Journal of Experimental Psychology: Human Perception and Performance

Trade-Offs Between Visual Sampling and Memory in Stable and Changing Worlds

Luzi Xu, Surya Gayet, Andre Sahakian, Jacqueline P. Gottlieb, Stefan Van der Stigchel, and Chris L. E. Paffen

Online First Publication, March 30, 2026. https://dx.doi.org/10.1037/xhp0001395

CITATION

Xu, L., Gayet, S., Sahakian, A., Gottlieb, J. P., Van der Stigchel, S., & Paffen, C. L. E. (2026). Trade-offs between visual sampling and memory in stable and changing worlds. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. https://dx.doi.org/10.1037/xhp0001395

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Human Perception and Performance

ISSN: 0096-1523

https://doi.org/10.1037/xhp0001395

Trade-Offs Between Visual Sampling and Memory in Stable and Changing Worlds

Luzi Xu[1], Surya Gayet[1], Andre Sahakian[1], Jacqueline P. Gottlieb[2, 3, 4], Stefan Van der Stigchel[1],
and Chris L. E. Paffen[1]
[1] Experimental Psychology, Helmholtz Institute, Utrecht University
[2] Department of Neuroscience, Columbia University
[3] Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University
[4] Kavli Institute for Brain Science, Columbia University

In natural behavior, humans make trade-offs between sampling information from the visual environment and relying on memory. As is often the case, observers favor visual sampling when its cost is low (e.g., a sampling only takes a few saccades); but when the sampling cost is artificially increased (e.g., by imposing longer waiting times), they favor visual working memory over visual sampling. Studies investigating this ubiquitous trade-off have neglected the stability in the real world, where repetitive patterns may engage a different memory system: long-term memory (LTM). Two competing hypotheses were derived from previous studies: when stable environments allow the formation of LTM (and sampling costs are relatively low), observers may either (a) continue to rely on visual sampling, or (b) choose to favor memory over visual sampling. We provide evidence for the latter hypothesis, based on a copying task in which participants reproduced an example display containing several colored polygons that either changed or remained stable over consecutive trials. Experiment 1 showed that, when the example display was repeated, the sampling frequency and durations decayed exponentially, eventually disappearing entirely. In Experiments 2 and 3, we replicated the reduction in sampling behavior when only half of the example polygons were repeated. Moreover, participants' recall of repeated items on a surprise memory test predicted this reduction in sampling behavior, confirming the involvement of LTM. Our findings demonstrate that repetitive patterns in stable visual environments make memory use preferable over visual sampling by reducing memory cost through the recruitment of LTM.

Public Significance Statement

When performing everyday tasks, we decide how extensively to rely on memory. For example, we can either memorize all the steps of a recipe at once, or only one step at a time. Recent work suggested that people tend to minimize memory use: when given the option, people only memorize one or two pieces of information; far below typical working memory capacity limits. These studies, however, overlooked the stability of natural environments, which may reduce the effort associated with memory use. In this study, we show that using memory is not always undesirable: it becomes the preferred option when visual information is repeated. Given enough repetitions, people even stop scrutinizing the outside world, and rely entirely on memory, storing far more information than we can hold in working memory. Although many studies explored how repetition strengthens memory, we show that repetitions also determine whether we use memory or not.

Paul E. Dux served as action editor.

Luzi Xu https://orcid.org/0000-0001-8340-5545

Surya Gayet https://orcid.org/0000-0001-9728-1272

Andre Sahakian https://orcid.org/0000-0003-0106-1182

All materials and data are openly available at the project's Open Science Framework page (https://osf.io/6rj9x/). The hypotheses, experimental design, and data analysis plan of Experiment 3 were formally preregistered on the Open Science Framework (OSF) prior to data collection (https://osf.io/5drs8/). This manuscript was first made available as a preprint on OSF in 2024 (https://osf.io/preprints/psyarxiv/e8wbs_v1). The content of the preprint and this version are largely the same, with minor revisions made for improving clarity.

We have no conflicts of interest to disclose. This project was supported by a China Scholarship Council (CSC) scholarship (202106380058).

Luzi Xu served as lead for conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing–original draft, and writing–review and editing. Surya Gayet served as lead for conceptualization, supervision, and writing–review and editing and served in a supporting role for formal analysis. Andre Sahakian contributed equally to formal analysis, investigation, and software and served in a supporting role for writing–review and editing. Jacqueline P. Gottlieb served as lead for conceptualization, supervision, and writing–review and editing. Stefan Van der Stigchel served as lead for conceptualization, supervision, and writing–review and editing. Chris L.E. Paffen served as lead for conceptualization, supervision, and writing–review and editing.

Correspondence concerning this article should be addressed to Luzi Xu, Experimental Psychology, Helmholtz Institute, Utrecht University, Heidelberglaan 1, 3584 CS, Utrecht, The Netherlands. Email: luzixu3@gmail.com

Keywords: memory, working memory, long-term memory, decision making, action

Supplemental materials: https://doi.org/10.1037/xhp0001395.supp

Over the last decades, numerous studies on visual working memory (VWM) have predominantly focused on its maximum capacity. Most of these studies used highly controlled laboratory tasks with fixed set sizes and presentation durations (e.g., Bays and Husain, 2008; Brady et al., 2011; Constantinidis & Klingberg, 2016; Cowan, 2012; Luck & Vogel, 1997; Oberauer et al., 2016; Vogel & Machizawa, 2004). Another body of studies found that, when given the freedom to use memory flexibly (as is the case during most real-world behaviors), observers often do not exhaust their working memory capacity. Instead, they tend to resample sensory inputs multiple times, encoding only one or two items into VWM with each inspection (here labeled samplings, e.g., Draschkow et al., 2021; Grinschgl et al., 2021; Hoogerbrugge et al., 2023; Melnik et al., 2018; Sahakian et al., 2023, 2025; Somai et al., 2020; Xu et al., 2025). The inclination to frequently sample the visual world and store only a limited amount of information internally instead is considered a *cognitive off-loading* strategy (Hu et al., 2019; Risko & Dunn, 2015) that alleviates the energy-consuming cognitive demand of loading information into VWM (Melnik et al., 2018; Somai et al., 2020). Supporting this view, the reliance on sampling decreases and the reliance on memory (VWM) increases when sampling costs are increased (e.g., by having observers wait longer to access the information, or by requiring them to walk for a longer distance toward the information; Ballard et al., 1995; Draschkow et al., 2021; Gray et al., 2006; Inamdar & Pomplun, 2003; Melnik et al., 2018; Sahakian et al., 2023, 2025; Somai et al., 2020; Qing et al., 2025). A key open question, however, concerns the role of long-term memory (LTM) in this process. Although the trade-off between visual sampling and memory (coined "sensory-mnemonic decisions"; Kumle et al., 2025) has mainly been investigated with continuously changing visual streams, the external world is often stable and contains many repetitive patterns. Although it is well-established that repeated exposure engages LTM (Carlisle et al., 2011; Gunseli et al., 2014; Reinhart & Woodman, 2014; Woodman et al., 2013), it is unclear whether the engagement of LTM through repeated exposure influences the trade-off between visual sampling and memory. Here, we investigated this question by having participants perform an unrestrained memory task, using ever-changing versus repeated displays.

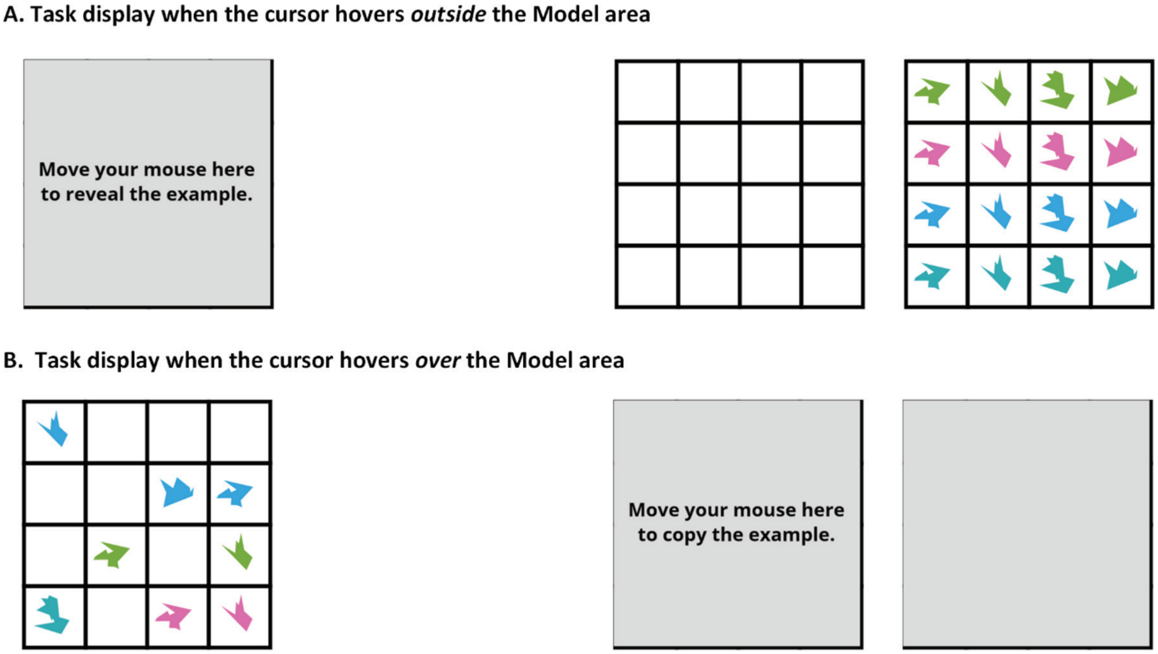
The existing literature supports two conflicting hypotheses regarding the effects of LTM on the trade-off between visual sampling and memory. One possibility is that the availability of repetitive patterns will reduce the reliance on sampling by lowering the costs of memory use. This hypothesis is supported by substantial evidence that visual repetition recruits LTM after only a handful of consecutive trials (Carlisle et al., 2011; Gunseli et al., 2014; Reinhart & Woodman, 2014; Woodman et al., 2013) and, in turn, LTM enhances visual memory (e.g., Brady et al., 2009; Olson et al., 2005; Umemoto et al., 2010), as well as related processes including attention (Chun & Jiang, 1998; Kristjánsson & Campana, 2010) and target recognition (Van Strien et al., 2005). Moreover, repeated exposure reduces neuronal firing rates and hemodynamic responses (Barron et al., 2016; Miller & Desimone, 1994), suggesting that LTM is less effortful and less metabolically costly compared with VWM (Brady et al., 2008, 2024; Greenwald et al., 2003; van Moorselaar et al., 2016).

Alternatively, it is possible that LTM engaged by the repetition of visual patterns will not reduce the reliance on sampling. This hypothesis is supported by abundant evidence that observers tend to rely on sampling rather than memory, even operating well below maximum VWM capacity (Ballard et al., 1995). Moreover, although LTM has behavioral benefits such as larger storage capacity and higher resistance to interference (e.g., Brady et al., 2008, 2024; Greenwald et al., 2003; Shiffrin & Atkinson, 1969; van Moorselaar et al., 2016), LTM still requires some cognitive effort, both in forming LTM representations (e.g., Kafkas & Montaldi, 2011; Paller & Wagner, 2002) and in retrieving information (e.g., Addante et al., 2011). Finally, previous studies have shown that, even when information has been successfully encoded in VWM, observers may still choose to (re)sample from the external world (e.g., Desender et al., 2018; Sahakian et al., 2023). In sum, based on the existing literature, it remains unknown whether the possibility of engaging LTM (in repetitive environments) influences the trade-off between visual sampling and memory.

To investigate this, we used the so-called copying task (Ballard et al., 1995; Sahakian et al., 2023; Somai et al., 2020), in which participants can freely choose between relying on memory or sampling behavior. In this task, participants are tasked to reproduce a configuration of items shown in a model by picking up items from a resource area one by one and placing them at the appropriate locations in the workspace using a cursor (Figure 1). If an item was placed incorrectly, it immediately flew back to the resource area, and participants could move the same or a different item until they correctly placed all items in the model. We refer to a "placement" as the act of lifting and dropping an item, and to a "trial" as the full set of placements by which a participant reproduced the full model in the workspace. Importantly, the model was hidden by default and was revealed only if participants moved the cursor over it. Thus, the number and durations of inspections required to place all items correctly reflect the extent to which participants relied on sampling or memory.

We tested the effects of visual repetition (i.e., presenting the same combination of visual features [i.e., color and shape] at the same locations across trials, or not) in three separate experiments. In Experiment 1, we used a task in which the exact same *model display* was presented across 20 consecutive trials (100% repeat condition), and compared the results with those of a recent study that used a similar task but with a new model display on each trial (Sahakian et al., 2025; i.e., "0% repeat" condition). In Experiment 2, we extended and verified the results by (a) using within-participant measures, (b) a more ecological setting in which only some but not all features of the environment were repeated across trials, and (c) directly testing the

Figure 1
Overview of the Copying Task Display



Note. Participants are required to reproduce the model display (left) by dragging items from the resources grid (right) to the correct position in the workspace (middle). (A) The copying task display when the cursor is *outside* the model display area. (B) The copying task display when the cursor is *inside* the model display area. See the online article for the color version of this figure.

engagement of LTM using a surprise memory test. Finally, in a preregistered Experiment 3, we replicated the results of Experiment 2, while excluding the confound of *resource grid* repetition.

Experiment 1

In Experiment 1, we tested whether the occurrence of visual repetitions across trials would affect the reliance on sampling versus memory. To this aim, we obtained data from a group of participants for whom the entire model remained fixed across 20 trials, and compared it to a recent experiment in which a separate group of participants performed the copying task but using a new model each trial (Sahakian et al., 2025).

Method

Participants

Sixty-one participants (45 women and 16 men, mean age = 21.16 years, $SD = 2.01$) were successfully recruited from our university student population and received course credit for participation. All participants had self-reported normal or corrected-to-normal vision and no color blindness. All experiments and procedures were approved by the Ethics Committee of the university.

We decided a priori to include as many participants as the experimenters could gather within the time frame available for data collection (this experiment was conducted as part of a student thesis), with a minimum of 38 participants. This minimum sample size was determined through a power analysis based on the effect sizes obtained in a copying task study (Sahakian et al., 2023). Using G*Power 3.1 (Faul et al., 2009) for a one-way analysis of variance (ANOVA; omnibus *F test*), suggested that a minimum total sample size of $N = 38$ is required to achieve 80% power ($\alpha = 0.05$, $df = 1$) to detect an effect size of at least $f = 0.48$—which, based on Sahakian et al. (2023), was the effect size for the number of errors, the smallest effect size among all output measures of interest (number of inspections, inspection durations, and errors). A post hoc power analysis showed that the smallest effect size in our data was $f = 0.29$, and the actual sample size that we used achieved ≥95% power ($\alpha = 0.05$, $df = 1$) (one-way ANOVA, omnibus *F test*).

To establish a nonrepetition baseline for comparison with our experiment where models repeated across trials, we leveraged a preexisting data set from Sahakian et al. (2025), a visually nearly identical experiment in which models did not repeat across trials. Thus, we performed a between-subjects comparison, incorporating their data ($N > 100$) as a 0% repetition condition. This approach allowed for an efficient and well-powered initial investigation of the repetition effect. The Sahakian et al. (2025) data set comprised 104 participants (48 women, 53 men, 3 nonbinary; mean age = 35.52, $SD = 11.63$).

Apparatus and Stimuli

The apparatus, stimuli, and procedure in the current experiment were similar to those in Sahakian et al. (2025), allowing for a direct comparison between the two data sets. Both experiments were programmed in the code editor Visual Studio Code (Version 1.75; https://code.visualstudio.com) using the JavaScript libraries jsPsych (Version 7) (de Leeuw, 2015) and Fabric.js (Version 5.2; https://www.fabricjs.com). All participants completed the experiment using a laptop and a computer mouse. Data analysis was performed in MATLAB (R2021a; The MathWorks, Natick, MA) and JASP (Version 0.18.1; Love et al., 2019).

The stimulus sets for both experiments included 400 items consisting of all possible combinations of 20 shapes retrieved from Arnoult (1956), and 20 hues retrieved from the HSLuv color space (https://www.hsluv.org) (Figure 2). The hues were selected to have consistent saturation and luminance (95% and 65%, respectively) and be spaced 18° apart, starting at 0° on the 360-degree color wheel. To construct the resource area on each trial (Figure 1), we randomly selected 16 items from the 400-item set, with the constraints that the items represented four unique shapes and four unique hues, and there was a minimum distance of 54° (three steps) between any two hues. As shown in the right resource grid of Figure 1, items were consistently structured by color and shape, such that each row contained four different shapes in the same color, while each column contained a single shape in four different colors. This canonical structure served as the baseline layout for the resource grid across all experiments. To construct the model display, we randomly selected six items from the resource area. We allowed each item to be used up to two times within a given model display (i.e., trial), so that the correct placement of one item would not exclude this item from being used again, while avoiding excessive repetitions of items.

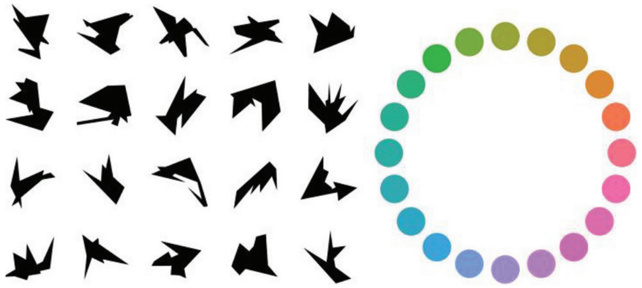
Procedure

After the instructions and two practice trials, participants started the main task. On each trial, participants were shown a model (the *left* grid; Figure 1) and were asked to reproduce it by sequentially dragging each model item from the resource area (the *right* grid) and dropping it onto the workspace (the *middle* grid). As noted earlier, we use the terms "placement" to refer to the positioning of each item and "trial" to refer to the entire sequence of placements that copied the model into the workspace.

For each placement, when a participant picked up an item and hovered it over the workspace, the nearest grid cell of the workspace was highlighted in yellow. If the item was correct (had the correct shape and color for that grid cell), it was pinned to the center of the cell; but if it was incorrect, it flew back to its original location in the resource area. When participants placed all model display items (i.e., six items in Experiment 1; eight items in Experiments 2 and 3) correctly in the workspace, the trial ended, and participants proceeded to the next trial by clicking a button.

Figure 2

The 20 Shapes and 20 Colors That Were Combined to Create the Stimuli in the Experiment



Note. See the online article for the color version of this figure.

The model area was by default covered by a gray square, but the cover was removed when the participant moved the cursor over it, providing a direct measure of the participants' decisions on how many times and how long to inspect the model.

The key experimental manipulation was that, in the 100% repeat trials of the present experiment, we selected six repeated items from the full set of 400 unique combinations (20 shapes × 20 colors) for each participant, and used those six items in the same configuration throughout all 20 trials that the participant performed. Specifically, both the visual features (color and shape) of the six repeated items and their locations in the model grid remained identical across all 20 trials. Contrastingly, in the 0% repeat condition (retrieved from Sahakian et al., 2025), a new set of six items was selected and randomly positioned to create a unique model display on each trial. Aside from the key manipulation of visual repetition, the present procedure and that of Sahakian et al. (2025) differed in minor aspects that are not expected to confound the cross-experiment comparisons. These are detailed and justified in the Interim Discussion of Experiment 1.

Data Analysis

Our four outcome measures were as follows: (a) the number of inspections per trial (i.e., the number of times the model display was uncovered for inspection); (b) individual inspection duration (i.e., the time for which the model was viewed *per inspection*); (c) the number of placement errors (i.e., the number of times an item was incorrectly placed on any empty grid in the workspace); and (d) the average trial duration (i.e., from the start of a trial until the last correct placement). Measures (c) and (d) indicate general performance quality, whereas measures (a) and (b) indicate the reliance on visual sampling (more reliance indicated by more inspections and longer inspection duration).

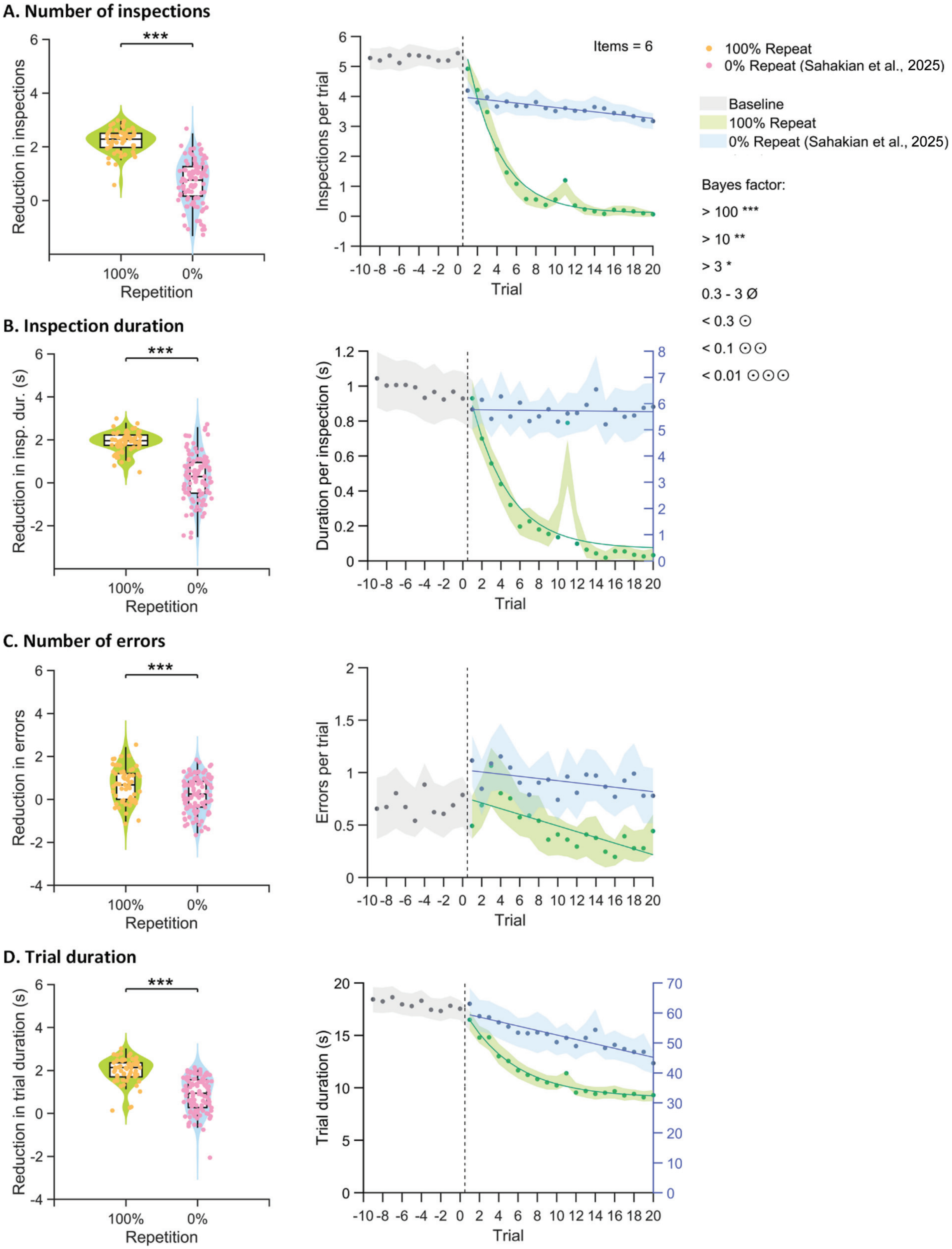
Thus, for the comparison of two different data sets, we first normalized the data for each condition of each participant and computed the reduction in each output measure (i.e., number of inspections, inspection durations) from the first trial to the 20th (last) trial. A greater reduction in output measures (a) and (b) in one condition (e.g., 100% repeat trials) than the other (i.e., 0% repeat trials), for example, indicates a greater reduction in visual samplings caused by the repetition of displays across trials (Figure 3A and 3B, *left*). Besides, to describe the characteristics of the progression of each outcome measure over 20 consecutive trials, we also fit the data of each output measure of each condition to a linear model and an exponential model and tested which is the best fit. More details for data analysis are provided below.

After this normalization procedure, we set out to compare the change in outcome measures over time between experiments, following these steps:

(1) We fit the data for each output measure of each participant across trials to a linear model in which b denotes the slope and a denotes the intercept, see Equation 1. After the fitting procedure, we obtained the predicted (rather than observed) values of each output measure for each trial within participants.

Figure 3

Results of Experiment 1



(*figure continues*)

$$(x) = a + bx \tag{1}$$

(2) We then calculated the reduction in output measures across trials based on the model fits. Specifically, we subtracted the predicted value of the *last* trial (i.e., the 20th trial) from the predicted value of the *first* trial. The difference between the first and last trials indicates the extent of improvement in participants' performance over the course of the experiment.

$$\partial_{SD} = \hat{y}_z^1 - \hat{y}_z^N, \tag{2}$$

where ∂_{SD} reflects the reduction in the outcome measures y (e.g., duration per inspection) over *N* trials expressed in standard deviations, *N* reflects the total number of trials (20), and $\hat{y}_z^n$ reflects the predicted value of the *z*-scored outcome measure *y*, based on the linear fit described in equation (1), for trial *n*. In other words, to quantify the amount of decrease over trials in standardized units, we fit a linear function to the *z*-scored data, and we subtract the predicted value on trial 20 from the predicted value on Trial 1. As the predicted values of the first and last trials are based on a model fit to all trials, the difference between these two "trials" effectively captures the decreasing trend in output measures while being robust to trial-by-trial nuisance variability. Note that this metric does not describe the pace of decrease, only the (relative) amount of decrease over trials.

The reduction (fraction) in the output measures (i.e., the number of inspections, duration per inspection, the number of errors, and trial duration) of 100% repeat condition (of the current experiment) versus 0% repeat condition (Sahakian et al., 2025) were then compared in a Bayesian ANOVA, with model repetition (100% repetition vs. 0% repetition) as between-subjects factor.

Separately from this linear fitting procedure aimed at quantifying the magnitude of change across trials, we also conducted model comparisons to compare whether model repetitions would also qualitatively alter the change in performance across trials. Specifically, we tested whether the change in performance over trials in the 0% repeat and 100% repeat experiments was better explained by a linear or an exponential decrease. To this end, the group means of each output measure were fitted to a linear model as shown in Equation 1, and an exponential decay model as shown in Equation 3, in which *a* denotes the overall magnitude of the function, *b* denotes the decay rate, and *c* denotes the plateau of the decay.

$$\hat{y}(x) = ab^x + c \tag{3}$$

We used Bayesian Information Criterion (BIC) values for model comparison, while accounting for model complexity. The BIC values for the linear models (1) and the exponential models (3) were compared in a bootstrap analysis. In each bootstrap iteration, we extracted a certain number (i.e., 20) of trials from the participants' data with replacements, and we fitted the data to the exponential model and the linear model, respectively. This was done separately for each experiment and each outcome measure. After each of 10,000 iterations, we retrieved BIC values for the two models and conducted a Bayesian paired-sample *t* test to test whether BIC values systematically differed between the linear and exponential models. A lower BIC value indicates better model fit.

For all statistical analyses using Bayesian statistics, we used the JASP software with the default priors (for ANOVAs: fixed effects = 0.5, random effects = 1, covariates = 0.354; for *t* tests: a Cauchy distribution with a scale parameter of 0.707) and consistently setting the seed value to 1 for reproducibility. For each ANOVA, we conducted a so-called analysis of effects across matched models following the approach outlined by Mathôt (2017). This method involves comparing models containing the specific effect of interest to equivalent models without that effect. The *Inclusion Bayes Factor* ($\mathrm{BF_{incl}}$) obtained from this analysis indicates evidence strength, classified as inconclusive (values of < 1), anecdotal (1–3), moderate (3–10), strong (10–30) or very strong (>30) evidence in favor of the presence of the effect in question as opposed to its absence (Kass & Raftery, 1995).

Results

To assess the influence of LTM on the trade-off between sampling and memory, we analyzed how our outcome measures changed across trials in the current experiment, in which the model display was repeated ("100% repeat condition"), versus Sahakian et al. (2025), in which the model was new in each trial ("0% repeat condition"). We compared the overall change in each outcome measure (the difference between the first and 20th trials; Figure 3A–3D, *left*) and the time course of the change across trials (Figure 3A–3D, *right*).

Number of Inspections per Trial

The number of inspections tended to decline across trials and this decline was much larger in the present experiment (yellow, "100% repeat condition") than in the data of Sahakian et al. (2025, pink, "0% repeat condition"). A simple model-based analysis showed that the number of inspections between the first and 20th trials declined in both tasks, but declined much more in the 100% repeat than in the 0% repeat experiment, with a Bayes Factor ($\mathrm{BF_{incl}}$) of 8.29×10^{26}, indicating extremely strong evidence for an effect of repetitions (Figure 3A, *left*). In the 100% repeat condition, participants used, on average, 2.26 fewer inspections on the last

Figure 3 (*continued*)

Note. Panels (A)–(D) depict the results of different outcome measures—number of inspections, duration per inspection, the number of errors, and trial duration, respectively. The present study started with 10 baseline (0% repeat) trials in which the model display was not repeated (in gray), following by 20 experimental trials in which the same model display was repeated across consecutive trials (100% repeat condition; in green). The practice trials (in gray) illustrate the sudden change in each outcome measure (within the same group of participants) as the model display starts repeating after the practice trials (in green). The 20 trials of the 100% repeat condition were contrasted to 20 trials from Sahakian et al. (2025) in which a different model display was used on each trial (0% repeat; in blue). The right panels depict the progression of each outcome measure over 20 consecutive trials, whereas the left panels depict the reduction (fraction) of each outcome measure across these 20 consecutive trials, separately for the 100% repeat and 0% repeat conditions. The extent of the box plots shows the upper/lower quartile, and the whiskers extend to the most extreme data point within 1.5 interquartile range above/below the upper/lower quartile; individual dots in the left panels show individual participant means; color contours outside the box plots show the probability density function of these participant means. Individual dots in the right panels show the group means per trial, with shaded areas showing the 95% confidence interval of the means. Lines show the best fitting linear or exponential functions. See the online article for the color version of this figure.

trial compared with the first trial (*SD* = 0.40), whereas this reduction was only 0.69 (*SD* = 0.80) in the 0% repeat condition.

More detailed analysis of the time-course (Figure 3A, *right*) showed that, in the 100% repeat condition, the number of inspections had a rapid decline that was better fit by an exponential model (Methods, Equation 1; BIC = 0.77, *SD* = 0.17) than a linear model (Equation 3, BIC = 2.80, *SD* = 0.10; model comparison, $\text{BF}_{10} = \infty$). In comparison, in the 0% repeat condition, the number of inspections showed a modest decline that was better fit by a linear model (linear model BIC = −0.51, *SD* = 0.40; exponential model BIC = −0.41, *SD* = 0.41; $\text{BF}_{10} = \infty$). Thus, model repetition causes a faster, exponential decline in the number of inspections compared with the slower, linear decline due to practice effects alone. Note that the "bump" in inspection number at trial 10 was likely due to "optional" self-paced break that participants took between Trials 10 and 11, suggesting that participants needed one trial to "get back" to the task.

The exponential model fits in the 100% repeat condition produced a plateau of 0.108 (*SD* = 0.06) inspections per trial. Thus, participants on average had a very low asymptotic rate of inspection of approximately once every 10 trials and 49 out of 61 participants performed no inspections at all in the last five trials of the task, confirming that, when the model was repeated across trials, participants dramatically reduced or even stopped external sampling, relying nearly exclusively on memory.

Duration per Inspection

Similar to the number of inspections, the duration per inspection also decreased across trials, with this increase being much larger in the 100% repeat versus 0% repeat conditions (Figure 3B). The reduction in inspection duration between the first and last trials was larger in the "100% repeat" versus "0% repeat" condition (1.88, *SD* = 0.47 vs. .27, *SD* = 1.09; $\text{BF}_{\text{incl}} = 1.41 \times 10^{18}$; Figure 3B, *left*). As for the number of inspections, the time course of inspection duration was better fit by the exponential model in the "100% repeat" condition (exponential: BIC = −0.47, *SD* = 0.23; linear: BIC = −0.27, *SD* = 0.21; $\text{BF}_{10} = \infty$) but by the linear model in the 0% repeat condition (linear: BIC = 1.61, *SD* = 0.46; exponential: BIC = 1.72, *SD* = 0.46; $\text{BF}_{10} = \infty$).

Number of Errors per Trial

The number of errors declined from the first to the last trial, and this decline was larger in 100% repeat versus 0% repeat conditions (0.69, *SD* = 0.78 vs. 0.21, *SD* = 0.81; $\text{BF}_{10} = 83.20$; Figure 3C, *left*). The time course of this decline was best captured by a linear model in both conditions (100% repeat condition: linear model BIC = −0.48, *SD* = 0.30, exponential model BIC = −0.37, *SD* = 0.33, $\text{BF}_{10} = \infty$; 0% repeat condition: linear model BIC = −0.40, *SD* = 0.35, exponential model BIC = −0.30, *SD* = 0.37, $\text{BF}_{10} = \infty$). Thus, the reductions in the number and duration of samplings in the 100% repeat condition did not come at the expense of decreased accuracy but instead reflected increased efficiency of memory use.

Trial Duration

Consistent with the decline in the number of inspections, inspection durations, and errors, total trial duration declined from the first to the last trial, and this decline was larger in the 100% repeat than the 0% repeat condition (2.00, *SD* = 0.58 vs. 0.87, *SD* = 0.79; $\text{BF}_{10} = 6.61$; Figure 3D, *left*). The time course of the decline was better fit by an exponential model in the 100% repeat condition (exponential: BIC = 1.59, *SD* = 0.32; linear: BIC = 3.09, *SD* = 0.21; $\text{BF}_{10} = \infty$; Figure 3D, *right*) but was better fit by a linear model in the 0% repeat condition (linear: BIC = 5.23, *SD* = 0.43, exponential: BIC = 5.26, *SD* = 0.48; $\text{BF}_{10} = \infty$).

Interim Discussion

In Experiment 1, we repeated the same model display over 20 consecutive trials to examine how participants traded off the reliance on visual sampling versus memory. We showed that, in comparison to an experiment without model repetition, participants relied more on visual memory, as they inspected the model display less often and for shorter durations and had very low rates of inspection by the end of the task. Despite this reduction in sampling, completion time and error rate decreased, showing that the reduced reliance on sampling was accompanied by higher efficiency.

This conclusion is tempered by several caveats. First, the experiment used a between-subjects design, leaving open the possibility that the results were influenced by individual differences between the two participant groups (although this seems unlikely, given the within-subject drop-off observed after the 10 baseline [0% repeat] trials; Figure 3, gray to green). Second, several methodological differences exist between our study and that of Sahakian et al. (2025). Our experiment included 10 additional baseline (0% repeat) trials and a self-paced break, both absent in theirs. Conversely, a key feature of their design—a delay before revealing the model display—was not incorporated into ours. Critically, these differences do not challenge our core findings. Although they may influence absolute performance levels, they cannot account for the relative changes in our measures of interest across conditions. In fact, the imposed waiting times in Sahakian et al.'s (2025) design are more likely to suppress visual sampling, which works against our hypothesis that the sampling in Sahakian et al. (2025) (0% model repetition) will be higher than ours. Despite this, we emphasize that such cross-experimental methodological differences are best eliminated using a direct, within-subject comparison, and an identical task in the two conditions of interest. Third, we did not directly test LTM, that is, whether participants indeed remembered the repeated items and whether memory accuracy would predict performance on the copying task. Last, in natural settings, environments rarely stay completely unchanged across time, as was the case in our 100% repeat condition, and instead contain mixtures of items that remain constant or change.

In Experiment 2, we addressed these issues by using a new version of the copying task in which half the items in the model display remained fixed throughout all trials, whereas the others changed across trials ("50% repeat"). Importantly, we contrasted performance in this 50% repeat condition within-participant with a 0% repeat condition, in which none of the items were repeated. In addition, we added a surprise memory test for the repeated items at the end of the task to directly test LTM.

Experiment 2

Method

Participants

A new group of 42 participants (21 women and 21 men, mean age = 28.52 years, *SD* = 4.23) was recruited via Prolific (https://www

.prolific.co). As in Experiment 1, all participants self-reported normal or corrected-to-normal vision. In addition, we only recruited participants who (a) indicated to be fluent in English (as the instructions were given in English), (b) had an approval rate higher than 95%, and (c) had not taken part in earlier pilot versions of this experiment.

Based on the sample size of Experiment 1, we aimed for a minimum sample size of 38 participants. A post hoc power analysis (repeated-measures [RM] ANOVA, within-between interaction) indicated that our final sample size of 42 achieved ≥80% power ($\alpha = 0.05$, $df = 1$) for the smallest observed effect size of $f = 0.13$ across all outcome measures (i.e., number of inspections, inspection durations, number of errors, and trial durations).

Apparatus and Stimuli

The task was implemented on the web service Gorilla (https://app.gorilla.sc) using stimuli identical to those in Experiment 1. We highly recommended that participants use a laptop or desktop computer along with a computer mouse to perform the experiment, and all participants confirmed following this recommendation.

Given that participants may have used different screen sizes, we instructed them to run a calibration procedure before the experiment to standardize stimulus size. In the calibration task, participants placed a credit card or any standard-sized card (typically 8.56 cm wide) against the screen and adjusted the size of a rectangle on the screen to match the dimensions of the card. This ensured that the light gray rectangle background (Figure 1) measured approximately 25 cm × 8.5 cm, so that each cell in the copying task grid measured 1 cm × 1 cm.

Procedure

The experimental design and procedures were the same as in Experiment 1 except for three differences. First, after completing instructions and two practice trials, each participant completed two blocks of 24 trials (48 trials in total). In one block ("50% repeat" block), half of the items in the model display were kept constant across all 24 consecutive trials (i.e., "repeated items"), while the locations and features of the other items were randomly selected and varied across trials (i.e., "nonrepeated items"). In the other block ("0% repeat"), the locations and features of all items were randomly selected on each trial from the full set of 400 unique combinations, as in the Sahakian et al.'s (2025) study analyzed in Experiment 1. Block order was counterbalanced across participants, and participants were explicitly informed about the repetition manipulation at the start of each block.

Second, we increased the number of items from six to eight items in each model display, to increase statistical power for comparing between repeated and nonrepeated items in the 50% repeat block.

Third, after participants completed the main task, we added a surprise test to assess their memory for the repeated items. In this test, participants were shown the resource and workspace (but not model) areas, and were asked to select and place all four repeated items from the resources grid at their (repeatedly presented) locations in the workspace. Participants received no feedback, and the memory trial was complete once they placed four items in the workspace, regardless of whether a placement was correct or incorrect.

Data Analysis

As in Experiment 1, we extracted four outcome measures—namely, (a) number of inspections per trial; (b) duration per inspection; (c) the number of placement errors; (d) trial duration—and tested the time course of each measure as a function of trial, using the BIC to compare exponential and linear models while controlling for model complexity.

To these analyses, we added three additional tests specific to the present experiment. First, we compared each outcome measure in the 50% repeat and 0% repeat blocks using a Bayesian RM ANOVA with main factors of block type (50% vs. 0% repeat) and block order (50% repeat block administered first or second) to control for potential order effects. As block order is a nuisance factor that is not of interest to this study, we provided the complete report (with block order and its interaction with model repetition) only in the Supplemental Materials.

Second, we conducted Kendall's Tau rank correlation tests to assess the association between each output measure and performance on the surprise memory test.

Third, to test whether the placement of repeated items (in 50% repeat trials) was prioritized over nonrepeated items, we conducted a Bayesian RM ANOVA with time of placement (i.e., earlier vs. later placements within a trial) as the factor of interest and block order as a nuisance covariate.

Results

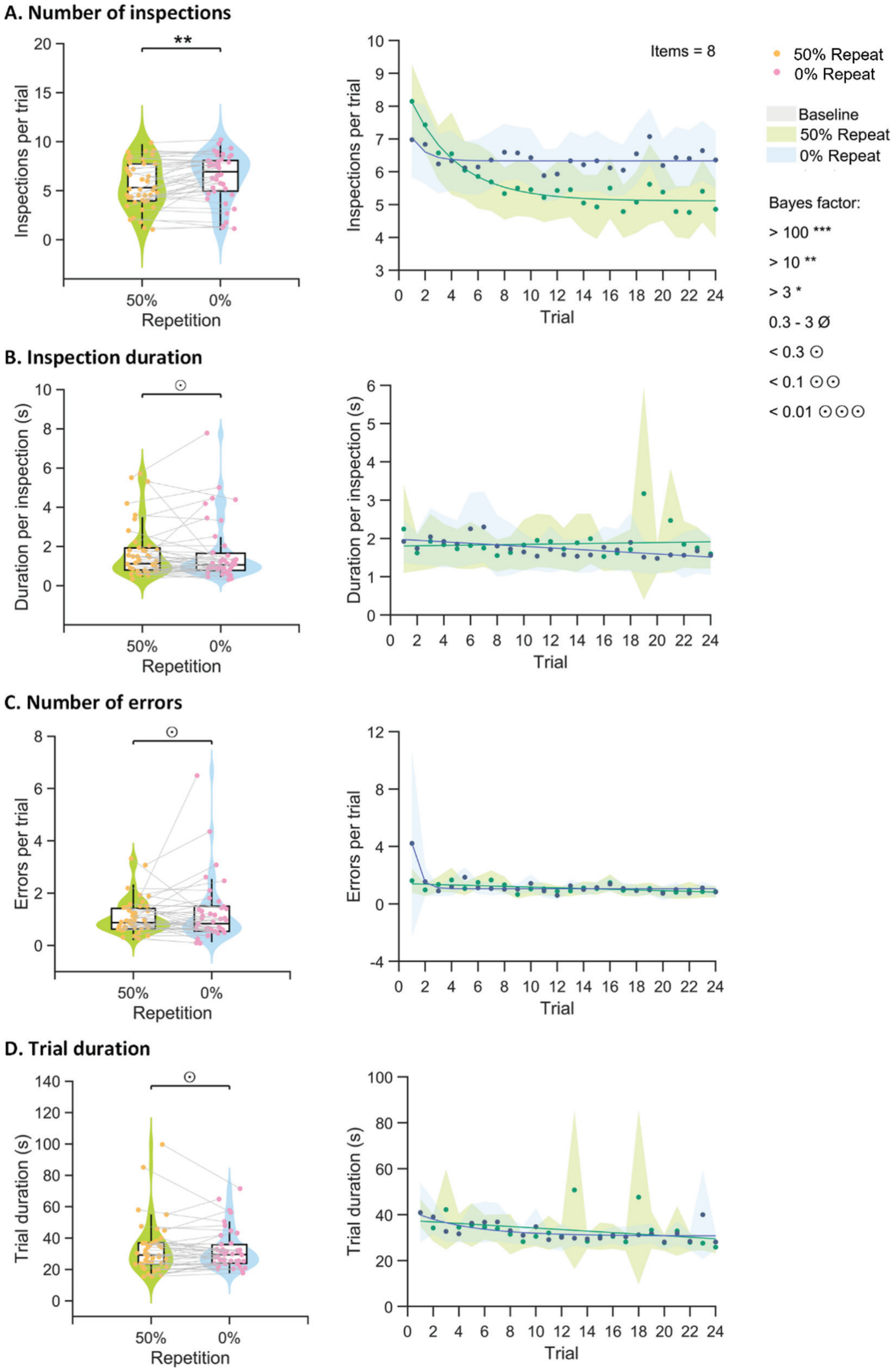
The Comparisons Between 50% Repeat and 0% Repeat Blocks

Focusing first on comparisons between 50% repeat and 0% repeat blocks (without distinguishing between repeated and nonrepeated items), we found only weak or moderate evidence that the blocks differed in terms of duration per inspection (50% repeat: 1.71, $SD = 1.38$; 0% repeat: 1.66, $SD = 1.52$; $BF_{incl} = 0.24$), errors per trial (50% repeat: 1.11, $SD = 0.66$; 0% repeat: 1.21, $SD = 1.19$; $BF_{incl} = 0.26$), and trial duration (50% repeat: 33.37, $SD = 16.82$; 0% repeat: 32.57, $SD = 12.56$; $BF_{incl} = 0.26$; Figure 4). We also found no evidence that the time courses of these measures differed between the conditions across trials. However, the number of inspections per trial was lower in 50% repeat versus 0% repeat blocks (5.62, $SD = 2.39$ vs. 6.38, $SD = 2.45$, respectively), which constituted strong evidence that including repeated items reduced overall sampling rates ($BF_{incl} = 15.05$). No significant main effect of block order nor interaction effects of block order for any output measure were found (all $BF_{incl} < 3$; for details, see Supplemental Materials).

Moreover, the number of inspections decreased across trials; this decline was better fit by an exponential model in both blocks (exponential vs. linear models, 50% repeat blocks, $BF_{10} = \infty$; 0% repeat blocks, $BF_{10} = 2.28$) and reached lower asymptotes in 50% repeat than in 0% repeat blocks (5.09, $SD = 0.41$ vs. 7.26, $SD = 12.17$, $BF_{10} = 9.78$). Thus, maintaining half the items fixed across trials reduced the overall reliance on visual sampling compared with a situation in which all items vary from trial to trial.

Figure 4

Results of Experiment 2



(figure continues)

Memory Test for Repeated Items

We next analyzed performance on the surprise memory test to verify whether participants memorized the repeated items across trials. Here, a correct placement refers to the correct selection of both the feature and location of a repeated item. On average, participants *correctly placed* 2.71 ($SD = 1.54$) out of the four repeated items during this test, amounting to a 67.75% ($SD = 0.38$) correct recall rate. This level was well above chance, whether chance was defined as fully random behavior $\left(3.49 \times 10^{-10} = \left(\frac{1}{16} \times \frac{1}{16}\right) \times \left(\frac{1}{16} \times \frac{1}{15}\right) \times \left(\frac{1}{16} \times \frac{1}{14}\right) \times \left(\frac{1}{16} \times \frac{1}{13}\right); \mathrm{BF}_{10} = 2.93 \times 10^{11}\right)$ or based on a more generous assumption that participants had partial memory for only one location and one feature $\left(\text{colororshape} \left[3.90 \times 10^{-3} = \left(\frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4}\right); \mathrm{BF}_{10} = 2.59 \times 10^{11}\right]\right)$.

To determine whether LTM performance predicted copying task behavior, we examined across-participant correlations between recall rates and copying task outcomes in 50% repeat blocks (Figure 5). Recall rates did not significantly correlate with inspection duration (Kendall's $\tau = 0.05$, $\mathrm{BF}_{10} = 0.22$, 95% confidence interval [CI] = [−0.15, 0.25]; Figure 5B) but correlated negatively with the number of model inspections ($\tau = -0.36$, $\mathrm{BF}_{10} = 55.06$, 95% CI = [−0.54, −0.14]; Figure 5A), number of errors (Kendall's $\tau = -0.40$, $\mathrm{BF}_{10} = 171.22$, 95% CI = [−0.17, −0.57]; Figure 5C) and trial durations (Kendall's $\tau = -0.26$, $\mathrm{BF}_{10} = 3.80$, 95% CI = [−0.05, −0.45]; Figure 5D). These findings indicate that LTM reduced the need for sampling and increased the efficiency of memory use in the copying task.

It should be noted that our sample size ($N = 42$) was relatively limited for estimating correlations across participants. Correlation analyses with such samples are prone to both false positives (spuriously inflated estimates) and false negatives (i.e., failure to detect smaller effects; Schönbrodt & Perugini, 2013). Accordingly, some caution is needed in interpreting the present correlation results (but see Experiment 3 for a preregistered replication of the key findings). The current sample size provides adequate power only to detect effects exceeding approximately $|r| \approx 0.28$ ($\approx$ |Kendall's $\tau| \approx 0.19$). Although our key prediction—the correlation between recall rate and sampling frequency (Kendall's $\tau = -0.36$; Pearson's $r \approx 0.54$)—was well above this threshold, nonsignificant correlations (e.g., recall rate vs. inspection duration) may reflect insufficient power to detect more modest effects rather than the true absence of such relationships.

Repeated Items Were Placed Earlier in a Trial

Our task design does not allow us to compare sampling behavior for repeated and nonrepeated items separately, as each sampling revealed the entire display and we could not identify the specific item that a participant was viewing during an inspection. Within the 50% repeat block, however, we could analyze the order in which participants chose to place repeated and nonrepeated items to gain insights about their strategy. This analysis can distinguish between two opposite hypotheses that are grounded in the literature. On the one hand, memory interference and decay increase over the course of a trial and affect VWM more than LTM (van Moorselaar et al., 2016; Woodman et al., 2013), suggesting that participants may tend to place nonrepeated items first to protect them from these detrimental effects. On the contrary, Sahakian et al. (2023) provide evidence that, when performing the copying task, participants start with items of higher certainty, suggesting that they will place repeated items first.

To distinguish between these hypotheses, we separately analyzed the earlier and later stages of a trial and examined in which stage participants tended to place more repeated items (out of the four repeated and 4 nonrepeated items in 50% repeat blocks). The colormap in Figure 6A shows the percentage of participants who placed a repeated item (rather than a nonrepeated item) in each of the eight correct placements within a trial (*x*-axis), and each of the 24 trials of the 50% repeat block (*y*-axis). Across placements, the color progresses from yellow (more participants placing repeated items) to blue (more participants placing nonrepeated items), showing that participants tended to place the repeated items first.

To formally test this, we binned the 24 trials into four bins of six trials, and we divided the eight correct placements per trial into early and late bins (the first and last four correct placements, respectively). For each of these bins (each comprising four correct placements and six trials), we computed the percentage of repeated items placed per participant. A Bayesian RM analysis confirmed that the fraction of repeated items was much higher in the early versus the later placements ($\mathrm{BF}_{\mathrm{incl}} = 59.52$). Moreover, this tendency was more pronounced in the later versus earlier trial bins (Figure 6A), as would be expected if it reflected LTM use. This was confirmed by a significant interaction between trial bin and placement bin ($\mathrm{BF}_{\mathrm{incl}} = 5.52$).

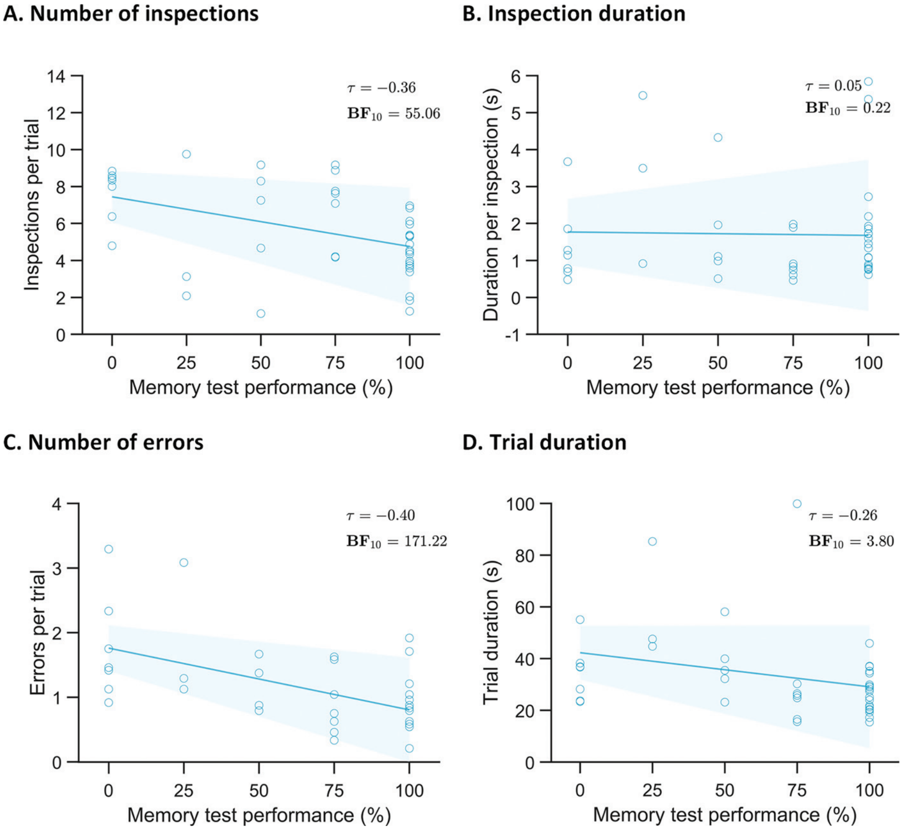
In a final analysis, we again analyzed the correct placements in the 50% repeat block, but now separately for each consecutive *inspection* of the model display (Figure 6B). For each inspection (e.g., the third inspection), we computed the (cumulative) percentage of placements that were repeated items (e.g., out of all items that were correctly placed before the fourth inspection). This allowed us to also analyze placements that participants made before *any* glance at the model display (0 inspection). Notably, participants placed 18.03% ($SD = 0.26$) of repeated items before any model display, a rate that was significantly higher than 0 ($\mathrm{BF}_{\mathrm{incl}} = 354.46$), while for nonrepeated items this percentage was 0.00%. Thus, participants selectively placed some repeated

Figure 4 (*continued*)

Note. Panels (A)–(D) depict the results of different outcome measures—number of inspections, duration per inspection, the number of errors, and trial duration, respectively. The left panels depict the comparison of each output measure between the 50% repeat and 0% repeat conditions, while the right panels depict progression of each outcome measure over consecutive trials. The extent of the box plots shows the upper/lower quartile, and the whiskers extend to the most extreme data point within 1.5 interquartile range above/below the upper/lower quartile; individual dots in the left panels show individual participant means; color contours outside the box plots show the probability density function of these participant means. Individual dots in the right panels show the group means per trial, with shaded areas showing the 95% confidence interval of the means. Lines show the best fitting linear or exponential functions. See the online article for the color version of this figure.

Figure 5

The Correlation Between Memory Test Performance and Outcome Measures of the Copying Task in the 50% Repeat Blocks of Experiment 2



Note. Panels (A)–(D) show the output measures—number of inspections, duration per inspection, number of errors and trial duration respectively. Circular markers show memory test performances (x-axis) against mean copying task outcome measure (y-axis) of individual participant. The line and shaded area show the linear regression fit across datapoints and the corresponding 95% confidence intervals. See the online article for the color version of this figure.

items purely based on memory, before any glance at the model display.

Interim Discussion

In Experiment 2, we found a decrease in sampling frequency when half of the items in the memory display repeated across trials. That is, the effect of visual repetitions on the trade-off between memory and sampling was replicated within-participants, and in a more natural setting where some but not all stimuli were repeated. In line with Experiment 1, the decrease in sampling was not accompanied by an increase in errors, and therefore, did not reflect a more liberal response tendency (an increased willingness to make errors). In contrast to Experiment 1, however, visual repetitions did not lead to a decline in inspection durations, number of errors, or completion time. One possible explanation for this discrepancy is that only half of the items were repeated in Experiment 2, which may obscure smaller effects. In addition, we found that repetitions of the memory displays not only increased the reliance on memory (vs. sampling) but also changed the prioritization of action: observers tended to place repeated items before nonrepeated items.

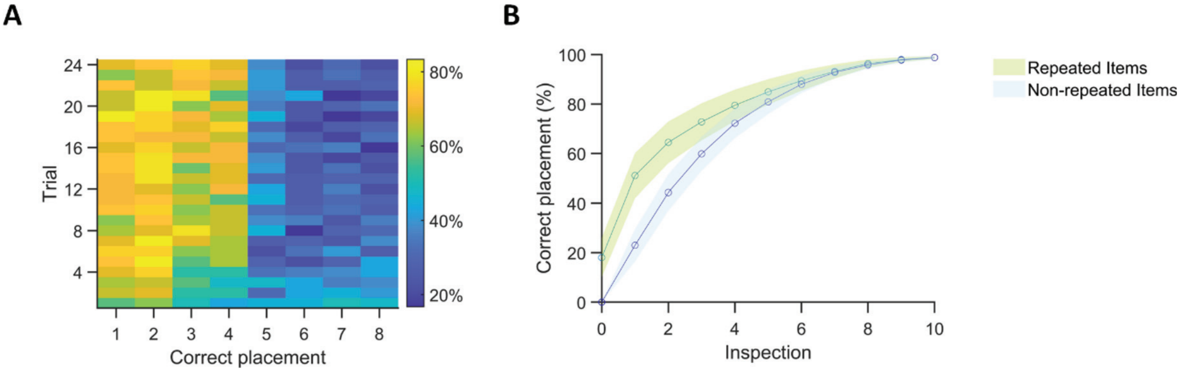
Importantly, we found a direct link between the recruitment of LTM for repeated items and a reduction in sampling for these items. Participants who correctly recalled more items in a surprise memory test about the repeated items, sampled model displays containing these items less frequently during the copying task. We also found that participants tended to (correctly) place repeated items earlier than nonrepeated ones in a trial, at times even before looking at the model display, confirming that they retrieved these items from LTM. Together, these findings provide a conceptual replication of Experiment 1, using within-participant analyses, and provide further support for the hypothesis that visual repetitions reduce the need to sample information from the external world, through the engagement of LTM.

Experiment 3

In Experiments 1 and 2, we found that when the items in the model grid were repeated across trials, participants needed fewer inspections to copy the model display. However, a potential confound exists in the design: when certain items were repeated across trials in the model

Figure 6

Correct Placements Across Trials and Inspections for Repeated vs. Non-repeated Items



Note. (A) The left figure shows the percentage of participants that placed a repeated item (rather than a nonrepeated item) in each of the 24 trials of the "50% repeat" block (*y*-axis) and for each of the eight correct placements in a trial (*x*-axis). (B) This graph depicts the (cumulative) percentage of correctly placed repeated items (in green) and nonrepeated items (in blue), after each consecutive *inspection* of the model display (on the *x*-axis) out of the four items in each condition. The data at Inspection 0 reflects placements that were made before participants viewed the model display. Individual dots show the cumulative group means per inspection (in B). The shaded areas represent the 95% confidence intervals of the means. See the online article for the color version of this figure.

grid (i.e., in the 50% repeat blocks), these items were *also* repeated in the resource grid (and at the exact same location). This repetition of items in the resource grid could have specifically facilitated the placement of repeated items by (a) building up a memory for the layout of the resource grid and/or (b) building up a memory for the motor actions required to pick up and place these repeated items.

The goal of Experiment 3 was to replicate the key findings of Experiments 1 and 2 (i.e., reduced samplings due to model repetition), while excluding the confound of resource grid repetitions. We chose to replicate Experiment 2 because of its superior within-subject design, and the ability to directly compare repeated and nonrepeated items within-trials. We prespecified our key hypothesis (based on Experiment 2) that participants would conduct fewer model inspections (but not at the cost of longer inspections or more errors) in 50%-repeated blocks compared with 0%-repeated blocks, while eliminating any systematic differences in the resource grids between conditions.

The hypotheses, experimental design, and data analysis plan of Experiment 3 were formally preregistered on the Open Science Framework (OSF) before data collection (https://osf.io/5drs8/).

Method

Participants

A new group of 54 participants (27 women and 27 men, mean age = 28.52 years, *SD* = 4.23) were recruited via Prolific, following the same approach as in Experiment 2. The sample size was based on the effect size for the difference in number of inspections between repeat and nonrepeat blocks in Experiment 2 (Cohen's *d* = 0.5), which requires a sample of 54 participants for an experimental power of 95% ($\alpha = 0.05$, two-tailed paired *t* test; G*Power (Version 3.1; Faul et al., 2009).

Apparatus, Stimuli, and Procedure

The experiment used a procedure similar to Experiment 2, including both "50% repeat" and "0% repeat" blocks. The key modification was that stimulus repetitions in the resource grid were now also fully matched between "50% repeat" and "0% repeat" blocks. Specifically, for each participant, a single set of 16 items was selected for the entire experiment (including 50% and 0% repeat blocks), comprising four unique shapes and four unique hues (randomly selected from the same 20 shapes × 20 unique hues stimulus pool) with a minimum separation of 54° (three steps) between any two hues.

Crucially, on every trial and in both conditions, we randomized all item locations within the resource grid by independently shuffling the order of its rows and columns. Thus, on a new trial, a given item could appear on each of 16 locations with equal probability, but each row would still contain four different shapes in the same color, whereas each column contained a single shape in four different colors. This process ensured that participants could not learn the spatial position of any (repeated or nonrepeated) item, nor the motor actions required to reach it, while preserving the canonical structure of the resource grid.

Data Analysis

According to the preregistered research plan, the main goal of Experiment 3 was to test whether participants performed fewer model inspections when half of the model items repeated across trials (compared with blocks with no repetitions). Furthermore, the reduced model inspections should not be accompanied by an increase in model inspection duration, nor by an increase in error, which would indicate a change in response strategy rather than a reduced reliance on visual sampling. Therefore, as in Experiment 2, we conducted Bayesian RM ANOVA comparing all four measures, including (a) number of inspections per trial; (b) duration per inspection; (c) the number of placement errors; (d) trial duration between the 50% repeat and 0% experimental blocks. For consistency with Experiment 2, we also replicated several secondary analyses.

Results

Comparisons Between 50% Repeat and 0% Repeat Blocks

As in Experiment 2, the number of inspections per trial was lower in 50% repeat blocks than in 0% repeat blocks (6.04, *SD* = 2.80 vs. 6.57, *SD* = 2.62; BF_{incl} = 16.42, respectively; Figure 7A, *left*), whereas no differences were found for inspection duration (50% repeat: 2.23, *SD* = 2.06; 0% repeat: 2.17, *SD* = 2.07; BF_{incl} = 0.22), errors per trial (50% repeat: 1.06, *SD* = 1.17; 0% repeat: 1.16, *SD* = 1.46; BF_{incl} = 0.24), or trial duration (50% repeat: 39.21, *SD* = 16.66; 0% repeat: 39.95, *SD* = 15.94; BF_{incl} = 0.23). Furthermore, no main effect or interactions with block order was observed for any output measure (all $BF_{incl} < 3$) except for a significant interaction between order and trial durations ($BF_{incl} > 100$), suggesting a general practice effect whereby the second block was completed faster (for details, see Supplemental Materials).

As in Experiment 2, the number of inspections decreased across trials; this decline was better fit by exponential models in both blocks (exponential vs. linear models, 50% repeat blocks $BF_{10} = \infty$; 0% repeat blocks, $BF_{10} = 2.28 \times 10^{174}$) and reached lower asymptotes in 50% repeat versus 0% repeat blocks (5.09, *SD* = .41 vs. 7.26, *SD* = 12.17, $BF_{10} = 9.78 \times 10^{65}$; Figure 7A, *right*).

These results replicate the findings of Experiment 2, demonstrating that the reduction in sampling behavior can be attributed to visual repetition in the model grid, rather than repetitions in the Resource grid.

Memory Test for Repeated Items

In the surprise memory test, participants *correctly placed* 1.96 (*SD* = 1.49) out of the four repeated items during this test, amounting to a 49.00% (*SD* = 0.37) correct recall rate, which was well above chance ($BF_{10} = 1.27 \times 10^5$) indicating that participants formed LTM representations for repeated items.

Correlation analyses replicated the main finding of Experiment 2, providing strong evidence for a negative correlation between recall rates in the surprise memory test and the number of inspections per trial (Kendall's $\tau = -0.33$, BF_{10} = 64.64, 95% CI = [−0.48, −0.13]; Figure 8A) and thus provide further evidence that LTM reduced the need for sampling. Recall rates, however, were not strongly correlated with inspection durations (Kendall's $\tau = 0.23$, BF_{10} = 3.62, 95% CI = [0.05, 0.40]; Figure 8B), number of errors: (Kendall's $\tau = -0.01$, BF_{10} = 0.18, 95% CI = [−0.19, 0.17]; Figure 8C) or trial durations: Kendall $\tau = 0.02$, BF_{10} = 0.18, 95% CI = [−0.16, 0.20]; Figure 8D), perhaps reflecting the lack of statistical power for medium to small correlations (our sample size [$N = 54$] provided 80% power to detect correlations with absolute magnitudes greater than Kendall's τ = 0.24, equivalent to $r = 0.36$).

Repeated Items Were Placed Earlier in a Trial

As in Experiment 2, in the 50% repeat block, participants tended to place repeated items earlier in a trial than nonrepeated items, and this tendency increased across trials (Figure 9A). Analysis of binned data replicated the results of Experiment 2, producing strong evidence for a main effect of early/late placement (BF_{incl} = 12.96) and an interaction between trial and placement bins (BF_{incl} = 648.44). Participants placed 14.08% (*SD* = 0.25) of repeated items correctly before *any* glance at the model (Figure 9B). This rate was significantly higher than 0 (BF_{incl} = 181.44) and in stark contrast to the 0.00% placement of nonrepeated items, thus showing that participants often placed repeated items purely based on LTM, before even looking at the model display.

Transparency and Openness

Materials (including figures and demos) and data from all three experiments are openly available at the project's Open Science Framework page (https://osf.io/6rj9x/).

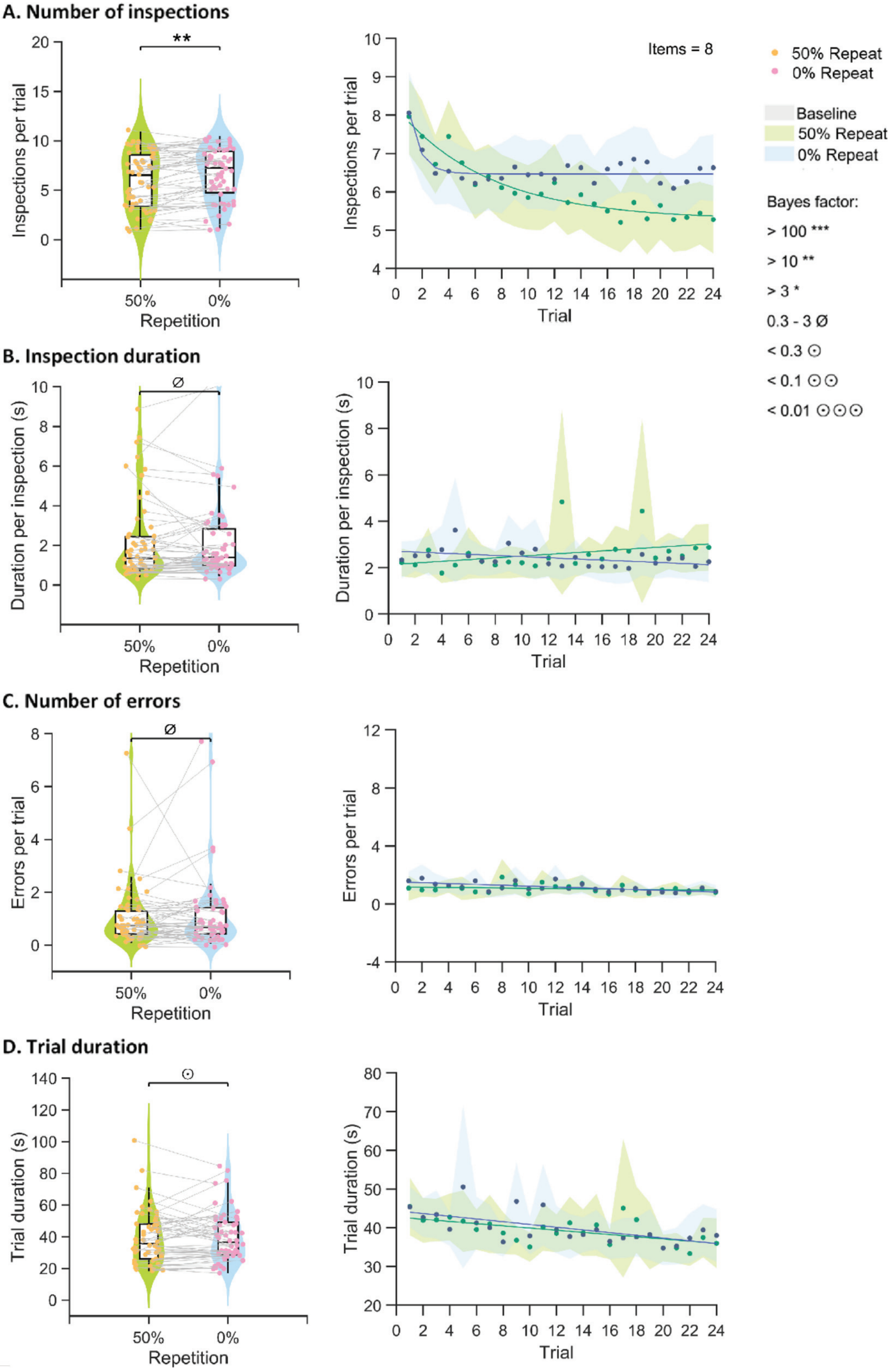
General Discussion

In everyday behavior, we are usually free to decide whether to load visual information into memory for later use, or to obtain (sample) it from the external world once it is needed. In previous literature, the investigations of this "sensory-mnemonic" trade-off have been primarily concerned with continuously changing visual inputs, overlooking the stability of visual streams in real-world settings. To address this, the present study established a case where (all or only part of the) task-relevant information was repeated over time and examined whether this influenced the use of visual memory (vs. visual sampling). To gauge reliance on external visual sampling versus memory, we tracked how often and how long participants viewed (i.e., sampled) freely accessible task-relevant information when performing a copying task. The results of Experiment 1 show that the complete repetition of visual displays across trials led to a steep exponential decay in the frequency and durations of visual sampling; eventually, most participants no longer resorted to visual sampling at all, relying fully on memory. In Experiments 2 and 3, we replicated the reduced sampling frequency in the situation where repeated items were intermixed with nonrepeated items, akin to real-world scenarios. These findings go beyond the well-established effect of visual repetition in improving memory (e.g., Brady et al., 2009; Olson et al., 2005; Umemoto et al., 2010), by showing that visual repetition also guides the *use* of memory, rendering the reliance on visual memory preferable to visual sampling.

The much higher reliance on memory in the model repetition (vs. no repetition) conditions of the present experiment was likely driven by the engagement of LTM. We here refer to LTM as a counterpart to the short-term memory (STM) system that actively holds and manipulates information that is currently in use, and has a brief retention duration (e.g., a few seconds) unless actively rehearsed (Atkinson & Shiffrin, 1968). We propose three lines of evidence that the present task design with visual repetitions indeed engaged LTM: First, it is well-established that repetition of the same visual display recruits LTM after only several successive trials (Carlisle et al., 2011; Gunseli et al., 2014; Reinhart & Woodman, 2014; Woodman et al., 2013). This is indicated by a neural signature reflecting the gradual change from recruiting VWM to LTM: A drop in the contralateral delay activity amplitude (indicating a decreased reliance on VWM) accompanied by a continued increase in recall performance, implicating an increased reliance on LTM (e.g., Woodman et al., 2013). Second, STM has an extremely limited capacity (e.g., ∼four items/features, Cowan, 2001), whereas participants in Experiment 1 loaded up to six complex items that contained multiple features (including color, shape,

Figure 7

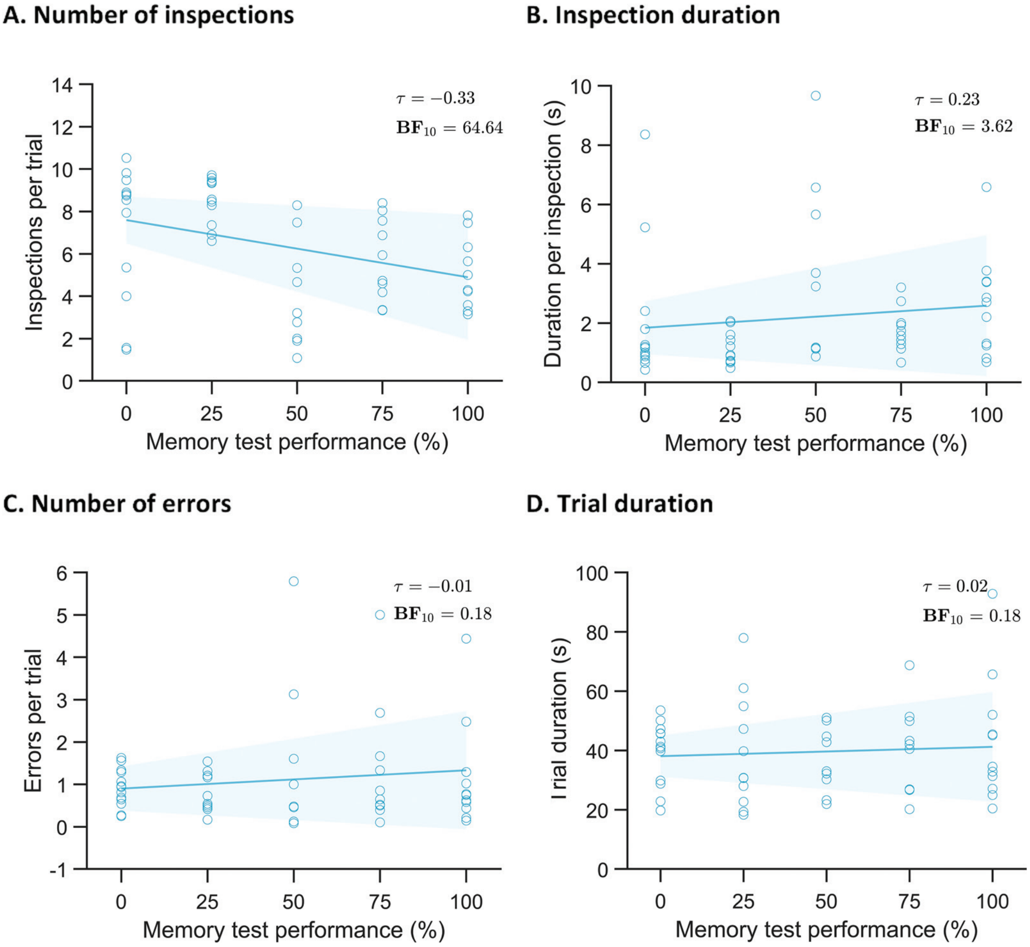
Results of Experiment 3



(figure continues)

Figure 8

The Correlation Between Memory Test Performance and Outcome Measures of the Copying Task in the 50% Repeat Blocks of Experiment 3



Note. Panels (A)–(D) show the output measures—number of inspections, duration per inspection, number of errors and trial duration respectively. Circular markers show memory test performances (x-axis) against mean copying task outcome measure (y-axis) of individual participant. The line and shaded area show the linear regression fit across datapoints and the corresponding 95% confidence intervals. See the online article for the color version of this figure.

and location) into memory, which would have been nearly impossible if only STM was recruited. Third, we provided a direct test of the involvement of LTM in Experiments 2 and 3, where we did not inform the participants beforehand and probed their recall of repeated items in a surprise memory test after the main experiment. Participants could have dropped these items from memory after finishing the copying task, but they still had a high recall rate of repeated items, suggesting that those items were stored in LTM.

It is striking that, when LTM was engaged through repetition, observers abandoned sampling entirely and fully relied on memory. Previous studies using continuously changing visual stimuli found that memory use consistently plateaued at only one to two items per sampling (Draschkow et al., 2021; Sahakian et al., 2023, 2025; Somai et al., 2020) even when sampling was more effortful. In contrast, sampling remained relatively effortless in our experiments, as it only required a few saccades, minor cursor

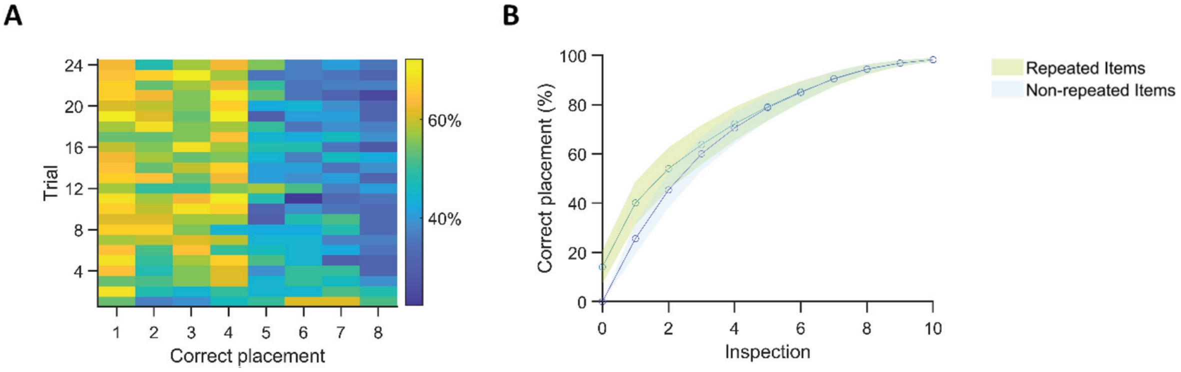
---

Figure 7 (*continued*)

Note. (A)–(D) depict the results of different outcome measures—number of inspections, duration per inspection, the number of errors, and trial duration, respectively. The left panels depict the comparison of each output measure between the 50% repeat and 0% repeat conditions, whereas the right panels depict progression of each outcome measure over consecutive trials. The extent of the box plots shows the upper/lower quartile, and the whiskers extend to the most extreme data point within 1.5 interquartile range above/below the upper/lower quartile. Individual dots in the left panels show individual participant means; color contours outside the box plots show the probability density function of these participant means. Individual dots in the right panels show the group means per trial, with shaded areas showing the 95% confidence interval of the means. Lines show the best fitting linear or exponential functions. See the online article for the color version of this figure.

Figure 9

Correct Placements Across Trials and Inspections for Repeated vs. Non-repeated Items



Note. (A) The left figure shows the percentage of participants that placed a repeated item (rather than a nonrepeated item) in each of the 24 trials in the "50% repeat" block (*y*-axis) and each of the eight correct placements in a trial (*x*-axis). (B) This graph depicts the (cumulative) percentage of correctly placed repeated items (in green) and nonrepeated items (in blue), after each consecutive *inspection* of the model display (on the *x*-axis) out of four items total per condition. The data at Inspection 0 reflects placements that were made before participants viewed the model display. Individual dots show the cumulative group means per inspection (in B). The shaded areas represent the 95% confidence interval of the means. See the online article for the color version of this figure.

movements, and no waiting time (as in Melnik et al., 2018; Somai et al., 2020). Despite the relatively effortless sampling, participants in our experiments chose to rely heavily on memory, loading up to six complex polygons. One might attribute this greater memory reliance solely to LTM's larger capacity compared with VWM. However, this capacity-based explanation is insufficient because it leaves a critical question unanswered: why is VWM consistently underutilized in previous studies? If VWM's capacity ceiling were the primary constraint, we would have observed participants in previous studies maximize their memory use to approach VWM's approximately four-item limit, rather than plateauing at only one to two items (Draschkow et al., 2021; Sahakian et al., 2023). This underutilization indicates that a factor beyond mere capacity governs the strategic choice between sampling and memory.

Why, then, did we observe the strategy of abandoning sampling altogether, rather than sticking to the predominant sampling strategy seen in prior work? We suggest that a critical difference lies in the cognitive costs of the memory system recruited: whereas changing stimuli engage VWM, the repeated displays that we used enable the recruitment of LTM—which makes it less costly (and therefore, more favorable) to rely on memory. This idea follows from a body of work arguing that the encoding, maintenance, and retrieval of LTM representations is less effortful than that of VWM. During the encoding phase, participants can build memory representations of repeated items based on prior visual experience (see the benefits of repetition on memory, e.g., Hintzman, 1976; Umemoto et al., 2010; Young & Bellezza, 1982) instead of building the representations from scratch. Fewer resources are also required to maintain these items (Brady et al., 2024; van Moorselaar et al., 2016; Woodman et al., 2013), and a large set of activated LTM representations can be maintained without interfering with VWM contents (Saltzmann et al., 2024). This suggests that once activated, LTM representations can be readily maintained (and retrieved) with minimal costs to concurrent cognitive operations. After information has been stored in LTM, participants can directly retrieve items from LTM for reuse—the need to resample from the external world is thus eliminated.

It is noteworthy, however, that visual repetition by itself does not necessarily lead to LTM recruitment, nor does LTM necessarily assist tasks that require real-time maintenance of information, such as copying tasks. We propose several key factors that enabled visual repetitions to engage LTM in our copying task. First, prior work suggested that LTM formation generally requires repetition of the same feature-location conjunctions, which allows for the creation of fully bound representations (Logie et al., 2009). The binding of features and locations for repeated items in our study, therefore, may play a crucial role in engaging LTM. Second, previous studies also emphasized interference-free consolidation as a prerequisite for LTM (Logie et al., 2009). Contrary to this, we found that LTM for repeated items was successfully formed even in the presence of distracting nonrepeated items. We propose that this apparent discrepancy can be explained by the fact that the consistent binding of features and locations for repeated items facilitated a process known as chunking. Given that repeated items were bound in both features and locations, participants could store them as single, integrated units, or "chunks." In fact, prior work has uncovered a reliance on chunking during copying tasks, by showing that memory encoding for current targets was impaired when the spatial configuration of surrounding items in the same model was disrupted (Xu et al., 2025). Chunking can reduce the number of discrete items to be maintained (Musfeld et al., 2024a), thereby helping shield the memory representations from interference by concurrently presented distractors (Adam et al., 2024). Furthermore, the nature of our copy task—which required participants to memorize and act on each item of the model grid—likely induced deeper processing of repeated items compared with typical change detection paradigms (e.g., Olson & Jiang, 2004). This deeper encoding facilitates retrieval by creating a more robust memory representation that is less susceptible to interference (Beck & Van Lamsweerde, 2011). Future studies could further address critical boundary conditions, such as whether the repetition effect persists with only one repeated dimension (thereby breaking down binding and preventing chunking), or when trials with repeated and novel grids are intermixed (thereby testing the limits of its resistance to interference).

Interestingly, previous works (Musfeld et al., 2023, 2024b; Ngiam et al., 2019; Souza & Oberauer, 2022) suggest that explicit awareness of repetition may be a prerequisite for LTM to assist real-time maintenance of task-relevant information (Musfeld et al., 2023, 2024b; Ngiam et al., 2019; Souza & Oberauer, 2022), as required in the copying task that we used. Specifically, previous studies suggested that the amount of memorized information can massively transcend VWM capacity limits, only when repetitions are explicitly noted (Adam et al., 2024; Ngiam et al., 2019). In line with this idea, our data provides behavioral evidence strongly suggestive of explicit awareness of visual repetitions. Specifically, even without prior instructions about which items were repeated items, participants correctly recalled nearly half of the four repeated items (Experiment 2: 2.71 items; Experiment 3: 1.96 items). Moreover, we observed that a majority of participants exhibited explicit knowledge of the repetitions, by successfully placing repeated items before any glance of the model grid (Experiment 2: 47.62%; Experiment 3: 37.04%) or, as in Experiment 1 (98.36%), having no inspection of the model at all. This pattern strongly indicates that participants actively retrieved and used their explicit knowledge of the repeated items during the task. That said, the level of LTM reliance observed in our study likely represents a conservative estimate of its full potential. If explicit instructions about the repeated items were provided, we might have observed a more complete shift away from working memory sampling and toward near-perfect LTM recall.

The present study makes several important contributions to the literature on sensory-mnemonic trade-offs: First, we highlight the role of LTM in resolving this trade-off, which has long been neglected by the literature (e.g., Draschkow et al., 2021; Sahakian et al., 2023, 2025). This helps the field to shift from a VWM-centric to a dual-system (VWM and LTM) perspective. Second, we add that the sensory-mnemonic trade-off is not only dependent on the costs of sampling (as extensively investigated before: Draschkow et al., 2021; Sahakian et al., 2023, 2025), but also on the cost of memory use. This possibility had not been explicitly tested before and was only supported by a previous study in which healthy individuals sampled less frequently from the external world than individuals with Korsakoff amnesia (i.e., a patient population with a lower working memory capacity) when doing the same copying task (Böing et al., 2023). We here kept the sampling cost constant and directly tested this hypothesis, showing that the use of memory can become the preferred strategy when the memory cost is reduced through LTM engagement. These results further support theories stating that a continuous decision is made for balancing the costs of memory and sampling (Draschkow et al., 2021; Van der Stigchel, 2020). Third, our findings reveal that the sensory-mnemonic trade-off extends beyond the decision to sample to strategically guide the temporal order of actions. Specifically, we observed a clear prioritization of repeated items over nonrepeated items for action. Even more so, in Experiments 2 and 3, we showed that participants started acting on repeated items, even before the first inspection of the visual display. This initial placement of repeated items may reflect that repeated items in LTM are more accessible than nonrepeated items, and that participants are more certain about these items (see Sahakian et al., 2023). Moreover, placing repeated items first could reflect a cognitive off-loading strategy (Grinschgl et al., 2021); upon completing the repeated items, observers can actively exclude them from their task goals, freeing up attentional resources and streamlining subsequent sampling.

In conclusion, we show that, in a highly predictable and structured visual environment, akin to the real world, the cost of memory becomes lower because of repetition, encouraging participants to rely more on visual memory than on visual sampling. We suggest that the engagement of LTM over STM allows observers to store multiple items in memory at a lower cost, which reduces and may even eliminate the need for visual sampling altogether.

References

Adam, K. C., Zhao, C., & Vogel, E. K. (2024). Behavioral signatures of the rapid recruitment of long-term memory to overcome working memory capacity limits. *Memory and Cognition*, *52*(8), 1816–1832. https://doi.org/10.3758/s13421-024-01566-z

Addante, R. J., Watrous, A. J., Yonelinas, A. P., Ekstrom, A. D., & Ranganath, C. (2011). Prestimulus theta activity predicts correct source memory retrieval. *Proceedings of the National Academy of Sciences, USA*, *108*(26), 10702–10707. https://doi.org/10.1073/pnas.1014528108

Arnoult, M. D. (1956). Familiarity and recognition of nonsense shapes. *Journal of Experimental Psychology*, *51*(4), 269–276. https://doi.org/10.1037/h0047772

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), The psychology of learning and motivation: Advances in research and theory (Vol. 2, pp. 89–195). Academic Press.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*(1), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Barron, H. C., Garvert, M. M., & Behrens, T. E. (2016). Repetition suppression: A means to index neural representations using BOLD? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1705), Article 20150355. https://doi.org/10.1098/rstb.2015.0355

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854. https://doi.org/10.1126/science.1158023

Beck, M. R., & Van Lamsweerde, A. E. (2011). Accessing long-term memory representations during visual change detection. *Memory and Cognition*, *39*(3), 433–446. https://doi.org/10.3758/s13421-010-0033-4

Böing, S., Ten Brink, A. F., Hoogerbrugge, A. J., Oudman, E., Postma, A., Nijboer, T. C., & Van der Stigchel, S. (2023). Eye movements as proxy for visual working memory usage: Increased reliance on the external world in Korsakoff syndrome. *Journal of Clinical Medicine*, *12*(11), Article 3630. https://doi.org/10.3390/jcm12113630

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*(4), 487–502. https://doi.org/10.1037/a0016797

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

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

Brady, T. F., Robinson, M. M., & Williams, J. R. (2024). Noisy and hierarchical visual memory across timescales. *Nature Reviews Psychology*, *3*, 147–163. https://doi.org/10.1038/s44159-024-00276-2

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *Journal of Neuroscience*, *31*(25), 9315–9322. https://doi.org/10.1523/JNEUROSCI.1097-11.2011

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

Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience*, *17*(7), 438–449. https://doi.org/10.1038/nrn.2016.43

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2012). *Working memory capacity*. Psychology Press.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*(5), 761–778. https://doi.org/10.1177/0956797617744771

Draschkow, D., Kallmayer, M., & Nobre, A. C. (2021). When natural behavior engages working memory. *Current Biology*, *31*(4), 869–874.e5. https://doi.org/10.1016/j.cub.2020.11.013

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Gray, W. D., Sims, C. R., Fu, W. T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113*(3), 461–482. https://doi.org/10.1037/0033-295X.113.3.461

Greenwald, A. G., Abrams, R. L., Naccache, L., & Dehaene, S. (2003). Long-term semantic memory versus contextual memory in unconscious number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 235–247. https://doi.org/10.1037/0278-7393.29.2.235

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, *74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060

Gunseli, E., Olivers, C. N., & Meeter, M. (2014). Effects of search difficulty on the selection, maintenance, and learning of attentional templates. *Journal of Cognitive Neuroscience*, *26*(9), 2042–2054. https://doi.org/10.1162/jocn_a_00600

Hintzman, D. L. (1976). Repetition and memory. *Psychology of Learning and Motivation*, *10*, 47–91. https://doi.org/10.1016/S0079-7421(08)60464-8

Hoogerbrugge, A. J., Strauch, C., Nijboer, T. C. W., & Van der Stigchel, S. (2023). Don't hide the instruction manual: A dynamic trade-off between using internal and external templates during visual search. *Journal of Vision*, *23*(7), Article 14. https://doi.org/10.1167/jov.23.7.14

Hu, X., Luo, L., & Fleming, S. M. (2019). A role for metamemory in cognitive offloading. *Cognition*, *193*, Article 104012. https://doi.org/10.1016/j.cognition.2019.104012

Inamdar, S., & Pomplun, M. (2003). *Comparative search reveals the tradeoff between eye movements and working memory use in visual tasks. In* Proceedings of the annual meeting of the cognitive science society (Vol. 25, No. 25). https://escholarship.org/uc/item/4k01x26b

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. https://doi.org/10.1080/17470218.2011.588335

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kristjánsson, Á., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception, & Psychophysics*, *72*(1), 5–18. https://doi.org/10.3758/APP.72.1.5

Kumle, L., Nobre, A. C., & Draschkow, D. (2025). Sensorimnemonic decisions: choosing memories versus sensory information. *Trends in Cognitive Sciences*, *29*(4), 311–313. https://doi.org/10.1016/j.tics.2024.12.010

Logie, R. H., Brockmole, J. R., & Vandenbroucke, A. R. (2009). Bound feature combinations in visual short-term memory are fragile but influence long-term learning. *Visual Cognition*, *17*(1–2), 160–179. https://doi.org/10.1080/13506280802228411

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*(2), 1–17. https://doi.org/10.18637/jss.v088.i02

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Mathôt, S. (2017). *Bayes like a baws: Interpreting Bayesian repeated measures in JASP*. Retrieved July 26, 2022, from https://www.cogsci.nl/blog/interpreting-bayesian-repeated-measures-in-jasp.

Melnik, A., Schüler, F., Rothkopf, C. A., & König, P. (2018). The world as an external memory: The price of saccades in a sensorimotor task. *Frontiers in Behavioral Neuroscience*, *12*, Article 253. https://doi.org/10.3389/fnbeh.2018.00253

Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science*, *263*(5146), 520–522. https://doi.org/10.1126/science.8290960

Musfeld, P., Dutli, J., Oberauer, K., & Bartsch, L. M. (2024a). Grouping in working memory guides chunk formation in long-term memory: Evidence from the Hebb effect. *Cognition*, *248*, Article 105795. https://doi.org/10.1016/j.cognition.2024.105795

Musfeld, P., Souza, A. S., & Oberauer, K. (2023). Repetition learning is neither a continuous nor an implicit process. *Proceedings of the National Academy of Sciences, USA*, *120*(16), Article e2218042120. https://doi.org/10.1073/pnas.2218042120

Musfeld, P., Souza, A. S., & Oberauer, K. (2024b). Testing expectations and retrieval practice modulate repetition learning of visuospatial arrays. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(5), 740–758. https://doi.org/10.1037/xlm0001298

Ngiam, W. X., Brissenden, J. A., & Awh, E. (2019). "Memory compression" effects in visual working memory are contingent on explicit long-term memory. *Journal of Experimental Psychology: General*, *148*(8), 1373–1385. https://doi.org/10.1037/xge0000649

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799. https://doi.org/10.1037/bul0000046

Olson, I. R., & Jiang, Y. (2004). Visual short-term memory is not improved by training. *Memory and Cognition*, *32*(8), 1326–1332. https://doi.org/10.3758/BF03206323

Olson, I. R., Jiang, Y., & Moore, K. S. (2005). Associative learning improves visual working memory performance. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 889–900. https://doi.org/10.1037/0096-1523.31.5.889

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, *6*(2), 93–102. https://doi.org/10.1016/S1364-6613(00)01845-3

Qing, T., Strauch, C., Van Maanen, L., & Van der Stigchel, S. (2025). Shifting reliance between the internal and external world: A metaanalysis on visual-working memory use. *Psychonomic Bulletin and Review*, *32*(3), 1118–1130. https://doi.org/10.3758/s13423-024-02623-z

Reinhart, R. M., & Woodman, G. F. (2014). High stakes trigger the use of multiple memories to enhance the control of attention. *Cerebral Cortex*, *24*(8), 2022–2035. https://doi.org/10.1093/cercor/bht057

Risko, E. F., & Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and Cognition*, *36*, 61–74. https://doi.org/10.1016/j.concog.2015.05.014

Sahakian, A., Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (2023). Mountains of memory in a sea of uncertainty: Sampling the external world despite useful information in visual working memory. *Cognition*, *234*, Article 105381. https://doi.org/10.1016/j.cognition.2023.105381

Sahakian, A., Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (2025). Action consequences guide the use of visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *51*(1), 4–13. https://doi.org/10.1037/xlm0001326

Saltzmann, S. M., Eich, B., Moen, K. C., & Beck, M. R. (2024). Activated long-term memory and visual working memory during hybrid visual search: Effects on target memory search and distractor memory. *Memory and Cognition*, *52*(8), 2156–2171. https://doi.org/10.3758/s13421-024-01556-1

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review*, *76*(2), 179–193. https://doi.org/10.1037/h0027277

Somai, R. S., Schut, M. J., & Van der Stigchel, S. (2020). Evidence for the world as an external memory: A trade-off between internal and external visual memory storage. *Cortex*, *122*, 108–114. https://doi.org/10.1016/j.cortex.2018.12.017

Souza, A. S., & Oberauer, K. (2022). Promoting visual long-term memories: When do we learn from repetitions of visuospatial arrays? *Journal of Experimental Psychology: General*, *151*(12), 3114–3133. https://doi.org/10.1037/xge0001236

Umemoto, A., Scolari, M., Vogel, E. K., & Awh, E. (2010). Statistical learning induces discrete shifts in the allocation of working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1419–1429. https://doi.org/10.1037/a0019324

Van der Stigchel, S. (2020). An embodied account of visual working memory. *Visual Cognition*, *28*(5–8), 414–419. https://doi.org/10.1080/13506285.2020.1742827

Van Moorselaar, D., Theeuwes, J., & Olivers, C. N. (2016). Learning changes the attentional status of prospective memories. *Psychonomic Bulletin and Review*, *23*(5), 1483–1490. https://doi.org/10.3758/s13423-016-1008-7

Van Strien, J. W., Hagenbeek, R. E., Stam, C. J., Rombouts, S. A., & Barkhof, F. (2005). Changes in brain electrical activity during extended continuous word recognition. *NeuroImage*, *26*(3), 952–959. https://doi.org/10.1016/j.neuroimage.2005.03.003

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751. https://doi.org/10.1038/nature02447

Woodman, G. F., Carlisle, N. B., & Reinhart, R. M. (2013). Where do we store the memory representations that guide attention? *Journal of Vision*, *13*(3), Article 1. https://doi.org/10.1167/13.3.1

Xu, L., Sahakian, A., Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (2025). Latent memory traces for prospective items in visual working memory. *Journal of Experimental Psychology. Human Perception and Performance*, *51*(2), 164–177. https://doi.org/10.1037/xhp0001257

Young, D. R., & Bellezza, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(6), 545–559. https://doi.org/10.1037/0278-7393.8.6.545`

Received February 11, 2025
Revision received October 6, 2025
Accepted November 12, 2025 ■