



The veil of distraction: Mind-wandering and memorability jointly shape visual recognition and recall

Shivang Shelat^{1,2} · Brecken Marome¹ · Casey Lopez¹ · Barry Giesbrecht^{1,2} · Jonathan W. Schooler¹

Received: 19 April 2025 / Accepted: 6 October 2025
© The Psychonomic Society, Inc. 2026

Abstract

The shift of attention from external stimuli toward task-unrelated thoughts (TUTs), or mind-wandering, has been shown to disrupt encoding and subsequent retrieval of information. The present study examined how TUT during encoding affects memory for visual scenes that vary in memorability, an intrinsic stimulus property reflecting the likelihood of recognition across individuals. In Experiment 1, participants studied images of high or low memorability while intermittently reporting their attentional state. A subsequent recognition test revealed independent effects of mind-wandering and memorability on memory performance: TUT during encoding was associated with lower recognition accuracy and high memorability images were better recognized than low memorability images, but there were no interactions. In Experiment 2, we extended this paradigm to free recall using a novel drawing-based methodology. Participants studied and later drew scenes from memory, with recall quantified as the proportion of objects drawn relative to a perception baseline. The results showed that TUT impaired free recall while memorability enhanced free recall. Further evidence suggested that memorability only conferred a benefit to free recall when images were encoded during an on-task state. These findings indicate that both attentional state and image memorability are positively associated with recognition and recall. Although these two factors independently impact recognition, there is some indication that they may interact for recall, suggesting a need for focused attention to harness the mnemonic benefits of highly memorable stimuli.

Keywords Visual memory · Mind-wandering · Memorability · Recognition · Drawing

Introduction

Mind-wandering is a familiar phenomenon for us all: the moment when one's attention shifts away from a task and toward task-unrelated thoughts (TUTs; Smallwood & Schooler, 2015). Off-task thoughts consume many of our waking moments (Kane et al., 2007; Killingsworth & Gilbert, 2010; Seli et al., 2018), and in the past 20 years, these thoughts have been linked to various costs to well-being, psychomotor performance, and memory (Blondé et al., 2022; Mooneyham & Schooler, 2013; Randall et al., 2014). We access our long-term memory every day, whether to remember what to buy at the grocery store, to recollect

warm moments with a loved one, or to recall a diagram of the cell's sodium–potassium pump. If the merciless ebb and flow of the wandering mind can derail the encoding and later retrieval of these memories, it becomes critical to ask not only when such disruptions occur but what kinds of memory are affected; this is particularly true in the domain of visual memory, which remains underexplored.

Existing research has distinguished between forms of memory that guide recognition and recall based on their reliance on attention. Recognition, where the stimulus is typically perceptually available at test, can be supported by relatively automatic processes. Recall, where the representation has to be reconstructed from memory, relies more on attention at encoding (Jennings & Jacoby, 1993; Quamme et al., 2004; see also Brosowsky et al., 2021). For this reason, recall may be especially vulnerable to attentional lapses. However, measuring visual recall has been more challenging due to difficulties in data collection and the lack of a consistent operationalization (e.g., recalling a visual memory as a single descriptive word vs. an elaborated description;

✉ Shivang Shelat
sshelat@ucsb.edu

¹ Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106, USA

² Institute for Collaborative Biotechnologies, University of California, Santa Barbara, Santa Barbara, CA 93106, USA

Erdelyi & Becker, 1974; Madigan & Rouse, 1974; Shiffrin, 1973; Tabachnick & Brotsky, 1976).

A promising avenue to test visual recall is through drawing tasks, which can capture incredibly detailed internal representations that are difficult to measure otherwise (Bainbridge et al., 2019; Fan et al., 2023). Standardized methods now allow researchers to quantify the content of drawings in memory paradigms, typically by comparing memory-based drawings to perception-based baselines and scoring them via independent raters (Bainbridge, 2022). When paired with thought probe methods that prompt participants to intermittently report whether their thoughts were focused on the task or task-unrelated content, these drawing tasks can reveal how moment-to-moment lapses in attention shape the fidelity of visual memory. TUTs are associated with perceptual decoupling, a state in which external information receives diminished processing (Smallwood et al., 2008, 2011). Stimuli encountered during such lapses are encoded less robustly, as shown by both lab and real-world studies (Baird et al., 2014; Barron et al., 2011; Hollis & Was, 2016; Wammes et al., 2016). Together, these methods provide a powerful approach for testing whether attentional state during encoding influences the quality of visual memory.

While attentional state during encoding is an established determinant of memory, it is not the only predictor of what we remember. A growing body of work suggests that not all stimuli are remembered with the same likelihood; some images are more intrinsically memorable (i.e., they tend to be remembered across individuals regardless of who sees them; Bainbridge, 2019; Rust & Mehrpour, 2020). Memorability, defined as the degree to which a stimulus is remembered or forgotten across populations, has emerged as a robust predictor of recognition memory. Notably, it predicts recognition performance independently of attentional state during encoding (Roberts et al., 2024; Wakeland-Hart et al., 2022), suggesting that some visual stimuli may generate strong memory traces even in the absence of sustained attention. While memorability reliably predicts recognition, its influence on free recall appears more limited. Prior work has found no relationship between an image's memorability and the number of scene elements recalled from memory (Bainbridge et al., 2019). However, it remains unclear whether this dissociation holds when attentional state during encoding is taken into account. Given that free recall relies more heavily on attention-dependent processes than recognition, it is possible that attentional state may shape the extent to which memorability translates into lasting memory representations.

The present study

In this study, we examined the role of attention during encoding for visual recognition and free recall across high and low memorability scenes. In Experiment 1, we sought to replicate positive, independent associations between

attentional state, stimulus memorability, and recognition performance that were previously identified in Wakeland-Hart et al. (2022). Additionally, we used a between-subjects design, which enabled us to perform exploratory analyses on whether prolonged exposure to memorable or forgettable stimuli affected overall mind-wandering, which was not possible in Wakeland-Hart et al. (2022). Our prediction was that mind-wandering and memorability would not interact based on existing literature establishing that memorability is distinct from attention (Bainbridge, 2020). We hypothesized that

- (1) Recognition performance would differ for scenes encountered during TUT,
- (2) Recognition performance would differ for groups that encountered high versus low memorability scenes, and
- (3) TUT and memorability would be additive, such that recognition performance is best predicted by a model that includes both as independent predictors.

In Experiment 2, we tested whether attention and memorability would interact to predict later free recall performance, quantified as the proportion of objects in the original scene correctly drawn from memory relative to a perception drawing control. To do this, we integrated a mind-wandering probe procedure into the novel drawing framework proposed by Bainbridge (2022). Based on prior null effects of memorability on free recall identified by Bainbridge et al. (2019) and general patterns of null interactions between attention and memorability on retrieval (e.g., Bainbridge, 2020; Wakeland-Hart et al., 2022), we hypothesized that

- (1) Recall performance would differ for scenes encountered during high TUT,
- (2) Recall performance would not differ for high versus low memorability scenes, and
- (3) TUT would impair memory encoding for both high and low memorability scenes (no interaction).

To preview our results, all of the hypotheses for Experiment 1 were supported. TUT was associated with poorer recognition memory performance. Similarly, studying low memorability images resulted in poorer recognition performance when compared to high memorability images. A model comparison showed that the best-fitting statistical model used TUT and memorability as independent predictors, reinforcing their role as unique determinants of memory encoding. However, our hypotheses for Experiment 2 were only partially supported. High TUT during encoding was associated with poorer free recall performance. Against our prediction, high memorability conferred a benefit to free recall, and the detrimental effect of

TUT during encoding was pronounced for high memorability scenes. This suggests that focused attention may be crucial for memorability to confer a benefit to recollective memory encoding of objects within scenes.

Experiment 1: Mind-wandering and recognition

Methods

Participants

Using an α criterion = .05 and power = .80, we completed a sensitivity power analysis using G*Power (Faul et al., 2007) before data collection and found that a sample of 45 would detect small-to-medium effects ($d_z = 0.43$). Because we reasoned that online data might be noisier than in-person data, we preregistered a more conservative target of 60 participants, which is twice the usual sample size for prior drawing studies (Bainbridge et al., 2019, 2021; Hall et al., 2022). Although our preregistration specified stopping at 60 participants, we slightly over-recruited (67 total) to account for preregistered exclusions and ensure that 60 viable participants remained for analysis. Note that we originally planned on assessing drawings from memory by administering a drawing task before the recognition test in Experiment 1; however, due to a bug with the sketchpad function in jsPsych, we were not able to convert any of the base64 drawing data into images for viewing and analysis.¹ Elements of the procedure regarding this part of the design will not be mentioned further in this article.

Of the 67 students from the subject pool at the University of California, Santa Barbara that participated in our experiment, we removed six under our preregistered exclusion criteria (failed attention check, $n = 5$; false alarm rate three standard deviations (SDs) above the sample mean, $n = 1$). The study used an Institutional Review Board (IRB)-approved incomplete disclosure protocol: participants first consented knowing some task detail was withheld (the nature of the memory test), were fully debriefed after, and then chose whether their data could be retained; one declined and was removed. This left 60 adults (21 male, 39 female, 0 non-binary; mean [M] \pm SD; $M_{\text{age}} = 18.87 \pm 1.02$ years) who generally reported low drawing ability on a scale from 0 to 4 ($M = 0.90 \pm 0.95$). They were randomly placed into the high ($n = 30$) or low ($n = 30$) memorability condition and participated in exchange for course credit for a one-hour experiment. We required that participants have access to a

quiet space and a computer with a mouse (i.e., not a trackpad or touchscreen) to complete the study.

Materials

Scene stimuli A total of 72 images were used: three images that varied on memorability from each of 24 unique-scene categories. For the study phase of the experiment, participants saw either 24 high memorability or 24 low memorability images. These images were a subset of those used in Bainbridge et al. (2019) and are originally from the Scene Understanding Database (SUN; Xiao et al., 2010). Their memorability was previously quantified in a large-scale recognition study by Isola et al. (2011); high and low memorability bounds were determined as being at least 0.5 SDs above or below the average hit rate, respectively. Another 24 medium memorability images were taken from the dataset to serve as foils for the recognition portion of the study.

To further match the stimuli, we generally equated the number of visually distinct, nameable, separable objects larger than a 50-pixel diameter, which were provided by authors of the original paper. The high memorability set had an average of 18.38 ± 8.53 objects and the low memorability set had an average of 21.79 ± 10.33 objects.

Digit span task The forward digit span task presented a series of digits, starting at a length of 3, one at a time for 1 s each with no interstimulus interval (ISI). Digits were sampled from 0 to 9 without replacement so there were no repeats within a sequence. Participants had to type the digits in order into a single text field. Accuracy feedback was presented after each trial. The task used a 1:2 ladder design, such that a correct response increased the next trial's span length by 1 (maximum length = 9) and two consecutive incorrect responses decreased the next trial's span length by 1 (minimum length = 3). The purpose of this task was trifold: to sample a participant's maximal digit span, to insert a delay before the retrieval phase, and to ensure that the participant did not use working memory strategies (e.g., inner verbal rehearsal) to maintain image details in their head.

Procedure

Participants first underwent the study phase, where they studied 24 scenes sized at 500×500 pixels for 15 s each with an ISI of 0.5 s. They were informed that they should study the stimuli in preparation for a test later but were not told the type of test. Images were presented in blocks of four, where the latter three had an equal chance of being succeeded by a TUT probe. The probe wording was adapted from Kane and colleagues (2021) and had participants report whether they were thinking about the task, everyday things, current state of being, personal worries, daydreams, external

¹ Preregistered drawing task details for Experiment 1 can be found on the Open Science Framework at osf.io/e2sw6.

environment, or other. “The task” was defined as focusing on studying the images, “everyday things” as routine events in the recent or distant past or future, “current state of being” as physical or emotional states, “personal worries” as life concerns, “daydreams” as fantasies or thoughts disconnected from reality, “external environment” as objects or task-unrelated events in the room, and “other” as any thought not captured by the other categories. The first option is considered on-task (OT), and the others are considered TUT. Because there was one probe per block, each participant reported their mind-wandering a total of six times with a minimum of ~30 s and a maximum of ~90 s between probes. After finishing the study phase, participants responded to a retrospective scale item that asked them to rate the extent of their mind-wandering during the entire phase. They next completed 20 trials of the digit span distractor task.

Afterward, participants completed a recognition test in which they saw the original 24 encoded images intermixed with 24 foils of matched scene categories. Participants were instructed to indicate old/new for each image and had unlimited time to respond. In between responses, there was a 0.5 s ISI. They then completed a brief survey, entering demographic information (age, gender), rating their drawing ability in response to the question “Rate your drawing ability” on a 5-point scale (0 = very poor, 1 = below average, 2 = average, 3 = above average, 4 = excellent), and indicating if they approved their data being used for our analyses. A procedure schematic is presented in Fig. 1. The procedures were approved by the Human Subjects Committee at the University of California, Santa Barbara. Experiment 1 was preregistered on the Open Science Framework at osf.io/e2sw6.

Results and discussion

For Experiment 1, we used multiple statistical approaches to test our hypotheses. For tests using collapsed data that were non-normally distributed, we used Kendall’s tau tests and two-tailed Welch’s independent-samples t-tests assuming unequal variances. To handle non-normality, all t-tests were supported using a permutation approach (see Bullock et al., 2017), which involved shuffling the memorability condition labels 10,000 times and recalculating the t-statistic for each permutation. This generated a null distribution of t-statistics which we then compared to our observed t-statistic. The p-value was calculated as the number of permuted t-statistics greater than the observed statistic divided by the total number of permutations. For all t-tests, we report both the initial observed p-value (p) and the permutation-based p-value (p_{perm}). In addition to these correlation tests and groupwise comparisons, we adopted a logistic mixed-effect regression approach to model hit probabilities on memory trials based on scene memorability and corresponding attentional state during encoding. For each of these models, we retained the long format of our data and assigned subjectID as a random effect.

First, we calculated overall memory metrics based on hits and false alarms in the recognition test. Hit rate was calculated as (correct recognitions / [correct recognitions + misses]). False alarm rate was calculated as (false recognitions / [false recognitions + correct rejections]). To enable the calculation of other signal detection metrics, we followed recommendations by Macmillan and Kaplan (1985) and replaced rates of 0 with $(0.5 / 24)$ and of 1 with $([24 - 0.5] / 24)$. We selected 24 as our correction value because it was the number of stimuli available to encode

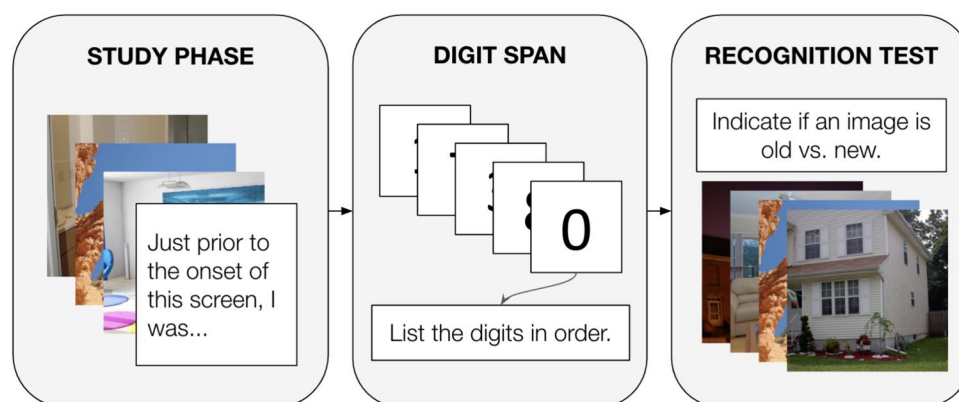


Fig. 1 Phases of the procedure for Experiment 1. Participants completed a study phase where they encoded 24 complex visual scenes. Then, after a digit span distractor task, they completed a recognition memory test

during the study phase. Afterward, we calculated d' , which is a corrected recognition rate that accounts for false alarm rate. We used the formula $(\text{qnorm}[\text{hit rate}] - \text{qnorm}[\text{false alarm rate}])$, where $\text{qnorm}()$ applies the inverse cumulative distribution function of the standard normal distribution to the input value. We also computed criterion as an index of response bias using the formula $(-0.5) \times (\text{qnorm}[\text{hit rate}] + \text{qnorm}[\text{false alarm rate}])$.

We moved on to calculate our TUT metrics. First, all instances where participants reported being focused on the task were coded as OT; the rest were classified as TUT. We calculated the proportion of TUT responses across all thought probes and transformed these proportions to z-scores. This was significantly correlated with z-scored responses to the single-item retrospective TUT report ($\tau_b = .475$, $p < .0001$). We then averaged these two TUT metrics together to create a single value that represented overall TUT.

We began our main analyses by performing exploratory correlation tests with d' and trait-level measures. Memory performance was not correlated with maximal digit span ($M = 7.78 \pm 1.09$; $\tau_b = .00068$, $p = .99$) or drawing ability ($\tau_b = -.066$, $p = .51$). Next, we tested our first hypothesis: recognition performance should differ for scenes encountered during mind-wandering. Overall TUT was significantly associated with a lower d' ($\tau_b = -.32$, $p = .00037$; see Fig. 2a) and not associated with criterion ($\tau_b = .049$, $p = .58$; see Fig. 2b). For completion, we also performed correlations with hit rate ($\tau_b = -.30$, $p = .0015$) and false alarm rate ($\tau_b = .29$, $p = .0020$).

After, we used a mixed-effects modeling approach to assess the trial-by-trial association between TUT and memory. Here, and in the rest of the article, we report

unstandardized betas. We fitted a logistic regression using the formula $\text{hit_one_back} \sim \text{tut} + (1 \mid \text{subjectID})$, where hit_one_back represents the memory performance for the scene presented directly before a thought probe. Participants were approximately 2.61 times likely to fail to recognize scenes encountered during TUT relative to OT ($b = -0.96$, $\text{SE} = 0.36$, $95\% \text{ CI} = [-1.67, -0.25]$, $p = .0080$). To examine the temporal sensitivity of thought reports, we also assessed whether responses were associated with the scene encountered before the probe. Our model was $\text{hit} \sim \text{tut} \times \text{stimulus_position} + (1 \mid \text{subjectID})$, where stimulus_position indicated whether a scene was one or two before the thought probe. There was no effect of stimulus_position ($b = -0.22$, $\text{SE} = 0.39$, $95\% \text{ CI} = [-0.98, 0.53]$, $p = .56$) nor an interaction between stimulus_position and TUT ($b = 0.35$, $\text{SE} = 0.46$, $95\% \text{ CI} = [-0.55, 1.24]$, $p = .45$). Again, scenes encountered during TUT were around 2.59 times more likely to be missed during test ($b = -0.95$, $\text{SE} = 0.35$, $95\% \text{ CI} = [-1.64, -0.27]$, $p = .0063$; see Fig. 2c). Together, the results indicate that TUT reports during encoding were associated with poorer recognition memory performance for up to two images presented before the thought probe.

Having established that mind-wandering impairs memory encoding, we tested our next hypothesis: memorability affects memory performance. We compared d' between the participants who encountered high memorability scenes and those who studied low memorability scenes (see Fig. 3a). As expected, studying high memorability scenes resulted in a higher d' when compared to low memorability scenes ($M_{\text{high}} = 2.72 \pm 1.17$, $M_{\text{low}} = 2.16 \pm 1.00$, $95\% \text{ CI } M_{\text{diff}} = [0.0053, 1.13]$; $t(56.71) = 2.022$, $p = .048$, $p_{\text{perm}} = .049$, $d = 0.52$). Participants who studied high memorability scenes also exhibited a lower criterion than those in the low memorability

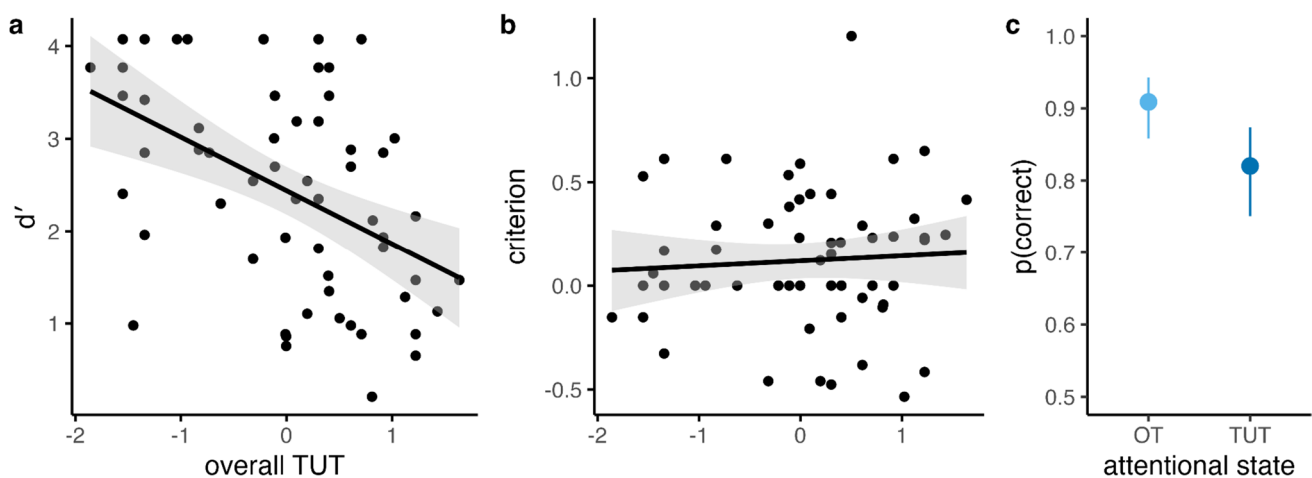


Fig. 2 Associations between overall task-unrelated thought (TUT) and two signal detection memory metrics: **(a)** d' and **(b)** criterion. **(c)** Logistic regression predictions for the TUT main effect using the model $\text{hit} \sim \text{tut} \times \text{stimulus_position} + (1 \mid \text{subjectID})$. The y-axis indi-

cates the probability of a correct recognition during test (i.e., a hit). For these figures and all afterwards, error shading and bars represent the 95% confidence interval

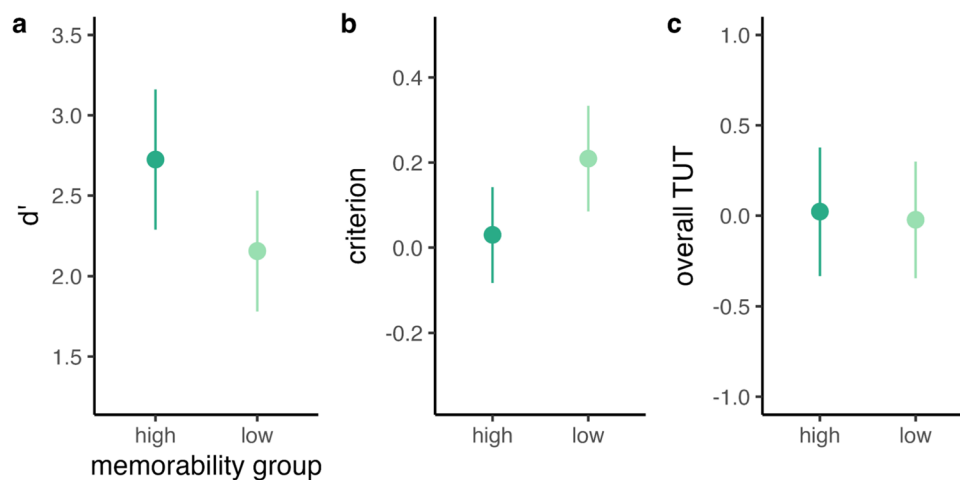


Fig. 3 Comparisons between memorability groups for (a) d' , (b) criterion, and (c) overall task-unrelated thought (TUT)

group ($M_{\text{high}} = 0.030 \pm 0.30$, $M_{\text{low}} = 0.21 \pm 0.33$, 95% CI $M_{\text{diff}} = [-0.34, -0.015]$; $t(57.42) = -2.19$, $p = .033$, $p_{\text{perm}} = .035$, $d = -0.57$; see Fig. 3b). Again, for completion, we compared hit rate and false alarm rate between the groups. The high memorability group had a higher hit rate ($M_{\text{high}} = 0.86 \pm 0.16$, $M_{\text{low}} = 0.77 \pm 0.16$, 95% CI $M_{\text{diff}} = [0.0017, 0.17]$; $t(58) = 2.042$, $p = .046$, $p_{\text{perm}} = .048$, $d = 0.53$) and a non-different false alarm rate ($M_{\text{high}} = 0.12 \pm 0.10$, $M_{\text{low}} = 0.14 \pm 0.12$, 95% CI $M_{\text{diff}} = [-0.079, 0.039]$; $t(56.25) = -0.68$, $p = .50$, $p_{\text{perm}} = .50$, $d = -0.18$) when compared to the low memorability group.

The between-subjects nature of this experiment enabled us to assess whether prolonged exposure to memorable or forgettable stimuli affected overall mind-wandering. To this end, we ran an exploratory comparison on the overall TUT metric. There were no differences between the high and low memorability groups in overall TUT ($M_{\text{high}} = 0.023 \pm 0.95$, $M_{\text{low}} = -0.023 \pm 0.86$, 95% CI $M_{\text{diff}} = [-0.42, 0.51]$; $t(57.45) = 0.19$, $p = .85$, $p_{\text{perm}} = .85$, $d = 0.050$; see Fig. 3c), indicating no evidence that processing stimuli of a particular memorability affects lapse rates.

Wakeland-Hart et al. (2022) concluded that tracking both sustained attentional state and stimulus memorability enables the best predictive model of memory performance. To corroborate this, we tested whether linear models that included TUT and memorability as independent main effects on memory performance (d') would provide the best statistical fit compared to models with each factor alone or their interaction (Table 1). We hypothesized that mind-wandering and memorability will be additive, such that recognition performance is best predicted by a model that includes both. Model fit was assessed using the Akaike Information Criterion (AIC), which penalizes overcomplexity (Akaike, 1973). A lower AIC indicates better fit, with differences greater than 2 considered substantial (Burnham & Anderson, 2004).

Table 1 A linear model comparison for the effect of task-unrelated thinking (TUT) and memorability on d' , arranged from lowest to highest AIC

Model	AIC	Δ AIC	R^2	Adjusted R^2
TUT + memorability	170.26	0	0.29	0.26
TUT \times memorability	170.89	0.63	0.30	0.26
TUT only	174.04	3.78	0.22	0.20
Memorability only	184.57	14.31	0.066	0.050

Table 2 Linear regression model results when fitting d' as a function of overall task-unrelated thought (TUT) and memorability

Term	beta	SE	95% CI	t	P
(Intercept)	2.74	0.175	[2.39, 3.09]	15.62	<.0001
Overall TUT	-0.59	0.14	[-0.86, -0.31]	-4.22	<.0001
Memorability condition (low)	-0.60	0.25	[-1.092, -0.099]	-2.40	.020
R^2	0.29				
Adjusted R^2	0.26				

SE = standard error of the mean, CI = confidence interval

The results showed that the numerically best-fitting model included TUT and memorability as independent predictors (AIC = 170.26; Table 2), though its fit was not substantially different from the interaction model (TUT \times memorability, AIC = 170.89). Both models outperformed those with TUT alone (AIC = 174.04) or memorability alone (AIC = 184.57). Despite the comparable AIC of the independent-predictor and interaction models, the interaction term in the latter was not significant ($b = 0.32$, $SE = 0.28$, 95% CI = [-0.24, 0.87], $p = .26$). As preregistered,

we performed exploratory interaction tests on criterion, hit rate, and false alarm rate. Overall TUT and memorability did not interact to affect criterion ($b = 0.0022$, $SE = 0.094$, $95\% CI = [-0.19, 0.19]$, $p = .98$), hit rate ($b = 0.013$, $SE = 0.043$, $95\% CI = [-0.073, 0.10]$, $p = .76$), or false alarm rate ($b = -0.020$, $SE = 0.031$, $95\% CI = [-0.082, 0.042]$, $p = .52$).

Based on the earlier result indicating that encoding up to two scenes before the thought probe is sensitive to TUT reports, we applied the same approach using the long version of our dataframe. Instead of modeling d' , we used logistic mixed-effects regressions with hit probability as the dependent variable. We performed the same AIC model comparison, controlling for stimulus_position as a fixed effect and assigning subjectID as a random effect in all models (Table 3).

The memorability-only model (AIC = 653.32) and TUT-only model (AIC = 650.04) were both outperformed by models with TUT and memorability as independent predictors (AIC = 645.19; Table 4) or as an interaction (AIC = 646.64). Again, although the independent-predictor and interaction models showed similar AICs, the interaction term in the latter was nonsignificant ($b = -0.38$, $SE = 0.51$, $95\% CI = [-1.39, 0.62]$, $p = .46$). These patterns in the results are consistent with the hypothesis that attentional state and memorability contribute independently to the probability of a correct memory trial during the recognition test.

Table 3 A logistic model comparison for the effect of task-unrelated thought (TUT) and memorability on hit probability, arranged from lowest to highest AIC

Model	AIC	Δ AIC	Marginal R^2	Conditional R^2
TUT + memorability	645.19	0	0.078	0.27
TUT \times memorability	646.64	1.45	0.073	0.27
TUT only	650.04	4.85	0.033	0.26
Memorability only	653.32	8.13	0.045	0.29

Table 4 Logistic mixed-effect regression model results when fitting hit as a function of task-unrelated thought (TUT) report and memorability

Term	beta	SE	95% CI	z	p
(Intercept)	2.740	0.337	[2.079, 3.401]	8.124	<.0001
TUT	-0.796	0.253	[-1.292, -0.301]	-3.153	.00162
Memorability condition (low)	-0.874	0.338	[-1.538, -0.211]	-2.584	.00978
Stimulus position (hit two back)	0.0216	0.208	[-0.386, 0.429]	0.104	.92
R^2 Marginal	0.08				
R^2 Conditional	0.04				

SE = standard error of the mean, CI = confidence interval

Experiment 2: Mind-wandering and free recall

The results from Experiment 1 suggested that memorability and mind-wandering during encoding are distinct predictors of future recognition memory test performance. Memory performance was worse in a group that studied low memorability scenes than in a group that studied high memorability scenes. Images encountered during TUT were less likely to be recognized later on. The effect of memorability on recognition did not vary across on-task versus off-task states, nor did exposure to high versus low memorability scenes significantly alter TUT rates.

We next implemented an in-person experiment aimed to test the effects of mind-wandering and memorability on free recall memory performance. We quantified free recall as the proportion of objects drawn from memory relative to objects drawn from perception. To restate our hypotheses, we predicted that

- (1) Recall performance would differ for scenes encountered during TUT,
- (2) Recall performance would not differ for high versus low memorability scenes, and
- (3) TUT would impair memory encoding for both high and low memorability scenes (no interaction).

These hypotheses were driven by prior work suggesting no effect of memorability on free recall (Bainbridge et al., 2019), and research indicating that attention and memorability are generally separable and do not interact to shape retrieval (Bainbridge, 2020; Wakeland-Hart et al., 2022). Notably, no prior study to our knowledge has examined the relationship between attention and memorability with a recall measure.

Methods

Participants

For Experiment 2, we chose our sample size solely based on precedent in previous drawing recall studies. We aimed for

40 participants, which is 10 more than the 30 typically used (Bainbridge et al., 2019, 2021; Hall et al., 2022) so that at least 30 viable datasets would remain after exclusions. We recruited all 40 individuals from the same participant pool as Experiment 1. Potential participants were not allowed to sign up for Experiment 2 if they had participated in Experiment 1.

We removed one participant for drawing only two scenes from memory during the free recall test. This left us with 39 adults (14 male, 23 female, 2 non-binary; $M_{\text{age}} = 20.46 \pm 1.80$ years) who largely reported low-to-average drawing ability from 0 to 4 ($M = 1.21 \pm 1.06$). A post hoc sensitivity analysis at $\alpha = .05$ and power = .80 confirmed that this sample size allowed us to detect small-to-medium effects ($d_z = 0.46$). All participants reported normal or corrected-to-normal color vision. They completed the experiment in exchange for either course credit ($n = 4$) or 20 USD ($n = 35$) for a one-hour session. We did not find any evidence that the type of compensation affected average memory performance ($p = .88$) or TUT ($p = .74$), but this should be interpreted conservatively since group counts were highly imbalanced.

Materials

Scene stimuli We used three high memorability (garden, playground, kitchen) and three low memorability (amusement park, living room, bedroom) scenes from the first experiment. The scenes were sized at 500×500 pixels and presented during the study phase in a randomized order. Each was on the screen for 15 s with a 0.5 s ISI, and a TUT probe appeared after each image. In their materials, Megla and colleagues (2025) provided the same scenes with annotations such that each nameable, visually distinct object had a red outline around it. We later showed each annotated object in each scene to research assistants, who scored whether that outlined object was present or not (see scoring procedure subsection below). Like in Experiment 1, scenes were generally equated on their number of objects; high memorability scenes had an average of 22.67 ± 15.82 objects and low memorability scenes had an average of 24.00 ± 3.00 objects.

Change detection task To gather a measure of visual working memory capacity k per participant, we administered 12 practice trials and 120 main trials of a change detection task with set sizes 2, 4, and 6 (Luck & Vogel, 1997). We computed two distinct accuracy proportions for change and no-change trials per set size. We then considered the accuracy proportion for change trials as a hit rate and the accuracy proportion for the no-change trials subtracted from 1 as a false alarm rate. For each set size, we calculated Cowan's k as $(\text{set_size} \times [\text{hit_rate} - \text{fa_rate}])$. Finally, we averaged those values to produce one k per person.

Current concerns task The current concerns task was designed by Kopp et al. (2015) to induce attentional fluctuations in a subsequent task. Like in the present experiment, it has been administered across an entire sample to elevate baseline TUT (Zhang et al., 2024). Participants were told the following: "Take a few minutes and think about your upcoming plans. In the space below, please write out a to-do list of the things that you have to do in the next 5 days. Please try to be as complete as possible." Participants were provided with a free response box to write out their plans. A button to continue to the next part of the experiment only appeared after three minutes to encourage responding. We informed participants that the content of what they wrote would not be analyzed or published.

Digit span task We used the same digit span task described in the Materials section of Experiment 1.

Experimental procedure

Upon entering the lab, participants were seated 88 cm away from a 48×27 cm monitor with a resolution of $1,920 \times 1,080$ pixels. Because the encoding phase of Experiment 2 was brief (six scenes; shorter than Experiment 1), we wanted to avoid floor-level TUT during study. Therefore, we first administered the change detection task and the current concerns induction to increase later mind-wandering rates. The change detection task also enabled exploratory correlations between visual working memory capacity and recall.

Participants then began the study phase. They were told that they would be interrupted by a probe after studying each image, and that the probe would ask them to rate the extent of their TUT while that particular image was on the screen. Like in Experiment 1, we defined mind-wandering as task-unrelated thoughts (TUTs). However, instead of using a categorical thought-content probe, we opted to use a slider scale to assess the degree of TUTs during each image's presentation. The probe asked participants: "How attentive were you?" Responses were on a 100-point slider scale from 0 (on task) to 100 (a lot of task-unrelated thoughts). Because of the small number of probes, it was critical to measure small-scale fluctuations in attentional state that might only be captured by slider responses. After being shown an example of the probe, participants were told that there would be a memory test on the images that they were studying, but they were not aware of what the test would entail.

After studying the images, participants completed the digit span task for the same reasons cited in Experiment 1. Participants were then informed of the nature of the test phase. The experimenter read out the following: "You will now draw as many of the pictures that you studied as you can from memory. You will be given as many sheets of paper as you need and unlimited time to draw. You should use

one sheet of paper per scene that you draw and only draw in the square outline. You can use colors and text to label parts of the scenes that might be hard to draw. Do you have any questions for me?” Participants were provided with 12 colored pencils (red, red-orange, orange, yellow, yellow-green, green, sky blue, blue, purple, white, brown, black), a sharpener, and sheets of paper with 5.21×5.21 in. borders on them. A reminder on the screen informed participants that they could label any details or objects in their drawings that they thought were unclear and that they could also use color, but to not insert colors unless they specifically recalled them in the image.

Once participants signaled the experimenter that they were done drawing, they proceeded to a phase where they copied the scenes while they were presented onscreen. This was to provide a within-subject and per-scene control for the memory drawings. The experimenter said: “You will now draw each of these pictures while they are up on the screen. You have unlimited time to draw each scene. Again, you should use one sheet of paper per scene that you draw and only draw in the square outline. You can use colors and text to label parts of the scenes that might be hard to draw. Do you have any questions for me?” Drawing times were not recorded during any phase.

After finishing copying the scenes, participants worked with the experimenter to label all memory and perception drawings. They were shown the original scenes onscreen and were instructed to tell the experimenter which of their drawings was supposed to represent which scene. The experimenter labeled each drawing, which created correspondence between all drawings and what participants intended to draw. Finally, participants reported their age, gender, and self-rated drawing ability. They provided consent for the use of their data in our analyses. A procedure schematic is presented in Fig. 4. The procedures were approved by the Human Subjects Committee at the University of California, Santa Barbara. Experiment 2 was not preregistered.

Scoring procedure

In a preliminary assessment, a research assistant noted if a drawing was low effort, contained color (i.e., used any color other than black), or used text. If a drawing did contain color or text, they tallied the number of colors used and the count of text labels. For scoring, three separate research assistants provided binary judgments on whether objects that were in the original scene were present in the drawing or not. For each object, the annotated scene described in the *Materials* section was presented with a scan of the drawing. All three of these latter scorers were completely blind to the experimental conditions and research questions.

We noticed that a subset of drawings were faint in scans, but not on the actual sheets. We rescanned to ensure these were not a fluke of the scanning machine. For the 46 drawings that still remained faint, we slightly increased the contrast of the scan to enhance the drawing’s visibility for our scorers. We also discarded three drawings that were not matched to any scene or that a participant described as a false memory.

First, to assess whether our object identification rating approach was reliable, we calculated Fleiss’ kappa for all ratings and for each scene individually. Fleiss’ kappa is appropriate to test interrater agreement when there are more than two raters and the administered rating is categorical (Fleiss, 1971); in our case, three raters provided a 0 if an object was not present and a 1 if it was. To interpret the kappa statistic κ , we used guidelines put forward by Landis and Koch (1977): 0.0–0.2 is slight agreement, 0.21–0.4 is fair, 0.41–0.6 is moderate, 0.61–0.8 is substantial, and > 0.8 is almost perfect. Ratings across all drawings yielded an excellent κ of 0.84, $p < .0001$. When we isolated the kappa test to each scene independently, estimates ranged from substantial to almost perfect ($\kappa_{\text{garden}} = 0.77$, $p < .0001$; $\kappa_{\text{kitchen}} = 0.85$, $p < .0001$; $\kappa_{\text{playground}} = 0.80$, $p < .0001$; $\kappa_{\text{amusement park}} = 0.82$, $p < .0001$; $\kappa_{\text{bedroom}} = 0.84$, $p < .0001$; $\kappa_{\text{living room}} =$

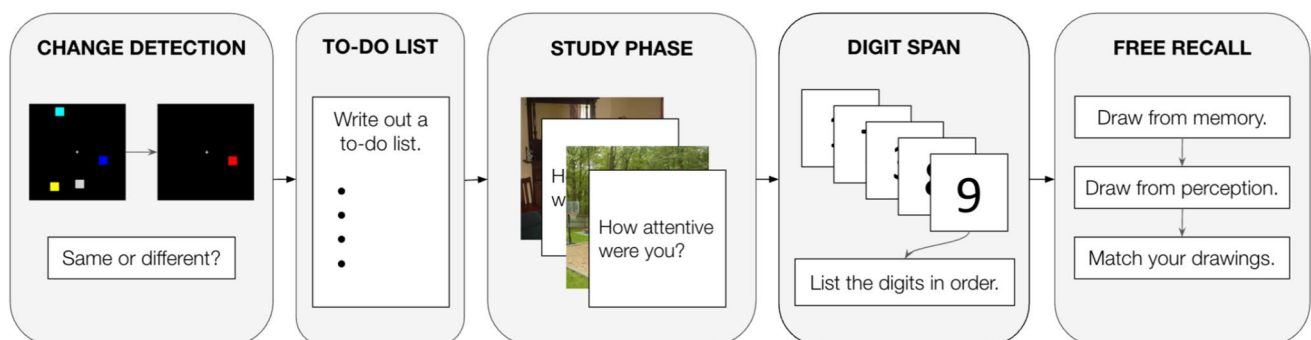


Fig. 4 Phases of the procedure for Experiment 2. Participants completed a change detection task, a to-do list “current concerns” induction, and a study phase where they encoded six complex visual

scenes. Then, after a digit span distractor task, they drew those scenes from memory and perception. Finally, they worked with the experimenter to match their drawings to the scenes they intended to draw

0.87, $p < .0001$). Note that these statistics were calculated while including drawings from a participant whose data was later removed for producing only two from memory (see *Participants* section).

With reliable object-presence estimates from our in-lab rating scheme, we followed with a “majority rules” approach to classify an object as present or not (e.g., Megla et al., 2025). If two or more raters agreed that an object was in a drawing, we coded it as present. Finally, we summed the count of present objects per drawing. A memory score $p(\text{objects drawn from memory})$ was calculated per scene as (count of objects drawn from memory / count of objects drawn from perception).

At this stage, we applied two filtering criteria. First, we removed one memory–perception pair since the participant drew more during memory than perception, which we attributed to a waning in motivation toward the end of the experimental session when the perception drawings took place. Second, we removed two memory–perception pairs because the experiment proctor noted that either their memory or perception drawing for that trial was low effort. The resulting data had 185 memory–perception pairs from 39 participants with corresponding memorability categories and TUT reports from the study phase.

Results and discussion

As in the first experiment, we performed exploratory Kendall’s tau tests with individual differences and average $p(\text{objects drawn from memory})$. Maximal digit span ($M = 7.44 \pm 1.14$; $\tau_b = -.11$, $p = .36$), drawing ability ($\tau_b = .14$, $p = .26$), and visual working memory capacity k ($M = 1.97 \pm 0.77$; $\tau_b = -.0095$, $p = .93$) were not significantly associated with free recall performance. Next, we used a linear mixed-effects modeling approach to test our primary hypotheses. Linear mixed models adhere to the within-subjects design and the continuous nature of the dependent variable (i.e., the amount of objects drawn from memory relative to those drawn from perception).

We fitted memory performance as a function of the TUT report during the study phase while assigning subjectID as a random effect. Higher TUT was associated with a lower proportion of objects drawn from memory ($b = -0.0016$, $SE = 0.00071$, $95\% \text{ CI} = [-0.0030, -0.00023]$, $p = .024$), indicating that deeper mind-wandering disrupts memory encoding. We next tested the effect of memorability on memory performance, anticipating that there would be no effect based on prior work (Bainbridge et al., 2019). However, our model revealed that participants had a lower drawn proportion of objects from memory for low memorability scenes than high memorability scenes ($b = -0.083$, $SE = 0.027$, $95\% \text{ CI} =$

$[-0.14, -0.031]$, $p = .0023$). This suggests that free recall performance, when operationalized as the count of objects drawn from memory relative to within-participant per-scene perception drawing controls, is sensitive to inherent stimulus memorability.

We performed an exploratory follow-up test that modeled the TUT report as a function of memorability, finding that TUT magnitude was similar during the study of high and low memorability scenes ($b = 2.85$, $SE = 2.67$, $95\% \text{ CI} = [-2.39, 8.09]$, $p = .29$). This reinforces the finding from Experiment 1 that exposure to scenes of varying memorability does not influence lapses.

Finally, we tested whether TUT and memorability interacted to affect free recall performance. The results from this model are presented in Table 5. This full model reaffirmed the main effects of memorability and TUT on memory but also yielded a nonsignificant interaction term ($p = .057$) that was near our threshold of .05. We performed a simple slopes analysis to establish what drove this pattern. This showed that TUT during the study of high memorability scenes diminished later memory performance (i.e., the negative slope was significantly different from 0; $b = -0.0028$, $SE = 0.00099$, $95\% \text{ CI} = [-0.0047, -0.00086]$, $t(180) = -2.842$, $p = .0050$), but this was not the case for the study of low memorability scenes ($b = -0.00057$, $SE = 0.00084$, $95\% \text{ CI} = [-0.0022, 0.0011]$, $t(181) = -0.68$, $p = .50$). In other words, TUT during encoding impaired free recall performance for highly memorable scenes, but not forgettable ones (Fig. 5). Given that the interaction term that motivated the analyses here was nonsignificant, this result should be interpreted carefully and requires future replication. Exemplar drawings from multiple different participants are presented in Fig. 6.

Table 5 Linear mixed-effect regression model results when fitting $p(\text{objects drawn from memory})$ as a function of task-unrelated thought (TUT) rating and memorability

Term	beta	SE	95% CI	t	p
(Intercept)	0.53	0.036	[0.45, 0.60]	14.52	<.0001
Memorability condition (low)	-0.137	0.040	[-0.22, -0.058]	-3.41	.00082
TUT rating	-0.0028	0.00098	[-0.0047, -0.00089]	-2.87	.0046
Memorability condition (low):TUT rating	0.0022	0.0012	[-0.00045, 0.0045]	1.92	.057
R ² Marginal	0.08				
R ² Conditional	0.37				

SE = standard error of the mean, CI = confidence interval

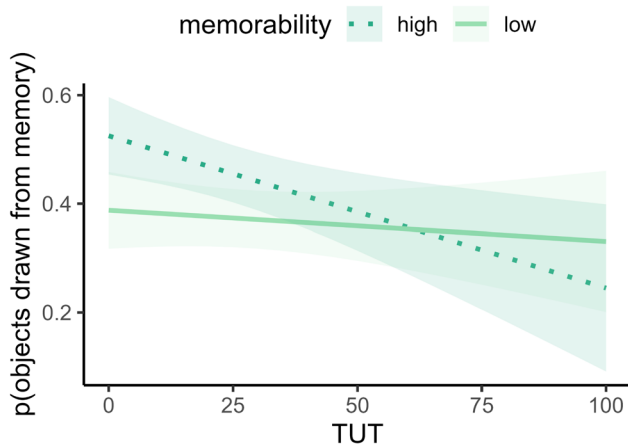


Fig. 5 Linear mixed-effect model predictions for the nonsignificant interaction between task-unrelated thought (TUT) and memorability on the proportion of objects drawn from memory. Follow-up simple slopes tests suggested that TUT during encoding may impair later free recall for high memorability scenes, but not for low memorability scenes

General discussion

Mind-wandering, or task-unrelated thought (TUT), is a ubiquitous experience with consequences for how we process our external world. Researchers have established that it impairs

encoding and reduces the likelihood of retrieval in later memory tests (Blondé et al., 2022). The present study tested whether mind-wandering differentially impacts memory encoding for complex visual scenes that are largely remembered (high memorability) or forgotten (low memorability) across populations. In Experiment 1, we tested memory using an old/new recognition test. In Experiment 2, we tested memory with a novel drawing protocol to assess free recall. The encoding phases in both experiments were interleaved with thought probes to assess fluctuations between on-task thought and TUT. Our results showed that both attentional state during encoding and inherent stimulus memorability are key determinants of recognition performance and the proportion of original scene objects correctly drawn from memory. Interestingly, memorability’s benefit to free recall was present only when the scene was encountered during a period of focused attention.

These findings show how stimulus- and observer-level factors jointly affect memory based on whether it is indexed by recall or recognition (see related perspectives in Jacoby et al., 1993; MacLeod & Kampe, 1996; Mandler et al., 1969; Uner & Roediger III, 2022; Yonelinas et al., 2022; Yonelinas, 2002). Recognition can be supported even when attention at encoding is reduced, since a correct “old” response may rely on a general sense of a prior encounter. Recall, on the other hand, depends more strongly on attention and requires reconstructing detailed information (Jennings &

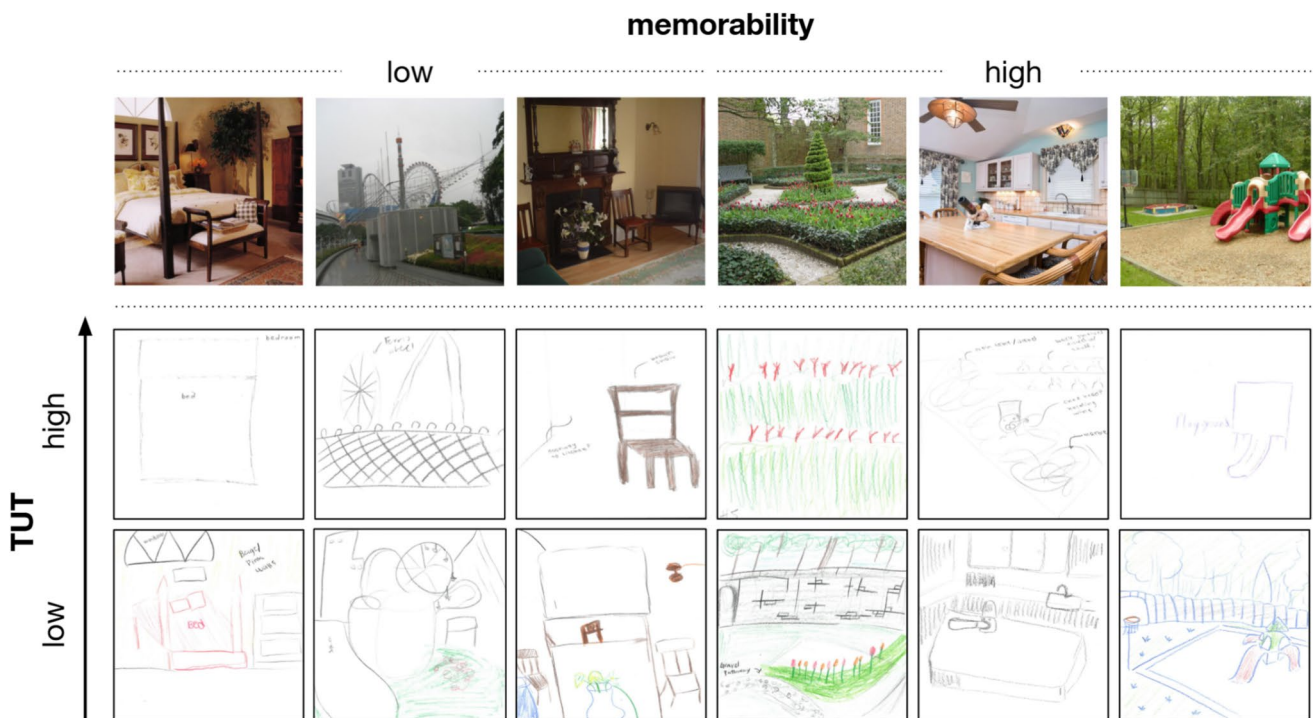


Fig. 6 Examples of memory drawings of the high and low memorability scenes encoded during moments of high and low task-unrelated thought (TUT) in the study phase

Jacoby, 1993; Yonelinas et al., 2022). The drawing protocol introduced by Bainbridge et al. (2019) offers an opportunity to quantify the fidelity of these detailed, recollective representations in visual memory. By integrating this method with thought probes, the present study is the first to assess how transient lapses in attention impair visual recall as measured through memory drawings. We found that TUT during encoding was associated with both poorer recognition and a reduced proportion of object-level details recalled through drawing. This not only replicates prior work showing that mind-wandering disrupts memory formation (Blondé et al., 2022), but extends it in key ways. While the recent review by Blondé et al. synthesized evidence mostly from verbal memory and recognition-based tasks, our results show that the impact of mind-wandering generalizes to visual memory as assessed through free recall. This advances the Bainbridge framework by showing that the memory drawing paradigm is sensitive to internal fluctuations in attention. More broadly, this study demonstrates how experience sampling can be paired with new methods to trace how attentional state shapes specific forms of memory.

Memorability, an intrinsic stimulus attribute that explains recognition performance across populations, has emerged in recent years as a powerful determinant of memory independent from attention (Bainbridge, 2019, 2020). It remains unknown what exactly drives memorability, although evidence suggests that its effect is distinct from priming, depth of encoding, cognitive control, and sustained attentional state (Bainbridge, 2020; Wakeland-Hart et al., 2022). Here, we contributed to other work that showed that lapses in sustained attention, presently operationalized as introspective reports of TUT, impede memory encoding for both high and low memorability scenes. This reinforces conclusions by Wakeland-Hart et al. (2022) that tracking both attentional state during encoding and inherent memorability of the material allows effective prediction of future recognition. We also extend Wakeland-Hart et al. by employing a between-subjects design rather than a within-subjects design; this enabled us to test and reject the possibility that exposure to scenes of varying memorability in Experiment 1 altered lapse rates.

Prior work has shown that memorability does not affect the likelihood of a scene being drawn from memory (Bainbridge et al., 2019). In a part of this previous study, participants viewed a set of complex scenes that varied by memorability before drawing them from memory and completing a recognition test. They found that memorability enhanced recognition but did not significantly impact recall, and that there was no correlation between the number of objects drawn for a given image and its recognition rate. Based on this, we hypothesized that memorability should have no effect on free recall. Contrary to our prediction, we found that freely recalled memories contained more object-level

detail for high memorability scenes than low memorability scenes. It is unsure whether our departure from Bainbridge et al. is due to noise or methodological deviations (e.g., we presented fewer stimuli for more time; our 6 scenes for 15 s each vs. their 30 scenes for 10 s each). A direct replication of Bainbridge et al. would help determine whether the observed divergence is due to design differences, statistical variability, or a genuine effect of memorability on object-level recall that was previously undetected.

Evidence from the verbal domain helps contextualize our positive finding. While memorability robustly predicts recognition when the stimulus is present at test, its role in recall, where the stimulus must be reconstructed, is less clear. Xie et al. (2020) showed that memorable words are retrieved more often and more quickly in a paired-associates verbal recall task. This provides some indication that memorability confers a benefit to reconstructive processes even when the stimulus is perceptually absent. Our results extend Xie et al.'s associative recall finding to the visual domain: memorability can benefit reconstructive processes for object-level detail in complex visual scenes. This supports the view that memorability is a stimulus-intrinsic property that benefits both recognition and recall, though possibly through different mechanisms.

A novel contribution of our work is that attention during encoding appears to explain some benefit of memorability on free recall. We showed that deeper TUT when encoding high memorability scenes was associated with fewer objects drawn from memory, but also that this effect was not present for low memorability scenes. Given that there were no signs of this interaction for recognition performance in Experiment 1, it is possible that TUT during study selectively impairs recollection of memorable scenes by disrupting the attentional processes required for encoding some rich, contextual detail which makes the stimulus memorable. This results in an interaction between TUT and memorability during recall tasks, where the advantage of high memorability is diminished during mind-wandering. However, recognition tasks remain unaffected as they may encourage a retrieval process that relies on familiarity, which is a process resilient to attentional lapses caused by TUT (Smallwood et al., 2007). One reason to be cautious about this result is that it emerged in a post hoc test conducted after identifying an interaction effect that only trended toward our significance threshold. Nonetheless, this result is worth elaborating on in future experimental attempts using drawing to quantify recollection-based visual memory while tracking attention during encoding.

Although our results add to the evidence that mind-wandering impairs encoding of complex visual scenes, we acknowledge recent work suggesting that some forms of learning are actually *enhanced* during mind-wandering (i.e., statistical learning; Simor et al., 2025; Vékony et al., 2025).

This shows the importance of distinguishing between memory systems when studying the consequences of unfocused thought. A recent review by Revsine and Bainbridge (2025) has also highlighted parallels between statistical regularities and memorability — moving forward, the field may benefit from rigorous consideration of all these factors.

Limitations

There are some limitations to consider. One is the use of different TUT probes across experiments: a binary content prompt in Experiment 1 and a 100-point slider in Experiment 2. No study to our knowledge has compared these formats directly, but each has strong predictive validity for performance (Martarelli & Ovalle-Fresa, 2024; see also Kane et al., 2021). Thought-probe design may shift overall TUT levels and TUT–performance coupling, so any comparisons we draw across experiments should be interpreted cautiously. Future work that manipulates probe types within a single experiment would help shed light on this uncertainty.

We originally planned within-subjects tests in Experiment 1 based on recognition and recall tests, but a software bug prevented this analysis. Experiment 2 then supplied new recall data from a separate sample. Because the experiments differed in probe format, trial count, and task structure, we cannot statistically compare interaction magnitudes across experiments; all cross-experiment comparisons are therefore qualitative. A future experiment that collects both outcome measures within-subjects will be needed to determine whether effects truly differ by test type (e.g., whether different types of memory vary in their dependence on attention, or whether the attention–memorability interaction is stronger for one test type than the other).

Another difference between experiments was the change detection and current concerns tasks administered prior to the study phase in Experiment 2, but not Experiment 1. We used these tasks to increase mind-wandering rates since the study phase was shorter. However, there is a possibility that these tasks may have also affected engagement, arousal, or memory in other ways. This again limits comparison between experiments; subsequent work could manipulate task order or include control conditions to isolate these effects.

A separate limitation concerns the low marginal R^2 values for our models. This indicates that our variables, although statistically reliable, only accounted for a small portion of variance in memory, which leaves much unexplained. We view this as a rich opportunity for future work to identify other variables (e.g., strategies, individual differences) and improved methods of attention tracking (Corriveau et al., 2025; Shelat et al., 2024; Schwartz et al., 2025) or memorability estimation (Needell & Bainbridge, 2022) to build more encompassing models of memory.

Conclusion

Our findings illustrate how shifts in attention during encoding interact with the intrinsic memorability of visual stimuli to shape memory outcomes across tasks. By combining assessments of internal cognitive state with complementary measures of recognition and recollection, this work offers an advance in understanding the interplay between observer-driven and stimulus-driven influences on memory. The integration of experience sampling with detailed measures of memory quality opens new avenues to trace the manner in which mind-wandering may shroud the encoding of visual perceptions.

Acknowledgements SS would like to thank Alvin Chen for his guidance with software implementation in this study.

Authors' contributions SS: conceptualization, methodology, data collection, data analysis, data interpretation, software, visualization, writing – original draft, writing – review and editing; BM: data collection, data interpretation, writing – review and editing; CL: data collection, visualization, writing – review and editing; BG: data interpretation, writing – review and editing, supervision; JS: methodology, data interpretation, writing – review and editing, supervision.

Funding SS is supported by a National Science Foundation Graduate Research Fellowship under grant 2139319.

Data availability All data and materials are available in an open-source repository.

Code availability All code is available in an open-source repository.

Declarations

Ethics approval The study obtained IRB approval from the Human Subjects Committee at the University of California, Santa Barbara.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication All included participants provided informed consent for the publication of findings from this study.

Competing interests The authors have no competing interests to declare.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceedings of the 2nd International Symposium on Information Theory* (pp. 267–281). Akademiai Kiado.
- Bainbridge, W. A. (2019). Memorability: How what we see influences what we remember. *Psychology of Learning and Motivation* (vol. 70, pp. 1–27). Academic Press.
- Bainbridge, W. A. (2020). The resiliency of image memorability: A predictor of memory separate from attention and priming. *Neuropsychologia*, 141, Article 107408.

- Bainbridge, W. A. (2022). A tutorial on capturing mental representations through drawing and crowd-sourced scoring. *Behavior Research Methods*, 54(2), 663–675.
- Bainbridge, W. A., Hall, E. H., & Baker, C. I. (2019). Drawings of real-world scenes during free recall reveal detailed object and spatial information in memory. *Nature Communications*, 10(1), 5.
- Bainbridge, W. A., Kwok, W. Y., & Baker, C. I. (2021). Disrupted object-scene semantics boost scene recall but diminish object recall in drawings from memory. *Memory & Cognition*, 49(8), 1568–1582.
- Baird, B., Smallwood, J., Lutz, A., & Schooler, J. W. (2014). The decoupled mind: Mind-wandering disrupts cortical phase-locking to perceptual events. *Journal of Cognitive Neuroscience*, 26(11), 2596–2607.
- Barron, E., Riby, L. M., Greer, J., & Smallwood, J. (2011). Absorbed in thought: The effect of mind wandering on the processing of relevant and irrelevant events. *Psychological Science*, 22(5), 596–601.
- Blondé, P., Girardeau, J. C., Sperduti, M., & Piolino, P. (2022). A wandering mind is a forgetful mind: A systematic review on the influence of mind wandering on episodic memory encoding. *Neuroscience and Biobehavioral Reviews*, 132, 774–792.
- Brosowsky, N. P., Murray, S., Schooler, J. W., & Seli, P. (2021). Attention need not always apply: Mind wandering impedes explicit but not implicit sequence learning. *Cognition*, 209, Article 104530.
- Bullock, T., Elliott, J. C., Serences, J. T., & Giesbrecht, B. (2017). Acute exercise modulates feature-selective responses in human cortex. *Journal Of Cognitive Neuroscience*, 29(4), 605–618.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304.
- Corriveau, A., Rosenberg, M., deBettencourt, M., & deBettencourt, M. T. (2025). Cognitive neuroscience of attention and memory dynamics. *OSF Preprint*. https://doi.org/10.31234/osf.io/n7tma_v1
- Erdelyi, M. H., & Becker, J. (1974). Hypermnnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology*, 6(1), 159–171.
- Fan, J. E., Bainbridge, W. A., Chamberlain, R., & Wammes, J. D. (2023). Drawing as a versatile cognitive tool. *Nature Reviews Psychology*, 2(9), 556–568.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378.
- Hall, E. H., Bainbridge, W. A., & Baker, C. I. (2022). Highly similar and competing visual scenes lead to diminished object but not spatial detail in memory drawings. *Memory*, 30(3), 279–292.
- Hollis, R. B., & Was, C. A. (2016). Mind wandering, control failures, and social media distractions in online learning. *Learning and Instruction*, 42, 104–112.
- Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? *CVPR 2011* (pp. 145–152). IEEE.
- Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, 122(2), 139.
- Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging*, 8(2), 283.
- Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science*, 18(7), 614–621.
- Kane, M. J., Smeekens, B. A., Meier, M. E., Welhaf, M. S., & Phillips, N. E. (2021). Testing the construct validity of competing measurement approaches to probed mind-wandering reports. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01557-x>
- Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, 330(6006), 932–932.
- Kopp, K., D’Mello, S., & Mills, C. (2015). Influencing the occurrence of mind wandering while reading. *Consciousness and Cognition*, 34, 52–62.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281.
- MacLeod, C. M., & Kampe, K. E. (1996). Word frequency effects on recall, recognition, and word fragment completion tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1), 132.
- Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, 98(1), 185.
- Madigan, S., & Rouse, M. (1974). Picture memory and visual-generation processes. *The American Journal of Psychology*, 87(1–2), 151–158.
- Mandler, G., Pearlstone, Z., & Koopmans, H. S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 8(3), 410–423.
- Martarelli, C. S., & Ovalle-Fresa, R. (2024). In sight, out of mind? Disengagement at encoding gradually reduces recall of location. *Quarterly Journal of Experimental Psychology*, 77(1), 42–56.
- Megla, E., Rosenthal, S. R., & Bainbridge, W. A. (2025). Drawings reveal changes in object memory, but not spatial memory, across time. *Cognition*, 254, Article 105988.
- Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Experimentale*, 67(1), 11.
- Needell, C. D., & Bainbridge, W. A. (2022). Embracing new techniques in deep learning for estimating image memorability. *Computational Brain & Behavior*, 5(2), 168–184.
- Quamme, J. R., Yonelinas, A. P., Widaman, K. F., Kroll, N. E., & Sauvé, M. J. (2004). Recall and recognition in mild hypoxia: Using covariance structural modeling to test competing theories of explicit memory. *Neuropsychologia*, 42(5), 672–691.
- Randall, J. G., Oswald, F. L., & Beier, M. E. (2014). Mind-wandering, cognition, and performance: A theory-driven meta-analysis of attention regulation. *Psychological Bulletin*, 140(6), 1411.
- Revsine, C., & Bainbridge, W. A. (2025). Memorability reflects statistical regularities of the environment. *Current Opinion in Neurobiology*, 94, Article 103095.
- Roberts, B. R., Pruijn, J., Bainbridge, W. A., Rosenberg, M. D., & deBettencourt, M. T. (2024). Memory augmentation with an adaptive cognitive interface. *Psychonomic Bulletin & Review*, 32(2), 1–12.
- Rust, N. C., & Mehrpour, V. (2020). Understanding image memorability. *Trends in Cognitive Sciences*, 24(7), 557–568.
- Schwartz, S. T., Yang, H., Xue, A. M., & Wagner, A. D. (2025). Attending to Remember: Recent Advances in Methods and Theory. *Current Directions in Psychological Science*, <https://doi.org/10.1177/09637214251339452>.
- Seli, P., Beaty, R. E., Cheyne, J. A., Smilek, D., Oakman, J., & Schacter, D. L. (2018). How pervasive is mind wandering, really? *Consciousness and Cognition*, 66, 74–78.
- Shelat, S., Schooler, J. W., & Giesbrecht, B. (2024). Predicting attentional lapses using response time speed in continuous performance tasks. *Frontiers in Cognition*, 3, 1460349.

- Shiffrin, R. M. (1973). Visual free recall. *Science*, *180*(4089), 980–982.
- Simor, P., Vékony, T., Farkas, B. C., Szalárdy, O., Bogdány, T., Brezóczki, B., & Németh, D. (2025). Mind wandering during implicit learning is associated with increased periodic EEG activity and improved extraction of hidden probabilistic patterns. *Journal of Neuroscience*. <https://doi.org/10.1523/JNEUROSCI.1421-24.2025>
- Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, *66*(1), 487–518.
- Smallwood, J., McSpadden, M., & Schooler, J. W. (2007). The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic Bulletin & Review*, *14*, 527–533.
- Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: The curious incident of the wandering mind. *Memory & Cognition*, *36*, 1144–1150.
- Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., Carlson, Jean M., & Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS one*, *6*(3), e18298.
- Tabachnick, B., & Brotsky, S. J. (1976). Free recall and complexity of pictorial stimuli. *Memory & Cognition*, *4*, 466–470.
- Uner, O., & Roediger, H. L., III. (2022). Do recall and recognition lead to different retrieval experiences? *The American Journal of Psychology*, *135*(1), 33–43.
- Vékony, T., Farkas, B. C., Brezóczki, B., Mittner, M., Csifcsák, G., Simor, P., & Németh, D. (2025). Mind wandering enhances statistical learning. *iScience*, *28*(2), 111703.
- Wakeland-Hart, C. D., Cao, S. A., deBettencourt, M. T., Bainbridge, W. A., & Rosenberg, M. D. (2022). Predicting visual memory across images and within individuals. *Cognition*, *227*, Article 105201.
- Wammes, J. D., Seli, P., Cheyne, J. A., Boucher, P. O., & Smilek, D. (2016). Mind wandering during lectures II: Relation to academic performance. *Scholarship of Teaching and Learning in Psychology*, *2*(1), 33.
- Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). Sun database: Large-scale scene recognition from abbey to zoo. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 3485–3492). IEEE.
- Xie, W., Bainbridge, W. A., Inati, S. K., Baker, C. I., & Zaghoul, K. A. (2020). Memorability of words in arbitrary verbal associations modulates memory retrieval in the anterior temporal lobe. *Nature Human Behaviour*, *4*(9), 937–948.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517.
- Yonelinas, A. P., Ramey, M. M., Riddell, C., Kahana, M. J., & Wagner, A. D. (2022). Recognition memory: The role of recollection and familiarity. *The Oxford Handbook of Human Memory*. Oxford University Press.
- Zhang, H., Miller, K. F., & Jonides, J. (2024). How does mind-wandering affect distractor suppression? *Visual Cognition*, *32*(9–10), 1–18.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open science and transparency Experiment 1 was preregistered at osf.io/e2sw6. Experiment 2 was not preregistered. All data, code, and materials for both experiments are deposited in the same repository as the preregistration.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.