt attention and attentional elaboration can explain why some incidental encoding tasks, such as visual search, will worsen memory compared to intentional memorization. In contrast, scene preference yielded enhanced memory above intentional encoding, while preserving very similar deployments of overt attention. As such, the incidental-encoding boost for preference likely draws upon more implicit encoding mechanisms—of which, we highlighted at least three possibilities. Together, these results suggest that factors affecting incidental memory could be different for different incidental encoding tasks and that implicit and explicit mechanisms of incidental memory should be differentiated.

Keywords: visual memory, eye movements, incidental memory, visual search, aesthetic preference

Acknowledgments

Commercial relationships: none.
Corresponding author: Marc G. Berman.
Email: bermanm@uchicago.edu.
Address: Department of Psychology, The University of Chicago, Chicago, IL, USA.

References

Anderson, J. R., & Lynne, M. R. (1979). An elaborative processing explanation of depth of processing. In S. Cermak & F. Craik (Eds.), *Levels of processing in human memory* (pp. 385–404). Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, *85*(3), 139–152.

Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition*, *17*(6–7), 1185–1204.

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.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–427.

Beck, M. R., Levin, D. T., & Angelone, B. L. (2007). Change blindness blindness: Beliefs about the roles of intention and scene complexity in change detection. *Consciousness and Cognition*, *16*(1), 31–51.

Beck, M. R., Peterson, M. S., & Angelone, B. A. (2007). The roles of encoding, retrieval, and awareness in change detection. *Memory and Cognition*, *35*(4), 610–620.

Beck, M. R., & van Lamsweerde, A. E. (2011). Accessing long-term memory representations during visual change detection. *Memory and Cognition*, *39*(3), 433–446.

Bernstein, L. J., Beig, S., Siegenthaler, A. L., & Grady, C. L. (2002). The effect of encoding strategy on the neural correlates of memory for faces. *Neuropsychologia*, *40*(1), 86–98.

Biederman, I., & Vessel, E. A. (2006). Perceptual pleasure and the brain. *American Scientist*, *94*(3), 247–253.

Bloom, L. C., & Mudd, S. A. (1991). Depth of processing approach to face recognition: A test of two theories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 556–565.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135.

Castelhano, M. S., & Henderson, J. M. (2005).

Incidental visual memory for objects in scenes. *Visual Cognition*, *12*(6), 1017–1040.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*(3):6, 1–15, doi:10.1167/9.3.6. [PubMed] [Article]

Choe, K. W., Blake, R., & Lee, S.-H. (2016). Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation. *Vision Research*, *118*, 48–59.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290.

Coin, C., & Tiberghien, G. (1997). Encoding activity and face recognition. *Memory*, *5*(5), 545–568.

Craik, F. I. M. (2002). Levels of processing: Past, present, and future? *Memory*, *10*(5–6), 305–318.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Danker, J. F., Tompary, A., & Davachi, L. (2017). Trial-by-trial hippocampal encoding activation predicts the fidelity of cortical reinstatement during subsequent retrieval. *Cerebral Cortex*, *27*(7), 3515–3524. doi.org/10.1093/cercor/bhw146.

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.

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, 1–18, doi:10.1167/14.8.10. [PubMed] [Article]

Drewes, J., Zhu, W., Hu, Y., & Hu, X. (2014). Smaller is better: Drift in gaze measurements due to pupil dynamics. *PloS One*, *9*(10), e111197.

Einhäuser, W., & Nuthmann, A. (2016). Salient in space, salient in time: Fixation probability predicts fixation duration during natural scene viewing. *Journal of Vision*, *16*(11):13, 1–17, doi:10.1167/16.11.13. [PubMed] [Article]

Goh, W. D., & Lu, S. H. X. (2012). Testing the myth of the encoding-retrieval match. *Memory & Cognition*, *40*(1), 28–39.

Grady, C. L., Bernstein, L. J., Beig, S., & Siegenthaler, A. L. (2002). The effects of encoding task on age-related differences in the functional neuroanatomy of face memory. *Psychology and Aging*, *17*(1), 7–23.

Grady, C. L., McIntosh, A. R., Rajah, M. N., & Craik, F. I. (1998). Neural correlates of the episodic encoding of pictures and words. *Proceedings of the National Academy of Sciences, USA*, *95*(5), 2703–2708.

Hannon, B., & Daneman, M. (2007). Prospective memory: The relative effects of encoding, retrieval, and the match between encoding and retrieval. *Memory*, *15*(5), 572–604.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, *16*(4), 219–222.

Henderson, J. M. (2011). Eye movements and scene perception. In S. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford handbook of eye movements* (pp. 593–606). Oxford, UK: Oxford University Press.

Henderson, J. M. (2013). Eye movements. In D. Reisberg (Ed.), *The oxford handbook of cognitive psychology* (pp. 69–82). New York, NY: Oxford University Press.

Hentschke, H., & Stüttgen, M. C. (2011). Computation of measures of effect size for neuroscience data sets. *The European Journal of Neuroscience*, *34*(12), 1887–1894.

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.

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.

Ibarra, F. F., Kardan, O., Hunter, M. R., Kotabe, H. P., Meyer, F. A. C., & Berman, M. G. (2017). Image feature types and their predictions of aesthetic preference and naturalness. *Frontiers in Psychology*, *8*, 632.

Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(7), 1469–1482.

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.

Kardan, O., Berman, M. G., Yourganov, G., Schmidt, J., & Henderson, J. M. (2015). Classifying mental states from eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1502–1514.

Kardan, O., Henderson, J. M., Yourganov, G., & Berman, M. G. (2016). Observers' cognitive states modulate how visual inputs relate to gaze control. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1429–1442.

Kolers, P. A. (1973). Remembering operations. *Memory & Cognition*, *1*(3), 347–355.

Kotabe, H. P., Kardan, O., & Berman, M. G. (2016). The order of disorder: Deconstructing visual disorder and its effect on rule-breaking. *Journal of Experimental Psychology: General*, *145*(12), 1713–1727.

Kotabe, H. P., Kardan, O., & Berman, M. G. (2017). The nature-disorder paradox: A perceptual study on how nature is disorderly yet aesthetically preferred. *Journal of Experimental Psychology: General*, *146*(8), 1126–1142. doi:10.1037/xge0000321.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442.

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology*, *3*(4), 525–551.

Luke, S. G., Smith, T. J., Schmidt, J., & Henderson, J. M. (2014). Dissociating temporal inhibition of return and saccadic momentum across multiple eye-movement tasks. *Journal of Vision*, *14*(14):9, 1–12, doi:10.1167/14.14.9. [PubMed] [Article]

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

Maxcey-Richard, A. M., & Hollingworth, A. (2013). The strategic retention of task-relevant objects in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 760–772.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533.

Murty, V. P., DuBrow, S., & Davachi, L. (2015). The simple act of choosing influences declarative memory. *The Journal of Neuroscience*, *35*(16), 6255–6264.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*(5–6), 389–395.

Nuthmann, A. (2017). Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic Bulletin & Review*, *24*(2), 370–392.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8):20, 1–19, doi:10.1167/10.8.20. [PubMed] [Article]

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.

Pajak, M., & Nuthmann, A. (2013). Object-based saccadic selection during scene perception: evidence from viewing position effects. *Journal of Vision*, *13*(5):2, 1–21, doi:10.1167/13.5.2. [PubMed] [Article]

Pertzov, Y., Avidan, G., & Zohary, E. (2009). Accumulation of visual information across multiple fixations. *Journal of Vision*, *9*(10):2, 1–12, doi:10.1167/9.10.2. [PubMed] [Article]

Poirier, M., Nairne, J. S., Morin, C., Zimmermann, F. G. S., Koutmeridou, K., & Fowler, J. (2012). Memory as discrimination: A challenge to the encoding–retrieval match principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 16–29.

Pomplun, M., Ritter, H., & Velichkovsky, B. (1996). Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception*, *25*(8), 931–948.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Ritchey, M., Wing, E. A., LaBar, K. S., & Cabeza, R. (2013). Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cerebral Cortex*, *23*(12), 2818–2828.

Roediger, H. L., & Guynn, M. J. (1996). Retrieval processes. In E. L. Bjork & R. A. Bjork (Eds.), *Human memory* (pp. 197–236). San Diego, CA: Academic Press.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*(9), 677–688.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you

look. *Journal of Vision*, *7*(14):16, 1–20, doi:10.1167/7.14.16. [PubMed] [Article]

Rugg, M. D., Johnson, J. D., Park, H., & Uncapher, M. R. (2008). Encoding-retrieval overlap in human episodic memory: A functional neuroimaging perspective. *Progress in Brain Research*, *169*, 339–352.

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, *1*(2), 103–113.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Smith, A. D., & Winograd, E. (1978). Adult age differences in remembering faces. *Developmental Psychology*, *14*(4), 443–444.

Spence, I., Wong, P., Rusan, M., & Rastegar, N. (2006). How color enhances visual memory for natural scenes. *Psychological Science*, *17*(1), 1–6.

Swets, J. A. (1973, Dec. 7). The relative operating characteristic in psychology: A technique for isolating effects of response bias finds wide use in the study of perception and cognition. *Science*, *182*(4116), 990–1000.

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*(3), 371–394.

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.

Tatler, B. W., & Tatler, S. L. (2013). The influence of instructions on object memory in a real-world setting. *Journal of Vision*, *13*(2):5, 1–20, doi:10.1167/13.2.5. [PubMed] [Article]

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, *3*(1):9, 86–94, doi:10.1167/3.1.9. [PubMed] [Article]

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.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373.

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*, 72–81.

Warrington, E. K., & Ackroyd, C. (1975). The effect of orienting tasks on recognition memory. *Memory & Cognition*, *3*(2), 140–142.

Wilkinson, G. N., & Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics*, *22*(3), 392–399.

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.

Wing, E. A., Ritchey, M., & Cabeza, R. (2015). Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. *Journal of Cognitive Neuroscience*, *27*(4), 679–691.

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(3), 181–190.

Wolfe, J. M., Horowitz, T. S., & Michod, K. O. (2007). Is visual attention required for robust picture memory? *Vision Research*, *47*(7), 955–964.

Wooding, D. S. (2002). Eye movements of large populations: II. Deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods*, *34*(4), 518–528.

Wyatt, H. J. (2010). The human pupil and the use of video-based eyetrackers. *Vision Research*, *50*(19), 1982–1988.

Yue, X., Vessel, E. A., & Biederman, I. (2007). The neural basis of scene preferences. *Neuroreport*, *18*(6), 525–529.