The extent and specificity of visual exploration determines the formation of recollected memories in complex scenes

Nico Broers

Institute of Psychology, University of Münster, Münster, Germany
Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany ✉

Wilma A. Bainbridge

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

René Michel

Institute of Psychology, University of Münster, Münster, Germany
Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany ✉

Elio Balestrieri

Institute of Psychology, University of Münster, Münster, Germany
Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany ✉

Niko A. Busch

Institute of Psychology, University of Münster, Münster, Germany
Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany ✉

Our visual memories of complex scenes often appear as robust, detailed records of the past. Several studies have demonstrated that active exploration with eye movements improves recognition memory for scenes, but it is unclear whether this improvement is due to stronger feelings of familiarity or more detailed recollection. We related the extent and specificity of fixation patterns at encoding and retrieval to different recognition decisions in an incidental memory paradigm. After incidental encoding of 240 real-world scene photographs, participants (*N* = 44) answered a surprise memory test by reporting whether an image was new, remembered (indicating recollection), or just known to be old (indicating familiarity). To assess the specificity of their visual memories, we devised a novel report procedure in which participants selected the scene region that they specifically recollected, that appeared most familiar, or that was particularly new to them. At encoding, when considering the entire scene, subsequently recollected compared to familiar or forgotten scenes showed a larger number of fixations that were more broadly distributed, suggesting that more extensive visual exploration determines stronger and more detailed memories. However, when considering only the memory-relevant image areas, fixations were more dense and more clustered for subsequently recollected compared to subsequently familiar scenes. At retrieval, the extent of visual exploration was more restricted for recollected compared to new or forgotten scenes, with a smaller number of fixations. Importantly, fixation density and clustering was greater in memory-relevant areas for recollected versus familiar or falsely recognized images. Our findings suggest that more extensive visual exploration across the entire scene, with a subset of more focal and dense fixations in specific image areas, leads to increased potential for recollecting specific image aspects.

Introduction

Visual memories can often appear as detailed and coherent records of our pasts. To form these memories we must visually explore our environments, because the size of the retina restricts high spatial resolution and color vision to an area of only 2° around fixation. In this sense, the details of visual memories must be based on specific viewing behavior: scene locations across the entire scene must be explored by successive fixations, which then must be integrated to form coherent memory representations. Indeed, eye-movement patterns are partially reinstated at memory retrieval to support extraction of originally encoded feature traces (Noton & Stark, 1971; Wynn, Olsen, Binns, Buchsbaum, & Ryan, 2018; Kragel & Voss, 2021). Moreover, observers better remember real-world scenes (i.e., scenes from everyday environments) when they are able to freely explore images rather than hold steady fixation in the center of an image (Henderson, Williams, & Falk, 2005; Damiano & Walther, 2019; Mikhailova, Raposo, Sala, & Coco, 2021). At memory retrieval, visual exploration typically becomes more efficient: less image space is explored in remembered compared to novel or forgotten scenes (Damiano & Walther, 2019; Smith, Hopkins, & Squire, 2006).

Hence, there is considerable evidence that the extent of visual exploration of a scene is strongly associated with recognition memory accuracy. However, correct recognition can be based on a feeling of familiarity or recollection of specific contextual details about the study event (Mandler, 1980; Yonelinas & Parks, 2007; Broers & Busch, 2021). Recollection and familiarity are often assessed with remember/know statements, whereby participants indicate after an old/new statement, whether they "remember" specific episodic details about the item (indicating recollection) or whether they only "know" that the item is old (indicating familiarity). It is currently unknown whether visual exploration of complex scenes only leads to better overall familiarity with a scene or whether it supports subsequent recollection of scene details. Do fixation patterns during first viewing determine what we specifically recollect about a scene? It is conceivable that more extensive fixation patterns across the whole scene support subsequent recollection rather than familiarity. This is because recollection is generally associated with more robust memories (Yonelinas & Parks, 2007) and the extent of exploration is strongly associated with memory performance. More specifically, an increased number of fixations at encoding improves later memory for the overall scene (Kafkas & Montaldi, 2011) and broader distributions of fixations over the entire image area are strongly related to subsequent memory accuracy (Damiano & Walther, 2019; Mikhailova et al., 2021). Thus it is reasonable to assume that the extent of exploration over the entire image area supports recollection of specific image details over and above mere familiarity with a stimulus.

However, it is also conceivable that more *condensed* fixation patterns are related to subsequent recollection of specific scene details. Kafkas and Montaldi (2011) used displays comprising only a single object cropped from its background and found that fixation patterns on locations of subsequently familiar objects were more dispersed compared to subsequently recollected objects. This finding suggests that more clustered fixation patterns lead to subsequent recollection of object-specific information. This reasoning has its limits when transferred to free scene viewing: fixation patterns on single objects likely differ from fixation patterns on complex real-world scenes. Eye-movements might traverse the entire scene more extensively, but fixations around certain image aspects might be more condensed in order to prioritize encoding of specific information. For instance, (Olejarczyk, Luke, & Henderson, 2014) investigated whether eye-fixation characteristics during a visual search task are predictive of later memory of those parts. Indeed, recognition memory was better for image parts with more and with longer fixations. Moreover, scene regions that attract more fixations at encoding are better recognized later (van der Linde, Rajashekar, Bovik, & Cormack, 2009). All in all, it is likely that both eye movement–related characteristics might determine what we ultimately recollect: more extensive exploration across a larger image area, with more focal fixation allocation in memory-relevant image regions.

In this work, we investigated how fixation patterns at encoding are related to specifically recollected memories of real-world scenes in comparison to broadly familiar images. We also considered how the intrinsic memorability of an image influences memory-related viewing behavior, because previous research has shown that the memorability of an image influences both viewing behavior (Bylinskii, Isola, Bainbridge, Torralba, & Oliva, 2015; Bainbridge, 2019; Lyu., Choe, Kardan, Kotabe, Henderson, & Berman, 2020) and the propensity to recollect specific image details (Broers & Busch, 2021). In an incidental memory paradigm, observers were free to explore a series of scene photographs. Notably, observers were unaware of the subsequent memory test to avoid memory-enhancing viewing strategies during encoding. In a surprise recognition test, participants reported whether an image was new, remembered (indicating recollection) or just known to be old (indicating familiarity). To assess the specificity of their visual memories, we devised a novel report procedure in which participants selected the scene region that they specifically recollected, that appeared most familiar, or that was particularly new to them.

Methods

Participants

The methods were preregistered before data acquisition at https://aspredicted.org/blind.php?x= vs7zn8. Sixty participants with normal or corrected-to-normal vision were recruited online and via campus-wide advertisements and were compensated for participation with course credit or money (8 €/h). The data of 1 participant was incomplete because of technical issues. Moreover, there were eye-tracking issues with an additional participant, preventing data recording of the dominant eye. These 2 participants were thus excluded from further analysis. Finally, we excluded 1 trial from further analysis because subjects recognized the depicted scene (Petronas Twin Towers in Kuala Lumpur, Malaysia). To keep with the requirements of our preregistration, we excluded 14 participants with an overall hit rate less than 50%, leaving 44 participants (29 female, mean age 25.02 years), 1 more than our preregistered sample size. The study was approved by the ethics committee of the Faculty of Psychology and Sports Science, University of Muenster (no. 2018-17-NBr). All participants gave their written consent to participate. All data are available at https://osf.io/hy753/.

Apparatus

Recordings took place in a medium-lit, sound-proof chamber. Participants placed their heads on a chin-rest and could adjust the height of the table to be seated comfortably. The distance between participant's eyes and the monitor was approximately 86 cm. Stimulus presentation was controlled with the *Psychopy v1.83.04* experimental software (Peirce, 2007), on a 24 Viewpixx/EEG LCD Monitor with a 120 Hz refresh rate, 1 ms pixel response time, 95% luminance uniformity, and 1920 × 1080 pixels resolution (33.76 × 19.38; www.vpixx.com). The experiment was controlled via a computer running Windows 10, equipped with an Intel Core i5-3330 CPU, a 2 GB Nvidia GeForce GTX 760 GPU, and 8 GB RAM. Responses were logged using a wired Cherry Stream 3.0 USB keyboard (www.cherry.de).

Eye tracking

Eye movements were monitored using a desktop-mounted Eyelink 1000+ infrared based eye-tracking system (https://www.sr-research.com/) set to 1000 Hz sampling rate (monocular). Pupil detection was set to centroid fitting of the dominant eye. The eye tracker was calibrated using a symmetric 14-point calibration grid adjusted to the area where stimuli were presented (outer points located 20 pixels inward from picture edges). Saccades, blinks, and fixations were defined via SR's online detection algorithm in "cognitive" mode. In addition, fixations shorter than 50 ms were discarded from subsequent analyses. To ensure that images were perceived retinotopically centered, participants were required to fixate a central fixation symbol before a stimulus was presented (3 radius around center).

Stimuli

A total of 240 color pictures, 700 × 700 pixels (12.63 × 12.63) in size, were selected from the memorability image database FIGRIM (Bylinskii et al., 2015). The memorability of an image is its likelihood of being remembered or forgotten, derived from large-scale online recognition memory studies (Isola, Xiao, Parikh, Torralba, & Oliva, 2014; Bylinskii et al., 2015). In such studies, it was found that images are consistently remembered or forgotten, across a large number of observers (Bainbridge, Isola, & Oliva, 2013; Isola et al., 2014; Bylinskii et al., 2015). Our selection included 240 everyday scenes from 3 broad memorability categories (high: 71%-90%; medium: 55%-71%; low: 21%-55%) and from a variety of semantic categories (e.g., highway, mountain, kitchen, living-room, bathroom) without close-ups of human or animal faces or added elements such as artificial text objects. Furthermore, images were selected such that each combination of memorability and semantic categories included the same number of images and was evenly distributed between new and old images in the retrieval phase. The assignment of scenes to the "old" (presented at encoding and retrieval) and "new" (presented only at retrieval) categories was counterbalanced across participants.

Procedure

Before the experiment, participants were required to pass the Ishihara Test for Color Blindness (Clark, 1924), which all participants passed. Image memory was tested in a recognition task with an incidental encoding block on the first day and a surprise memory test in a separate session one day later (see Figure 1).

Encoding block

Participants were instructed to simply observe the images to encourage natural viewing behavior without any memory-enhancing strategies (Kaspar & Koenig, 2011). That said, it must be acknowledged that in the absence of more specific instructions, participants are likely to set their own priorities and adopt their own "tasks" (Tatler & Tatler, 2013). After eye tracking calibration and validation, participants viewed 120

Incidental Encoding

"Simply observe and explore the images"

Retrieval

"Have you seen this image before?"

"Mark the image area relevant for your decision"

Old/New?
Remember/Know?

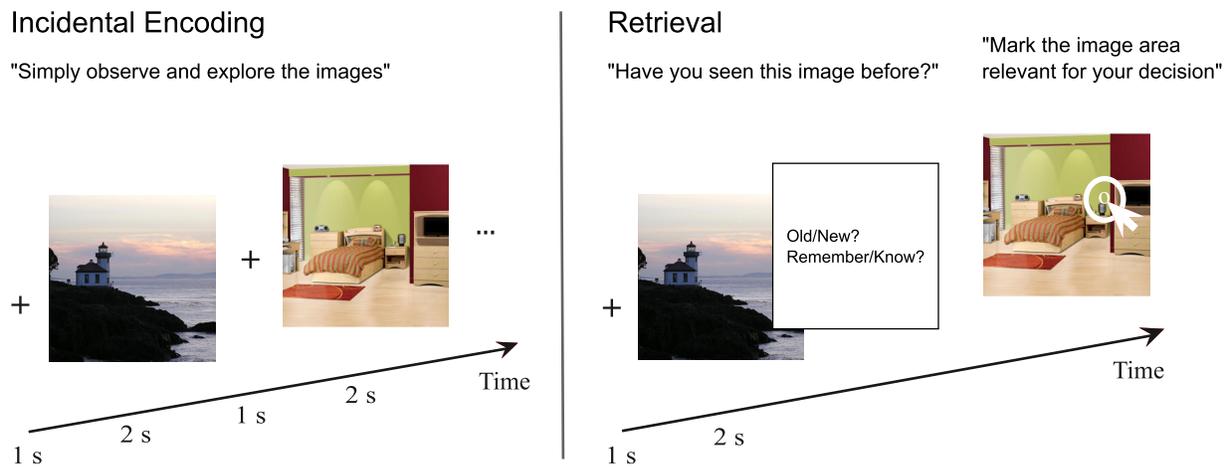Time

1 s   2 s   1 s   2 s

Time

1 s   2 s

Figure 1. Experimental procedure. Incidental encoding block: Trials started with central fixation on a fixation symbol. Upon fixation, an image was displayed at the center of the screen for 2000 ms. Participants were instructed to simply observe and to explore images when they appeared on the screen. Trials were interleaved by a blank 1000 ms intertrial interval. Testing block: As for the encoding block, participants first fixated a fixation symbol, after which an image was displayed for 2000 ms. Participants were again instructed to simply observe and to explore scenes when they appeared on the screen. Thereafter, participants were asked whether the image was "old" (presented in the first block) or "new" (not presented in the first block). After an "old" response, participants indicated whether they "remember" or "know" the picture. Thereafter, participants clicked on and drew a circle around the image area that informed their memory decision. When an image was reported as "remembered," they selected the specific recollected image area. When an image was reported as "known," they selected the image area that felt most familiar. In the case of new images, they selected the image area that appeared most novel.

images, each for 2000 ms, interleaved by a fixation cross shown for 1000 ms. The order of images was randomized across participants.

Testing block

Before the testing block, participants were thoroughly instructed and were given 6 practice trials to familiarize themselves with the procedure. Instructions for remember/know statements emphasized recommendations made by Migo, Mayes, and Montaldi (2012) to accentuate the distinction between recollection and familiarity. Specifically, know statements should be based on a feeling of familiarity for the scene as a whole, without any contextual knowledge about the encoding period. Remember statements on the other hand should be based on recollection of specific image aspects and the original encoding context. To this end, we carefully explained the concept definitions of remember/know statements. We also emphasized that remember statements do not need to refer to one particular object or feature but could also refer to multiple objects, features, or image parts. We emphasized that know statements can equally be based on high or low confidence in order to avoid a bias towards remember statements in states of high confidence.

After eye tracking calibration and validation, participants viewed 240 images, presented for 2000 ms

each, interleaved by 1000 ms intertrial intervals and a self-paced recognition test after each trial. In the recognition test, participants first indicated whether an image was old (i.e., appeared in the encoding block) or new. If the image was reported as old, participants had to indicate whether they remembered or just knew the image. They were then asked to click on and draw a circle around the image area that they specifically recollected or that was most familiar, respectively. If the image was reported to be new, participants were asked to click on and draw a circle around the image area that was particularly new to them or that they would have remembered if that image was old. The click defined the center of the circle. The size of the circle could be adjusted by dragging the mouse outward. Participants were allowed to correct the selected area and timing for this part was self-paced.

Analysis

As a measure of how extensively participants explored the images, we adapted a metric originally proposed by (Damiano & Walther, 2019), which quantifies the spatial dispersion of fixations based on the duration-weighted Euclidean distances between each fixation and the image center. Fixation distances were weighted by fixation durations following the rationale that longer fixations allow for more substantial

processing of the fixated location and are thus expected to have a stronger mnemonic impact. We adapted their formula by calculating this dispersion metric either relative to the centroid of all fixations be it located at the image center or elsewhere or relative to the centroid of all the fixations within the circle drawn by the subject. Specifically, we calculated for each trial the root-mean-square-distance (RMSD) between fixations and a reference point as:

$$RMSD = \sqrt{\frac{\sum_{f=1}^{n} d_f \left((X_f - X_j)^2 + (Y_f - Y_j)^2 \right)}{\sum_{f=1}^{n} d_f}}$$

where f enumerates the trial's individual fixations, d_f is the duration of each fixation f, (X_f, Y_f) are the fixations' coordinates in degrees of visual angle, and (X_j, Y_j) are the coordinates of the reference point. The $RMSD_{all}$ metric was computed including all fixations on each trial and with the centroid of all fixations as the reference point. The metric $RMSD_{circle}$ was computed only for fixations within the circled area and with the centroid of fixations within the circled area as the reference point. Here, we only considered trials in which at least 2 fixations fell into the circle's area (see Figure 2A for an illustration). All else being equal, larger RMSD values indicate a broader distribution or longer duration of fixations or both, whereas smaller values indicate more clustered or shorter fixations or both. Both RMSD metrics were calculated separately for trials of the encoding and retrieval blocks.

Moreover, we analyzed the proportion of fixations at encoding and retrieval that fell into the circle drawn by participants. Specifically, we calculated for each trial:

$$CFD = \frac{prop_{fix}}{prop_{area}}$$

where $prop_{fix}$ is the proportion of fixations that fell into the circled area independent of their duration and $prop_{area}$ is the proportion of the circled area relative to the total image area, according to the circle's radius r and the images' size s ($prop_{area} = \pi r^2 / s^2$). This metric adjusts for the fact that, even if the distribution of fixations was unrelated to where subjects drew the circles, larger circles would be expected to encompass more fixations. For example, if the majority of fixations fell into the circle, but the area of the circle covered almost the entire image, the circle fixation density (CFD) measure would be lower compared to a case where the same number of fixations fell into a smaller circle area (see Figure 2D for an illustration). Thereby it also adjusts for the possibility that circle areas differed between experimental conditions. Again, the CFD metric was calculated separately for trials of the encoding and retrieval blocks.

Fixation numbers, CFDs, $RMSD_{all}$ (but only in the retrieval phase) and $RMSD_{circle}$ violated assumptions of normality, so to analyze these measures we used Friedman rank sum tests, a nonparametric alternative to a 1-way analysis of variance, and Wilcoxon signed rank tests, a nonparametric alternative to a t-test. $RMSD_{all}$ from the encoding phase was normally distributed and thus compared with a 1-way analysis of variance. All post-hoc comparisons were Bonferroni corrected.

For the encoding phase, we compared the number of fixations, the $RMSD_{all}$, $RMSD_{circle}$, and the CFD between subsequently recollected, familiar, and forgotten trials. For the CFD specifically, we performed permutation tests to generate a distribution under the null hypothesis that the relationship between the circle location and fixation-related behavior is random (see 3A for an illustration). A total of 1000 permuted datasets were created by shuffling the click location labels in the original dataset at the single trial level within participants. Subsequently, using the same procedures as for the original data, CFDs were calculated for each participant, averaged across participants, and then compared with Wilcoxon Signed Rank Tests. We repeated this procedure for each permuted dataset, resulting in a probability distribution of CFDs under the null hypothesis. The estimated differences in means across permutations were then compared with the estimated difference in CFDs for subsequently recollected versus familiar scenes. Moreover, to replicate our findings we compared $RMSD_{all}$, $RMSD_{circle}$ and CFD between subsequently recollected and familiar highly memorable images to ensure that memorability-specific effects did not confound the relationship between eye-related behavior and subsequent memory. A comparison across memorability categories and within less memorable images was not feasible given the small number of recollected memories in low memorable images.

In the retrieval phase, we compared the number of fixations, circle sizes and $RMSD_{all}$ between recollected, familiar, forgotten, correctly rejected trials, and false alarms. Moreover, we compared CFDs and $RMSD_{circle}$ between recollected, familiar and falsely recognized images. Generally, we compared CFDs and $RMSD_{circle}$ only across conditions where observers (thought they) remembered an image, because these metrics depend on the selection that observers made in a scene.

Results

Memory performance

The proportion of old images correctly recognized as old (hit rate) increased with image memorability

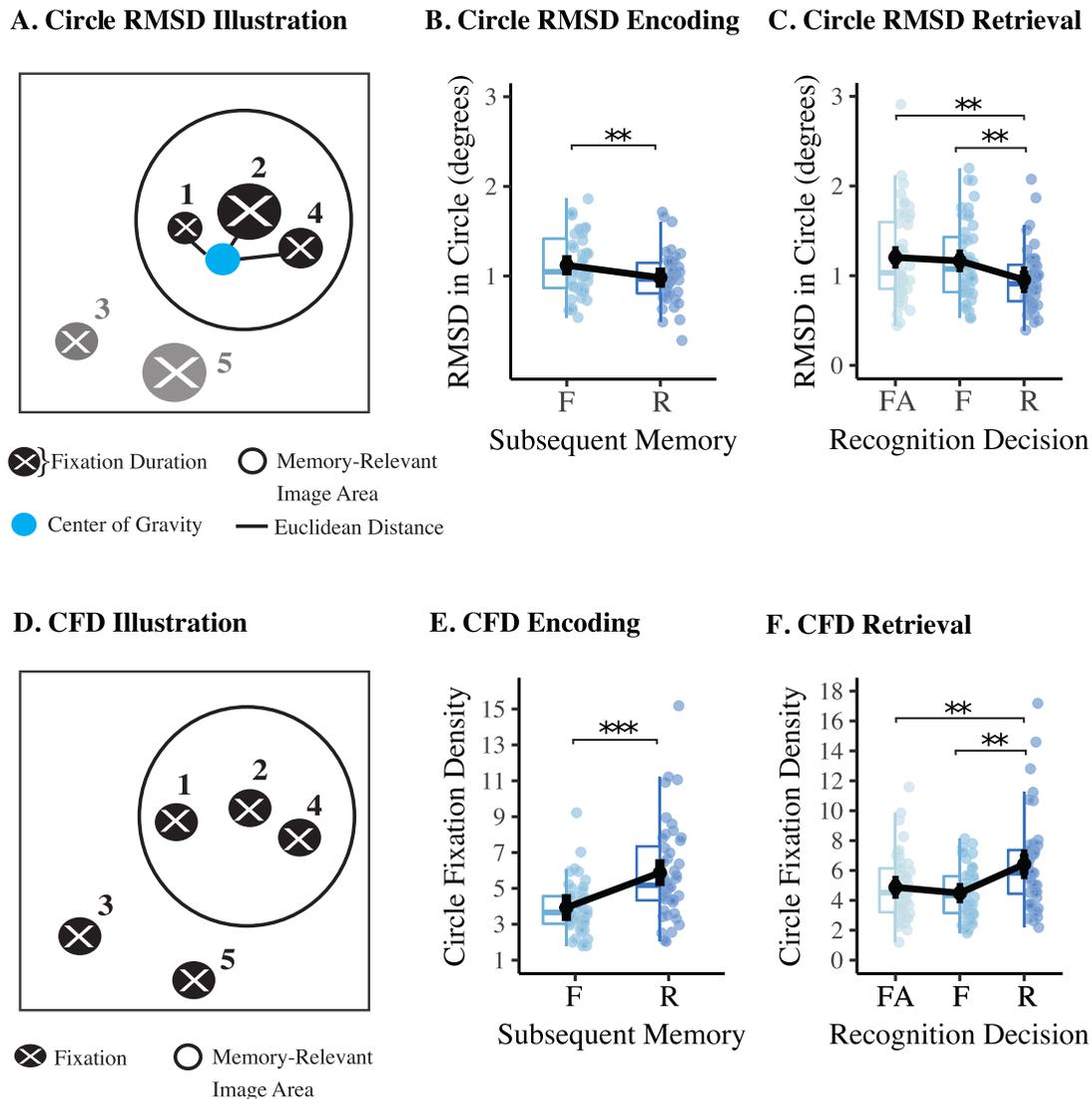# Fixation Cluster Size and Fixation Density in Memory-Relevant Areas

Figure 2. Fixation cluster and circle fixation density. Points correspond to single subjects, *black dots* and *error bars* represent the mean and within-subjects standard error, boxplots represent the underlying distribution. (A) Circle RMSD Illustration, exemplifying the average Euclidean distance between fixations and its center of mass of fixations within the circle area, weighted by fixation duration. (B) Circle RMSD between subsequently recollected versus familiar scenes. (C) Circle RMSD between recollected versus familiar versus falsely recognized scenes. (D) CFD illustration, the proportion of fixations within the circle relative to all fixations on the image are corrected by the size of the circle on the image. (E) CFD between subsequently recollected versus familiar scenes. (F) CFD between recollected versus familiar versus falsely recognized scenes. F, familiar; FA, false alarm; R, recollected. ***$p < .001$, **$p < .01$.

(low: 0.51; medium: 0.63; high: 0.71): $F(2, 86) = 72.11$, $p < 0.001$, $\eta_p^2 = 0.63$. Among those correctly recognized trials, the proportion of trials reported as "remembered" as opposed to "familiar" also increased with memorability (low: 0.36; medium: 0.44; high: 0.54): $F(2, 86) = 37.43$, $p < 0.001$, $\eta_p^2 = 0.47$. We also observed a small increase in false alarms for new images

with memorability: (low: 0.18; medium: 0.24; high: 0.24): $F(2, 86) = 14.45$, $p < 0.001$, $\eta_p^2 = 0.25$.

Memory-relevant selections

Circles indicating memory-relevant scene areas were differentially distributed across the entire image area

Permuted differences in CFD between subsequently recollected and familiar images

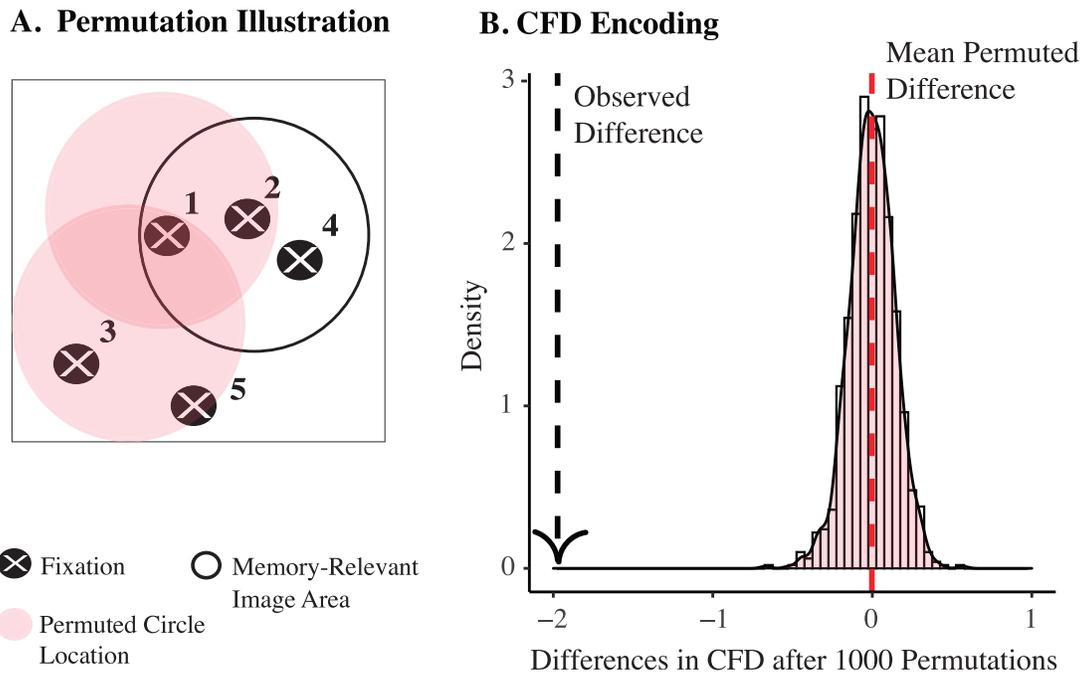### A. Permutation Illustration

B. CFD Encoding

Figure 3. CFD permutation test. (A) Permutation illustration. The circle location is shuffled across trials on a single-subject level. (B) Permutation test CFD at encoding. The distribution of estimated mean differences between subsequently recollected versus familiar CFDs under the null hypothesis. Our observed difference falls on the extreme left tail of the distribution, $p < 0.001$.

depending on the recognition decision ($RMSD_{all}$; $F(2,86) = 33.45$, $p < 0.001$, $\eta_p^2 = 0.16$). Circles indicating recollected scene areas were more distant from the image center, while circles indicating familiar areas or circles in falsely recognized images were placed closer to the image center. Specifically, memory-relevant selections were more distant from image center for recollected ($mean = 3.12°$) compared to familiar images ($mean = 2.59°$, $p < 0.001$) or false alarms ($mean = 2.65°$, $p < 0.001$). However, there was no difference between familiar images and false alarms, $p = 0.99$.

Additionally, we quantified for each image the root mean square distance between each individual circle and the centroid of all circles in that image. Thus this metric reflects how widely the circles were distributed across the image. Comparing this metric between circles indicating recollected scene areas and circles indicating familiar areas showed that recollected circles were indeed more widely distributed ($t(185) = 2.86$, $p = 0.005$).

It is intuitive to assume that the content or property making a given scene consistently memorable might also be the specific content that subjects consistently describe as most memory-relevant. To test this idea, we computed the Pearson correlation between the spatial distribution of circles across each image and the images'

memorability. However, the spatial consistency of the circles' distribution was not significantly correlated with memorability: $r = 0.01$; $p = 0.881$.

The area of these circles was smaller for recollected images ($mean_{radius} = 131.8$ pixels), compared to familiar ($mean_{radius} = 148.66$ pixels, $p = 0.003$) or falsely recognized images ($mean_{radius} = 145.99$ pixels, $p = .006$), $F(2,86) = 9.96$, $p < 0.001$, $\eta_p^2 = 0.19$). Again, there was no difference between familiar images and false alarms ($p = 0.84$). In sum, this indicates that smaller and more specific image areas were relevant for recollection compared to familiar memories.

Fixation-related statistics

Encoding

We found a significant effect of the number of fixations across subsequently recollected, familiar and forgotten scenes, $chi^2_{Friedman}(2) = 32.23$, $p < 0.001$, $\hat{W}_{Kendall} = 0.91$. Post-hoc comparisons with Bonferroni correction showed that participants made more fixations for subsequently recollected ($mean = 5.73$) versus familiar ($mean = 5.51$, $p < .001$) and forgotten scenes ($mean = 5.3$, $p < .001$). Additionally, participants made more fixations on subsequently familiar versus

forgotten scenes ($p < .001$). Thus, observers made more fixations when they successfully encoded information, especially when that information was subsequently recollected.

Moreover, we found a significant effect of $RMSD_{all}$ on the entire scene across conditions ($F(2,86) = 45.8$, $p < 0.001$, $\eta_p^2 = 0.52$). Post-hoc comparisons with Bonferroni correction revealed the largest dispersion for subsequently recollected ($meanRMSD_{all} = 2.74°$) versus familiar ($meanRMSD_{all} = 2.61°$, $p < 0.001$) and versus forgotten ($meanRMSD_{all} = 2.5°$, $p < 0.001$) scenes. In addition, $RMSD_{all}$ on subsequently familiar images was significantly larger compared to subsequently forgotten images ($p = 0.005$). This means that observers more extensively explored images when they successfully encoded information, especially when that information was subsequently recollected.

Furthermore, we found a larger density of fixations during encoding in subsequently selected memory-relevant image areas in subsequently recollected versus familiar scenes, $V_{Wilcoxon} = 4.68$, $p < 0.001$. In addition, fixations within the subsequently selected areas were more clustered (i.e., less dispersed) for later recollected ($meanRMSD_{circle} = 0.98°$) versus familiar images ($meanRMSD_{circle} = 1.12°$), $t(43) = 3.00$, $p = 0.004$ (see Figure 2A). A permutation test of memory-relevant image area location revealed that our observed difference in CFD between subsequently recollected and familiar images is indeed specific to the areas that participants reported as driving their memories (see Figure 3B), $p < 0.001$. This pattern of findings means that the interplay of more extensive exploration over the entire image area, with a larger density of more clustered fixations in memory-relevant image areas, is particularly associated with the formation of recollected memories.

Finally, we tested whether fixation patterns at encoding were influenced by the images' intrinsic memorability. Pearson correlations revealed that memorability was not correlated with the spatial dispersion of fixations across each image ($RMSD_{all}$: $r = 0.06$; $p = 0.390$), within the selected memory-relevant image areas ($RMSD_{circle}$: $r = 0.02$; $p = 0.739$), or with the proportion of fixations falling within the memory-relevant areas (CFD: $r = 0.10$; $p = 0.140$).

Retrieval

Comparing exploration behavior across recognition decisions yielded a strong effect on the number of fixations, ($F(4,172) = 21.26$, $p < 0.001$, $\eta_p^2 = 0.33$). After Bonferroni correction, recollected scenes received a smaller number of fixations ($mean = 5.47$) compared to familiar ($mean = 5.9$, $p < 0.001$), rejected-as-new ($mean = 5.99$, $p < 0.001$), falsely recognized ($mean = 6.03$, $p < 0.001$) and forgotten images ($mean = 5.98$, p

< 0.001). There were no significant differences among all other pairwise comparisons ($p > .05$). Moreover, there was no effect of recognition decision on the extent of dispersion of $RMSD_{all}$ ($chi^2_{Friedman}(4) = 8.44$, $p = 0.07$, $\hat{W}_{Kendall} = .05$).

Furthermore, we found a significant, albeit small, effect of recognition decision on the density of fixations in memory-relevant image areas, $chi^2_{Friedman}(2) = 12.41$, $p = 0.009$, $\hat{W}_{Kendall} = 0.15$. More precisely, we found a larger circle fixation density in recollected ($mean = 6.2$) compared to familiar ($mean = 4.49$, $p = 0.01$) and falsely recognized ($mean = 4.98$, $p = 0.03$) images. There was no difference however between familiar images and false alarms ($p = 0.99$). Fixations in the circle were more clustered for recollected ($mean = 0.95$) compared to familiar ($mean = 1.17$, $p = 0.006$) images and false alarms ($mean = 1.19$, $p = 0.004$), $chi^2_{Friedman}(2) = 13.68$, $p = 0.001$, $\hat{W}_{Kendall} = 0.14$ (see Figure 2B). Again, we found no difference between familiar images and false alarms ($p = 0.99$).

We also tested whether fixation patterns at retrieval were influenced by the images' intrinsic memorability. Pearson correlations revealed that memorability was not correlated with the spatial dispersion of fixations across each image ($RMSD_{all}$: $r = 0.02$; $p = 0.795$), within the selected memory-relevant image areas ($RMSD_{circle}$: $r = 0.04$; $p = 0.491$), or with the proportion of fixations falling within the memory-relevant areas (CFD: $r = 0.04$; $p = 0.556$).

To sum up, recollected images were associated with overall less fixation numbers compared to forgotten, familiar, and falsely recognized images, as well as new items. Moreover, recollected compared to familiar or falsely recognized images contain a larger fixation density in memory relevant image areas, with more clustered fixation patterns. This pattern of findings means that recollected scenes compared to all other recognition decisions were explored less (in terms of absolute number of fixations), with more focal fixation allocation to memory relevant image areas.

Discussion

It has long been established that eye movements during encoding are not a mere epiphenomenon but a determinant of the accuracy and precision of our visual memory (Ryan, Shen, & Liu, 2020). For instance, visual memory for the identity and for specific properties of objects is best for recently fixated objects but remains well above chance even if many objects are fixated thereafter (Hollingworth, 2004). When we form a coherent memory representation for an entire scene, such representations are more robust when exploration

is more extensive at first viewing (Damiano & Walther, 2019). The aforementioned findings highlight the general importance of eye movements and fixations in extracting visual information, thereby generally strengthening recognition memory. However, it is not well understood whether eye movements and fixations determine our familiarity with the scene as a whole or our recollection of specific image aspects. In this study, we confirmed previous research showing that the extent of scene exploration at encoding supports future memory retrieval (van der Linde et al., 2009; Olejarczyk, Luke, & Henderson, 2014; Damiano & Walther, 2019; Mikhailova et al., 2021). Importantly, we extend those findings by showing that scene exploration has a greater impact on future recollection compared to mere familiarity.

How is visual exploration across the entire scene related to remembering specific image details? Our results show that more extensive and more broadly distributed exploration on the entire scene is also related to recollection of specific image details. On the surface, our results are in contrast with the findings by (Kafkas & Montaldi, 2011), who showed the opposite pattern: overall more clustered fixation patterns were associated with recollection at test. However, while we used complex scenes comprising multiple objects on a natural background, Kafkas and Montaldi (2011) used displays comprising only a single object cropped from its background. Thus it is conceivable that the difference in overall fixation patterns arose due to a difference in complexity both regarding the stimulus material and the associated memory representations. More specifically, scenes can be understood and recognized based on multiple levels, from basic image properties such as edges or spatial frequencies to high-level features such as multiple object identities, spatial layouts, action affordances (Malcolm, Groen, & Baker, 2016) or the meaning of certain image regions (Lyu et al., 2020). Thus the potential for recollecting image aspects in complex scenes is greater compared to single object pictures, and such potential is more likely to be realized with greater visual exploration.

We also found that the density of fixations in memory-relevant areas was larger for subsequently recollected versus familiar scenes, with more clustered fixations within such areas. This pattern of findings suggests that the formation of recollected visual memories relies on an interplay of more elaborate exploration regarding the whole scene and more focal attention to scene regions that are later specifically recollected. It must be noted that this pattern of results does not imply that recollection for the whole scene was primarily driven by recollection of specific image areas. It instead implies that underlying memory representations for the whole scene are enriched by information extracted from more broadly dispersed fixation patterns but that more clustered fixations in

specific regions are responsible for a specific weighting of information, leading to subsequent recollection of specific image aspects. For instance, (Damiano & Walther, 2019) aimed to predict subsequent memory from fixations within pre-defined time bins in increments of 100 ms, to explore whether fixation-related behavior across the entire trial or across specific episodes during the trial was predictive of subsequent memory. It was found that subsequent memory was predicted at an above-chance level across the entire trial, suggesting that exploration differences between subsequently remembered and forgotten scenes exist throughout the entire trial. Hence, visual recollected memories become detailed via clustered fixation patterns, but they likely become robust via extensive visual exploration across the entire scene.

In addition, these results are largely robust against the impact of intrinsic image memorability. We found that intrinsic image memorability improved recognition accuracy and specifically increased the proportion of recollection as opposed to familiarity, corroborating previous findings (Broers & Busch, 2021). Highly memorable images have been shown to be viewed more consistently than less memorable images (Lyu et al., 2020) and memorability can be predicted from fixation patterns (Bylinskii et al., 2015). However, our analysis showed that the association between fixation patterns and recognition accuracy was not mediated by scene memorability, indicating that intrinsic memorability facilitated recognition via fixation-independent mechanisms. Moreover, circles indicating recognition-relevant image areas were not placed more consistently for highly memorable images. It may be plausible to expect that memories of images that are consistently remembered by most people are based on specific information that can be consistently reported. However, many highly memorable scenes do not actually contain a single, easy-to-localize object. Instead, memorability may be based on more distributed image features (e.g. a hazy sky, an unusual texture) or even non-localized, high level conceptual features (Needell & Bainbridge, 2022). If salient, low-level features drive eye movements, but not memorability (Isola et al., 2014), this may explain the lack of consistent associations between fixations and memorability. Moreover, it should be noted that the task to indicate a recognition-relevant image area is, strictly speaking, not an objective test of the content of a memory, but a meta-cognitive judgment about a memory. In this sense, our finding that these meta-cognitive judgments are not correlated with memorability is in line with previous demonstrations that people have only limited insight into how memorable an image is (Bainbridge, Isola, & Oliva, 2013; Isola et al., 2014).

It is important to mention at this point that our findings partially stand in contrast to conclusions

drawn by (Ramey, Henderson, & Yonelinas, 2020). They found that mere memory strength, compared with recollection, was more associated with more dispersed fixation patterns at encoding and less dispersed fixation patterns at retrieval. However, they found similar, if less powerful effects for recollection. In their work, memory strength was measured with confidence judgments. One might argue that the lack of memory strength measures in our study led to a stronger association between fixation patterns and memory-relevant selections for recollection compared to familiarity. The reason for this is that recollected memories have on average more highly confident hits compared to familiar memories (Yonelinas, Aly, Wang, & Koen, 2010). Instead, we would like to emphasize that memorability can be seen as a property of memory strength inherent in an image. Because our pattern of results is largely robust even among highly memorable images, we argue that not only the extent but also the specificity of spatial attention at encoding may determine increased potential for recollection over and above of mere familiarity.

Finally, fixation patterns at test exhibited a different signature compared to fixation patterns at encoding. Forgotten images were explored similarly compared to correctly judged new items, which are both explored more thoroughly and broadly compared to recollected images. More precisely, recollected scenes were explored more efficiently at retrieval, with a smaller number of fixations that were overall less dispersed. Thus, because scenes were either new or judged to be new in case of forgotten trials, exploration behavior was more extensive, suggesting that observers were searching for image information that matched an existing memory representation. It has long been thought that eye-movements at retrieval are used to reinstate the original spatiotemporal context encoded during first exposure (Noton & Stark, 1971; Ryan, Shen, & Liu, 2020), because eye movement patterns bear larger similarities (i.e., scanpath similarities) between viewing episodes for remembered versus forgotten scenes. However, scanpaths are not always repeated in correctly remembered scenes, but scenes are still remembered. It was therefore suggested that memory traces orchestrate eye movements to relevant image content regardless of whether scanpaths are repeated (Damiano & Walther, 2019). In line with that reasoning, overall exploration behavior did not differ, but memory-relevant areas had a larger fixation density and more condensed fixation clusters in recollected compared to familiar scenes. Hence, memory traces based on recollection led to specific allocation of attention to memory-relevant image content. Future research could study more specifically how image content differs among selections concerning different memory decisions.

Our results have exciting implications for the study of memory. Current models of recognition memory are agnostic towards the actual content of visual memory representations (Schurgin, 2018). Studying the content of memory-relevant areas in complex scenes may shed light on the kinds of representational content and structure supporting human visual memory. In this regard, a limitation of our study is that our procedure presupposed that participants have accurate insight into the features that drive familiarity. It is possible that the memory-relevant areas for familiar memories did not fully capture what parts of the image actually caused it to be recognized. In line with that reasoning is our finding that familiar image areas were closer to image center compared to recollected image areas. When an observer had no clear motivation about specific image content that drove their memory, they might have clicked close to the image center as a default option, simply because the gist of a scene often can best be inferred with information close to the image center. Moreover, the circles used in this study to mark the memory-relevant image areas probably did not always represent the precise boundaries of memory-relevant content. Future research could thus provide multiple, more fine-grained geometric shapes or snipping tools per trial that allow observers to more precisely select specific image aspects.

It should be noted that our results were obtained in an incidental memory paradigm with no specific instructions to remember any of the scenes, or to focus on any particular scene feature. As Tatler and Tatler (2013) have pointed out, in the absence of specific instructions, participants are likely to set their own priorities, potentially based on their idiosyncratic biases as to which image features (e.g., spatial layout) are worth looking at. Interestingly, Tatler and Tatler (2013) found that compared to a "free" viewing task, instructions to memorize a specific class of objects resulted in prioritized encoding of some object features (identity and position, but not color), indicating that different components of scene memories are differentially sensitive to task instructions. Thus it is conceivable that the relationship between fixation patterns, memory performance, and participants' judgments of recognition-relevant scene areas might be contingent on task instructions. Furthermore, future work could test whether our findings replicate in older or memory-impaired adults, where patterns of misses and false alarms have been shown to differ from that of the younger adults tested in the current study (Yeung, Ryan, Cowell, & Barense, 2013).

Unfortunately, we were not able to compare fixation patterns and selections of memory-relevant areas between specific scene categories such as man-made, indoor scenes comprising distinct objects (e.g., kitchens) and natural landscapes with fewer objects (e.g., mountains) due to the small number of exemplars per category. This comparison would be particularly interesting given that certain scene categories are

consistently more or less memorable (Bylinskii et al., 2015). It would be interesting for future studies to conduct a more detailed analysis of different scene categories using a suitable selection of scene exemplars.

Conclusion

In summary, we found that fixation patterns are strongly related to the formation and retrieval of specific memories that are consciously recollected. At retrieval, fixation patterns are generally, and in memory-relevant areas, more condensed, suggesting that underlying memory representations in concert with the oculomotor system guide attention efficiently to recollected image content. Thus viewing behavior does not only reinstate the original encoding context but is strategically implemented to match detailed memory representations with the outside world. In contrast, at encoding, subsequently recollected scenes were overall more explored but exploration behavior was more condensed in subsequent memory-relevant image areas. This interplay of more extensive but simultaneously more focal exploration behavior might be a potential solution by the visual system to build robust global representations for everyday scenes, with accentuated memory details to efficiently interact with them.

Open practices statement

Acknowledgments

References

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General, 142*(4), 1323.

Bainbridge, W A. (2019). Memorability: How what we see influences what we remember. In *Psychology of Learning and Motivation*, 1–27. St. Louis: Elsevier.

Broers, N., & Busch, N A. (2021). The Effect of Intrinsic Image Memorability on Recollection and Familiarity. *Memory & Cognition, 49*, 998–1018.

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research, 116*(November), 165–178.

Clark, J. H. (1924). The Ishihara test for color blindness. *American Journal of Physiological Optics, 5,* 269–276.

Damiano, C., & Walther, D. B. (2019). Distinct roles of eye movements during memory encoding and retrieval. *Cognition, 184*, 119–129.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition, 33*(1), 98–106.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short-and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance, 30*(3), 519.

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.

Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology, 64*(10), 1971–1989.

Kaspar, K., & Koenig, P. (2011). Viewing behavior and the impact of low-level image properties across repeated presentations of complex scenes. *Journal of Vision, 11*(13), 26–26.

Kragel, J. E., & Voss, J. L. (2021). Temporal context guides visual exploration during scene recognition. *Journal of Experimental Psychology: General, 150*(5), 873–889.

Lyu, M., Choe, K. W., Kardan, O., Kotabe, H. P., Henderson, J. M., & Berman, M. G. (2020). Overt attentional correlates of memorability of scene images and their relationships to scene semantics. *Journal of Vision, 20*(9), 1–17.

Malcolm, G. L., Groen, I. I.A., & Baker, C. I. (2016). Making sense of real-world scenes. *Trends in Cognitive Sciences, 20*(11), 843–856.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*(3), 252–271.

Migo, E. M., Mayes, A. R., & Montaldi, D. (2012). Measuring recollection and familiarity: Improving the remember/know procedure. *Consciousness and Cognition, 21*(3), 1435–1455.

Mikhailova, A., Raposo, A., Sala, S. D., & Coco, M I. (2021). Eye-movements reveal semantic interference effects during the encoding of naturalistic scenes in long-term memory. *Psychonomic Bulletin and Review, 28*(October), 1601–1614.

Needell, C. D., & Bainbridge, W. A. (2022). Embracing New Techniques in Deep Learning for Estimating Image Memorability. *Computational Brain & Behavior, 5*, 168–184.

Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research, 11*(9), 929–942.

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.

Peirce, J. W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8–13.

Ramey, M. M., Henderson, J. M., & Yonelinas, A. P. (2020). The spatial distribution of attention predicts familiarity strength during encoding and retrieval. *Journal of Experimental Psychology: General, 149*(11), 2046–2062.

Ryan, J. D., Shen, K., & Liu, Z.-X. (2020). The intersection between the oculomotor and hippocampal memory systems: Empirical developments and clinical implications. *Annals of the New York Academy of Sciences, 1464*(1), 115–141.

Schurgin, M. W. (2018). Visual memory, the long and the short of it: A review of visual working memory and long-term memory. *Attention, Perception, & Psychophysics, 80*, 1035–1056.

Smith, C. N., Hopkins, R. O., & Squire, L. R. (2006). Experience-dependent eye movements, awareness, and hippocampus-dependent memory. *Journal of Neuroscience, 26*(44), 11304–11312.

Tatler, B. W., & Tatler, S. L. (2013). The influence of instructions on object memory in a real-world setting. *Journal of Vision, 13*(February), 5.

van der Linde, I., Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2009). Visual memory for fixated regions of natural images dissociates attraction and recognition. *Perception, 38*(8), 1152–1171.

Wynn, J. S., Olsen, R. K., Binns, M. A., Buchsbaum, B. R., & Ryan, J. D. (2018). Fixation reinstatement supports visuospatial memory in older adults. *Journal of Experimental Psychology: Human Perception and Performance, 44*(7), 1119–1127.

Yeung, L.-K., Ryan, J. D., Cowell, R. A., & Barense, M. D. (2013). Recognition memory impairments caused by false recognition of novel objects. *Journal of Experimental Psychology. General, 142*, 1384–1397.

Yonelinas, A. P., Aly, M., Wang, W.-C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus, 20*, 1178–1194.

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROC s) in recognition memory: A review. *Psychological Bulletin, 133*(5), 800–832.