AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2022, Vol. 151, No. 12, 3114–3133
https://doi.org/10.1037/xge0001236

Promoting Visual Long-Term Memories: When Do We Learn From Repetitions of Visuospatial Arrays?

Alessandra S. Souza[1, 2] and Klaus Oberauer[2]
[1] Faculty of Psychology and Education Sciences, University of Porto
[2] Department of Psychology, University of Zurich

Repeated exposure is assumed to promote long-term learning. This is demonstrated by the so-called "Hebb-effect": when short lists of verbal or spatial materials are presented sequentially for an immediate serial recall test, recall improves with list repetition. This repetition benefit, however, is not ubiquitous. Previous studies found little or no performance improvement for repetitions of visuospatial arrays (e.g., arrays of colored squares). Across eight experiments with college students and Prolific samples, we investigated which factors promote visuospatial learning by testing all combinations of variables distinguishing between visual-array tasks (brief + simultaneous presentation + a single recognition test) and tasks showing the Hebb effect (slow + sequential presentation + recall test probing all items). Participants profited from repetitions when all items were tested with a recall procedure, but not if the test consisted of recognition. Hence, the key to promote long-term learning is to recall all of the memorized information over the short-term.

Keywords: Hebb learning, visuospatial arrays, working memory, recall, recognition

How do we learn *what* is *where* around us? Learning of visuospatial configurations arguably depends both on working memory and long-term memory. Working memory stores the small subset of information needed for ongoing processing. Long-term memory, conversely, holds representations for long periods of time. Although distinct, working memory and long-term memory are interlinked. For example, information attended to and retained longer in working memory is better recalled in long-term memory (Hartshorne & Makovski, 2019; Loaiza & McCabe, 2012; Souza & Oberauer, 2017; Zanto et al., 2016). Conversely, long-term memory knowledge increases working memory performance: memory span for words is larger than for nonwords (Hulme et al., 1991), and for familiar than unfamiliar visual items (Xie & Zhang, 2017; but see Wood & Simons, 2017).

One classical demonstration of the interdependency of working memory and long-term memory is the *Hebb-repetition effect* (Hebb, 1961): When a memory list is presented for serial recall, performance is limited by working memory capacity. When one list is repeated amid nonrepeated lists, recall of the repeated list increases over about eight repetitions, but recall of nonrepeated lists remain unchanged. Hence, as long-term memory traces accumulate, serial-recall performance increases. This effect has been replicated using verbal materials such as digits, letters, syllables, or words (Hebb, 1961; Oberauer et al., 2015; Page & Norris, 2009; Szmalec et al., 2012) as well as visual stimuli such as faces (Horton et al., 2008; Johnson et al., 2017; Johnson & Miles, 2019a, 2019b) and spatial locations (Couture & Tremblay, 2006; Sukegawa et al., 2019; Tremblay & Saint-Aubin, 2009).

However, the repetition benefit is not ubiquitous. A few studies have shown that repetitions do not promote learning of visuospatial arrays (Fukuda & Vogel, 2019; Logie et al., 2009; Olson et al., 2005; Olson & Jiang, 2004; Shimi & Logie, 2019). In visuospatial array tasks, participants have to learn which visual features appeared in which location. In this type of task, a visuospatial array with multiple elements is presented simultaneously for encoding, typically followed by a recognition test (change-detection) to probe memory. In Olson and Jiang (2004), change-detection of arrays of square locations did not improve over 24 repetitions. In Logie et al. (2009), change-detection of colored shapes did not improve after the same array was repeated 60 times in a row (with no other intervening arrays), and Shimi and Logie (2019) observed only modest improvements after the same array was repeated more than 30 times in a row. Fukuda and Vogel

This article was published Online First May 23, 2022.
Alessandra S. Souza https://orcid.org/0000-0002-1057-8426
This research was supported by a grant from the Swiss National Science Foundation to A. S. Souza (grant 100019_169302) and to K. Oberauer (grant 100014_149193). We thank the students in the Introductory Experimental Class of the University of Zurich in the Spring Semester of 2017 for working on the data collection of Experiments 1 to 5 as part of their course assignment, and Clara Overkott and Andra Arnicane for organizing the data collection. We also thank Philipp Musfeld for the help in programming the online experiments (Experiments 7 and 8). Parts of this data were presented in the 59th Annual Meeting of the Psychonomic Society.
Study preregistration (Experiments 1 to 5), materials, data, and analysis scripts are available at: https://osf.io/65bmd/
Correspondence concerning this article should be addressed to Alessandra S. Souza, Faculty of Psychology and Educational Sciences, University of Porto, Rua Alfredo Allen s/n, 4200-135 Porto, Portugal. Email: alessandra@fpce.up.pt

(2019) repeated arrays of colored squares 30 times each, but performance for repeated and novel unique arrays remained similar in a final test implemented at the end of the study. These results suggest that simple repetition is not enough to promote learning of visuospatial arrays. Yet it is unclear what sets visuospatial memory apart from other studies that found credible learning for repeated lists of verbal, visual, or even spatial information.

Why Do People Fail to Learn Visuospatial Arrays?

One possibility is that visuospatial arrays are genuinely difficult to learn. Alternatively, some feature of the commonly used visuospatial working memory tasks may either prevent learning, or the use of acquired knowledge. This is likely given that there are many procedural differences between the typical visuospatial working memory tasks and the Hebb-repetition paradigm. Typical visuospatial-arrays tasks comprise multiple items presented briefly (e.g., 150 ms) at once, whereas the Hebb paradigm is a standard serial-recall task: Memoranda are presented sequentially and slowly (usually 1 s/item). Sequential presentation allows attention to be focused on each element individually, and attended information is better remembered (Craik et al., 1996). Moreover, the slower presentation provides more time for consolidation (Ricker et al., 2018), and it implies longer retention in working memory, giving opportunity for maintenance processes to be applied on the memoranda which in turn may facilitate transfer to long-term memory (Hartshorne & Makovski, 2019). Another difference refers to the memory test: A single item is tested in the visuospatial tasks, typically with a recognition test, whereas all items are recalled in forward order in the Hebb paradigm. Trying to retrieve every element in turn may promote learning via the testing effect (Roediger & Butler, 2011; Sutterer & Awh, 2016), as recall also contributes to the Hebb effect (Cunningham et al., 1984; Oberauer & Meyer, 2009).

So far, it is unclear which of these features are critical to yield visual long-term memory learning. The study of Logie et al. (2009) suggested that visuospatial learning occurs when recall is verbal. This, however, leaves unclear what was the basis of the learning—the visual arrays, or a list of verbal labels. Compared to change-detection, Shimi and Logie (2019) observed steeper learning of a visuospatial array when features of all memory items were reconstructed (i.e., selected from a set of visual features). The many differences between the change-detection and the feature-reconstruction tests, however, preclude conclusions about which test characteristic(s) promote learning: It could be the switch from change detection to reconstruction, the fact that all items were tested in the reconstruction test, or the longer maintenance duration that the latter implies.

The Present Study

Inspired by the concept of metastudies (Baribault et al., 2018), we initially designed five experiments consisting of a total of 10 conditions that systematically probed for the role of three variables distinguishing between visuospatial-array and serial-recall tasks: (a) mode of presentation (simultaneous in visuospatial tasks vs. sequential in Hebb tasks); (b) maintenance time in working memory (fast presentation in visuospatial task vs. slow presentation in serial recall tasks); and (c) features of the memory test (a single test with a potential change in a single item in visuospatial tasks vs. serial recall of all items in Hebb tasks). Our approach was to test for their joint roles by creating conditions covering most of their combinations to determine which constellations were critical to promote visuospatial learning.

We modeled our study on the typical Hebb paradigm which has been the gold standard to demonstrate learning of repeatedly presented information. Therefore, in our studies we selected a single visuospatial array to be repeated across several cycles (miniblocks of three unique arrays + the repeated array), and we used a discrete recall procedure to test memory. In all of our experiments, participants learned arrays of colored squares presented scattered across an invisible grid. They had to remember which color appeared in which location. In our first series of experiments, participants reconstructed the correct color by picking the remembered color from a discrete set of nine possible colors. Hence our testing procedure more closely aligned with that of Hebb tasks.

To foreshadow the results of our first set of studies (Experiments 1–5), the most important determinant of learning was whether a single item or all items were tested: Only the latter condition consistently led to learning. Experiment 6 therefore focused on that variable, providing a parametric variation of the number of tests (one, three, or all six items), and demonstrates a gradual increase of learning as more items are tested. Finally, with Experiments 7 and 8 we address one outstanding difference between our procedure and the one previously used in the literature, that is the use of a recall test versus a recognition test. These studies show that when arrays are tested through change detection, learning did not occur even when we tested every single item of every array.

Experiments 1–5

We preregistered the research question, hypotheses, sample-size determination, inclusion and exclusion criteria, and method of Experiments 1–5 (https://osf.io/g3v7m).

Method

Participants

Experiments 1, 2, 4a, and 5 were conducted by undergraduate students enrolled in an experimental-psychology class. Each team of three students collected and analyzed data of 24 participants for one experiment. Sample-size was based on the number each group could reasonably recruit within the seminar time-frame. Our inferences were based on Bayesian estimation, and our goal was to have enough data to credibly estimate learning.

Experiments 1, 2, 4a, and 5 were conducted in a group lab with four laptops arranged in a row with dividers between them. Participants sat at comfortable distance from the laptop (ca. 40 cm away from the monitor). We replicated one experiment (Experiment 4b, N = 27) in a standard laboratory setting with individual booths, desktop computers (distance to the monitor ca. 54 cm), and an experienced research assistant.[1] Similar results as in Experiment 4a were obtained, and hence their data was collapsed (total N =

[1] We replicated this experiment as an initial analysis revealed ambiguous results. Results split by experiment (4a and 4b) are presented in the supplemental online materials.

51). Furthermore, we ran one experiment in the laboratory (Experiment 3) with the same set-up as in Experiment 4b.

Participants were between 18 and 35 years old, and had no self-reported color blindness. They took part in only one experiment as volunteers, either in exchange for course credit (Experiments 1, 2, 4a, and 5) or 15 CHF (Experiments 3 and 4b). Participants read and signed an informed consent form before the experiment, and they were debriefed at the end. The experimental protocol followed the ethical guidelines of the Institutional Review Board and did not require specific ethical approval.

Materials

Experiments 1 to 5 were programmed in MATLAB using the Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997). All our experimental tasks are available in our OSF project (https://osf.io/65bmd). Stimuli were modeled after Adam and Vogel (2017): The memory array consisted of six colored squares presented against a gray background (RGB 128 128 128). The squares (side = 60 pixels) appeared in a random subset of cells of an invisible, centered 6 × 6 grid. Colors were selected from nine values (RGB): white (255 255 255), black (0 0 0), blue (0 0 255), cyan (0 255 255), green (0 255 0), yellow (255 255 0), orange (255 128 0), red (255 0 0), and magenta (255 0 255).

For each participant, 154 unique memory arrays were created with the following constraints: (a) the six stimuli had different colors and were separated by a distance of at least one grid cell such that no colored squared touched another; and (b) the whole array differed from all other arrays by at least two color-location associations. Half of the arrays were assigned to each within-subject condition. For each condition, one array was *repeated* amid *unique arrays*, which only occurred once in the experiment. Each condition consisted of 20 (Experiments 1 and 5) or 24 (Experiments 2, 3, and 4) miniblocks of four trials: The first three trials presented unique arrays, and the fourth trial, the repeated array. This repetition schedule (i.e., presentation of the repeated array after a constant number of unique arrays) is frequently used in Hebb studies (Couture & Tremblay, 2006; Cumming et al., 2003; Gagnon et al., 2005; Oberauer et al., 2015; Oberauer & Meyer, 2009).

Procedure

The experiments were divided into a working memory phase and a long-term memory test phase. The long-term memory test phase was exploratory, and its results are presented in the online supplemental materials in our OSF project (https://osf.io/65bmd/).

Working Memory Phase. Each experiment had two working memory conditions (see Table 1, Column A) which were completed in separate blocks. Block order was counterbalanced across participants. Conditions in each experiment differed in only one variable. Across all five experiments, manipulations of three variables were implemented: (a) presentation mode (sequential vs. simultaneous); (b) total time in working memory (fast: 1,200 ms vs. slow: 6,000 ms); and (c) recall test (single-item, whole-report in random order, or whole-report in forward order). Figure 1 illustrates the general implementation of these manipulations.

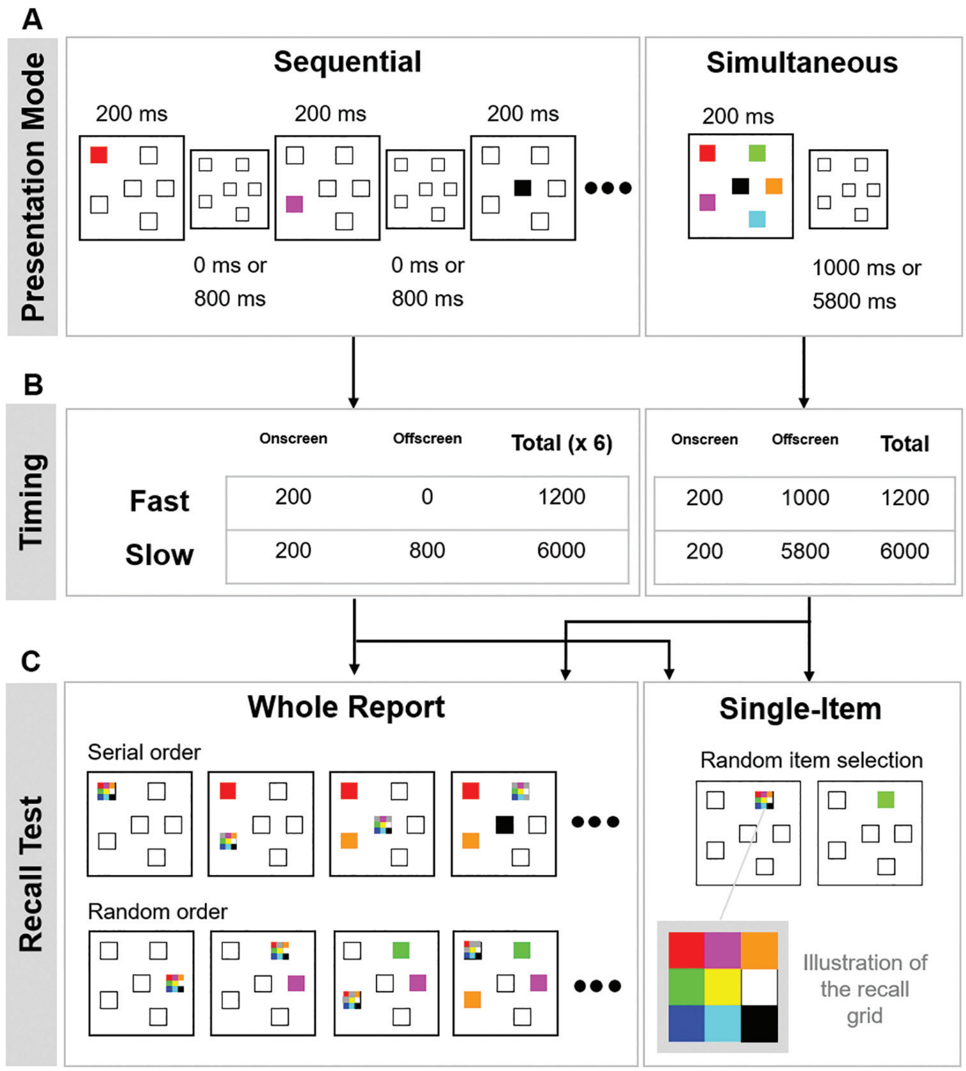
Table 1

Overview of (A) the Conditions Implemented in Experiments 1–5, and the Main Parameter Estimates (Mean and 95% HDI) for the (B) Models Fitted to the Full Data of Each Experiment, and (C) Fitted to the Data of Each of the 10 Conditions Separately

	A. Task parameters			B. Full model		C. Model by condition
Exp.	Presentation	Timing	Test	Array × Block	Array × Block × Condition	Array × Block
1	Sequential	***Fast***	Serial recall	**.21 [.13, .28]**	**.07 [.04, .09]**	**.15 [.06, .24]**
	Sequential	***Slow***	Serial recall			**.28 [.19, .38]**
2	Sequential	Fast	***Single-item report***	**.05 [.03, .07]**	**.04 [.03, .06]**	.00 [−.10, .10]
	Sequential	Fast	***Whole-report, Random***			**.16 [.06, .26]**
3	Simultaneous	Fast	***Single-item report***	.06 [−.01, .13]	**.10 [.07, .13]**	.08 [−.01, .19]
	Simultaneous	Fast	***Whole-report, Random***			**.16 [.07, .25]**
4	***Simultaneous***	Slow	Single-item report	.07 [−.01, .16]	**.15 [.08, .23]**	.07 [−.01, .16]
	Sequential	Slow	Single-item report			**.21 [.12, .32]**
5	***Simultaneous***	Slow	Whole-report, Random	**.19 [.11, .27]**	**.07 [.04, .10]**	**.10 [.04, .17]**
	Sequential	Slow	Whole-report, Random			**.29 [.16, .43]**

Note. HDI = highest density interval. In (A), within-subject conditions contrasted in each experiment are printed in italics, bold. In (B) and (C), parameters that were credibly different from 0 are printed in bold.

Figure 1
Illustration of the Task Parameters Across the Different Experiments.



Note. Panel A: Flow of events in the study phase for sequential and simultaneous arrays. During the blank intervals, placeholders remained onscreen. In the actual task, the screen background was gray. Panel B indicates the timings in effect in the conditions with fast and slow presentation rates (sequential presentation) or fast and slow retention intervals (simultaneous conditions). Panel C illustrates the test procedures used for the whole-report procedure either with serial- or random-ordered recall, and also the single-item report procedure. The inset in Panel C shows the response grid used. After the selection of each element, the selected color was displayed at the tested location. See the online article for the color version of this figure.

Each trial consisted of the presentation of a six-item array (study phase) followed by a probed-recall test (test phase). Trials were self-initiated with a mouse-button press. Thereafter, six placeholders in dark gray (RGB 112 112 112) appeared for 500 ms, indicating the locations of the upcoming colors. Placeholders remained onscreen during the study and test phases.

In the *sequential* conditions (Figure 1A), each memory item was presented individually for 200 ms. In the *fast* conditions, item offset was followed immediately by the next item. Given that six items were presented, this represents a total duration of the study phase of 1,200 ms (see Experiments 1 and 2). In the *slow* conditions, an 800-ms blank screen was inserted between items. This leads to a total duration of the study phase of 6,000 ms (see Experiments 1, 4, and 5). In the *simultaneous* conditions, all colors appeared together for 200 ms, followed by a retention interval of 1,000 ms (*fast* condition, total study + retention time = 1,200 ms; Experiment 3) or 5,800 ms (*slow* condition, total study + retention time = 6,000 ms; Experiments 4 and 5).

Figure 1C illustrates the testing procedures. At test, a checkerboard with nine colored squares (side = 20 pixels) arranged in a 3 × 3 grid appeared at the tested location (see inset in Figure 1C), and participants clicked on one of the colors. This color then appeared at the tested location, and it was replaced by a gray square in subsequent checkerboards, impeding its reuse in that trial. Participants could not correct their responses after entered. In the *whole-report* procedure, all memory items were tested either in the same order as presentation (Experiment 1) or in random order (Experiments 2, 3, and 5). In the *single-item report* procedure, a single randomly-selected memory item was tested (Experiments, 2, 3, and 4). The probability that each individual element was selected to be tested was therefore one sixth in this type of test. Hence over the repetitions, the same element would be repeatedly tested only one sixth of the times (i.e., four times). After recall was completed, visual feedback appeared for 1,500 ms: Correctly recalled colors remained onscreen, whereas incorrect colors turned dark-gray. Afterward, a new self-initiated trial started.

Long-Term Memory Test Phase. After completion of the working memory phase, participants were presented with an open-ended question inquiring whether they noticed something special about the experiment, and they typed a response. This was used to classify participants regarding their awareness of the repeated arrays. Afterward, participants were informed about the existence of the repeated arrays, and a recognition test followed. Twelve arrays were presented one-by-one, and participants indicated via keypress if this array was one of the repeated arrays (left-arrow keypress) or not (right-arrow keypress). The 12 arrays consisted of the two repeated arrays (one from each condition), two unique arrays presented in the course of the experiment, four completely new arrays, and four intrusion arrays which were similar to the repeated array. For each repeated array, one intrusion array differed from it in only one color (intrusion-1 array) and the other intrusion array differed from it in three colors (intrusion-3 array). This recognition test served to indicate to which degree participants had some long-term memory trace of the repeated array even if it may have not been sufficiently precise or accessible to increase performance in the working memory phase.

The results of the long-term memory phase are presented in the online supplemental materials available at the OSF. Overall, about half of the participants reported having noticed the repeated arrays. For most experimental conditions, there was ambiguous evidence that participants could distinguish the repeated arrays from new arrays, even when there was credible learning on that condition. We note that the retrospective nature of the report, and the fact that we tested both learned arrays from the two experimental blocks at the end of the entire experiment, limits the strength of any conclusions that can be drawn from this exploratory analysis.

Data Analysis

We assessed learning rate as the increase of accuracy over miniblocks for each array-type (repeated vs. unique). We examined whether accuracy for the repeated arrays increased above that for unique arrays across miniblocks, and whether that effect depended on the three variables manipulated: (a) presentation mode, (b) time in working memory, and (c) recall test.

We modeled the data of each experiment with Bayesian hierarchical generalized mixed models (BGLM) using the *brms* package (Bürkner, 2017) implemented in R (R Core Team, 2017). We modeled accuracy (n correct out of k responses) as a binomial distribution with a logit link-function using the following equations:

$$\begin{aligned} P(correct) = f(b_0 &+ b_1 \times block^{\alpha} + b_2 \times condition \\ &+ b_3 \times block^{\alpha}\, x\, arrayType \\ &+ b_4 \times block^{\alpha} \times condition \\ &+ b_5 \times block^{\alpha} \times arrayType \times condition) \\ n &\sim Binomial(k,\, P(correct)) \end{aligned}$$

Accuracy for repeated and unique arrays was predicted by miniblock (referred to as *block* in the equation, which varied from 1–24) and the within-subject manipulation implemented across the two experimental conditions. We also included the Block × Array-Type and Block × Condition Two-Way interactions, and a Block × Array-Type × Condition Three-Way interaction. A main effect of array-type was not included in the model because repeated and unique arrays were indistinguishable at the start of each condition. Learning of the repeated array therefore could only lead to an interaction of array-type with block. By omitting the main effect of array-type we forced the effect of learning to be fully captured by a single model parameter (i.e., the Array-Type × Block interaction), thereby making the test more sensitive to detect evidence for learning.

The block variable was entered in the model as a numeric predictor divided by its standard deviation to estimate its effect on the standardized effect-size scale. The predictors array-type (–1 = unique, 1 = repeated) and *condition* (–1 vs. 1) were contrast-coded. Note that for each experiment, the variable *condition* represents a different variable contrast (e.g., in Experiment 1 this represents the manipulation of presentation rate—fast vs. slow; see Table 1, Column A). The increase over blocks was modeled with a power function, with α determining the slope of change over miniblocks (linear for $\alpha = 1$; decelerating for $\alpha < 1$; or accelerating for $\alpha > 1$). This gave the model maximal flexibility to represent a monotonic learning curve. The models included random slopes for all predictors and their two-way interactions. For Experiment 2, random slopes for the two-way interactions led to bad convergence, hence we removed them. Regression coefficients had weakly informative Cauchy priors (location 0; scale .7), except the α parameter whose prior was a Gamma distribution (shape = 1; scale = 1).

Two parameters were of main interest: (a) the Block × Array-Type interaction, indicating that repeated and unique arrays diverged as repetitions accumulated, and (b) the Block × Condition × Array-Type interaction, indicating that learning differed between the within-subject conditions.

To follow up on the three-way interactions, we also fitted models separately to each of the condition contrasted in an experiment (i.e., presentation mode, time in working memory, or number of items tested). For these models, the main interest was in the Block × Array-Type interaction. It indicates whether in that particular condition, learning of the repeated array was credible. For the whole-report condition of Experiment 2, the model with the power function on block did not converge, so we fitted the model with a linear learning curve (i.e., fixing α to 1).

Parameters were estimated with three Markov chain Monte Carlo chains, each containing 1,000 warmup and 9,000 postwarm-up samples. The R-hat values for the parameters were below 1.05, indicating convergence. For inference, we took the mean of the parameter posterior and its 95% highest density interval (HDI), that is, the interval covering the credible parameter values given the data. When the HDI did not include 0, it was considered credibly different from 0 (Kruschke, 2013).

Results

Some conditions involved a single-item test for which performance could only be coded as correct or incorrect, hence for visualization we aggregated performance over four miniblocks (hereafter epoch). Figure 2 presents accuracy for the repeated and unique arrays over epochs in each condition together with model predictions (i.e., a posterior predictive check). As illustrated in the figure, the model captured the data well.

Table 1 presents the posterior estimates of the main parameters of interest for the models fitted to the data of both conditions of each experiment (Column B), and fitted to the data of each within-subject condition of that experiment separately (Column C). The online supplemental materials present all parameter estimates. For Experiments 1–5, the three-way interaction was credible, hence all three manipulated variables—presentation mode, time in working memory, and number of items tested—credibly affected the degree of visuospatial learning. The sizes and implications of these effects, however, differed. To identify which conditions produced credible learning, we considered the Block × Array-Type interactions for the models fitted to the data of each condition separately (Table 1, Column C).

Experiment 1

This experiment comprised conditions most similar to the traditional Hebb paradigm (i.e., sequential presentation and serial recall) with the only difference that we varied time in working memory between conditions (fast vs. slow). The slow condition represents the more typical scenario (Couture & Tremblay, 2006; Tremblay & Saint-Aubin, 2009) in which ample time is provided between the sequential presentation of the memoranda (i.e., with 1 s per item). The fast pace condition was introduced to match the timing typically used in visuospatial array tasks, which usually provide only a very brief interval for encoding. The slow presentation rate yielded better performance than the fast rate, replicating the presentation rate effect (Ricker & Hardman, 2017; Souza & Oberauer, 2018; Tan & Ward, 2008). Learning of the repeated array occurred in both conditions (see Figure 2), as revealed by their credible Array-Type × Block interactions (Table 1, Column C). The three-way interaction (Table 1, Column B) was due to the somewhat larger learning in the slow condition. This result indicates that providing more time for consolidation in working memory (200 ms vs. 1,000 ms) improved performance overall, and improved the learning of the repeated array; however, even with the fast presentation rate, participants learned the repeated arrays.

Experiments 2 and 3

Experiments 2 and 3 contrasted the single-item versus whole-report testing procedures in a within-subjects design. The two experiments differed only regarding the mode of presentation of the memoranda: Experiment 2 used sequential presentation, whereas Experiment 3 employed simultaneous presentation. In both experiments, learning was not credible for the single-item conditions, but occurred credibly in the whole-report conditions. Accordingly, Table 1, Column C shows no credible Array-Type × Block interactions in in the conditions with single-item tests, but credible interactions in the whole-report conditions. The presence versus absence of learning as a function of the single-item versus whole-report conditions, therefore, explains the credible three-way interactions obtained for these two experiments (Table 1, Column B).

Taking the results of Experiments 2 and 3 together, we can conclude that the mode of array presentation (sequential or simultaneous) was not critical for observing credible learning, as the same pattern of results was observed in both experiments. The critical variable to switch learning on and off was the number of elements tested (i.e., all vs. one).

Experiments 4 and 5

Experiments 4 and 5 contrasted simultaneous versus sequential presentation in a within-subjects design. The two experiments differed in terms of the type of memory test: Experiment 4 implemented only single-item tests; Experiment 5 only whole-report tests. In contrast to Experiments 2 and 3, Experiments 4 and 5 implemented a slow pace. In Experiment 4, only single-item tests were implemented, and learning was not credible in the simultaneous condition, but was credible in the sequential condition. In Experiment 5 all items were tested, and learning was credible in both conditions. In light of the results of Experiments 2 and 3, the lack of learning in one of the single-item conditions of Experiment 4, combined with the credible learning in both whole-report conditions of Experiment 5, corroborates the relevance of the number of elements tested for promoting learning. The only constellation in which the single-item test did not prevent learning was when it was combined with slow, sequential presentation of the memoranda.

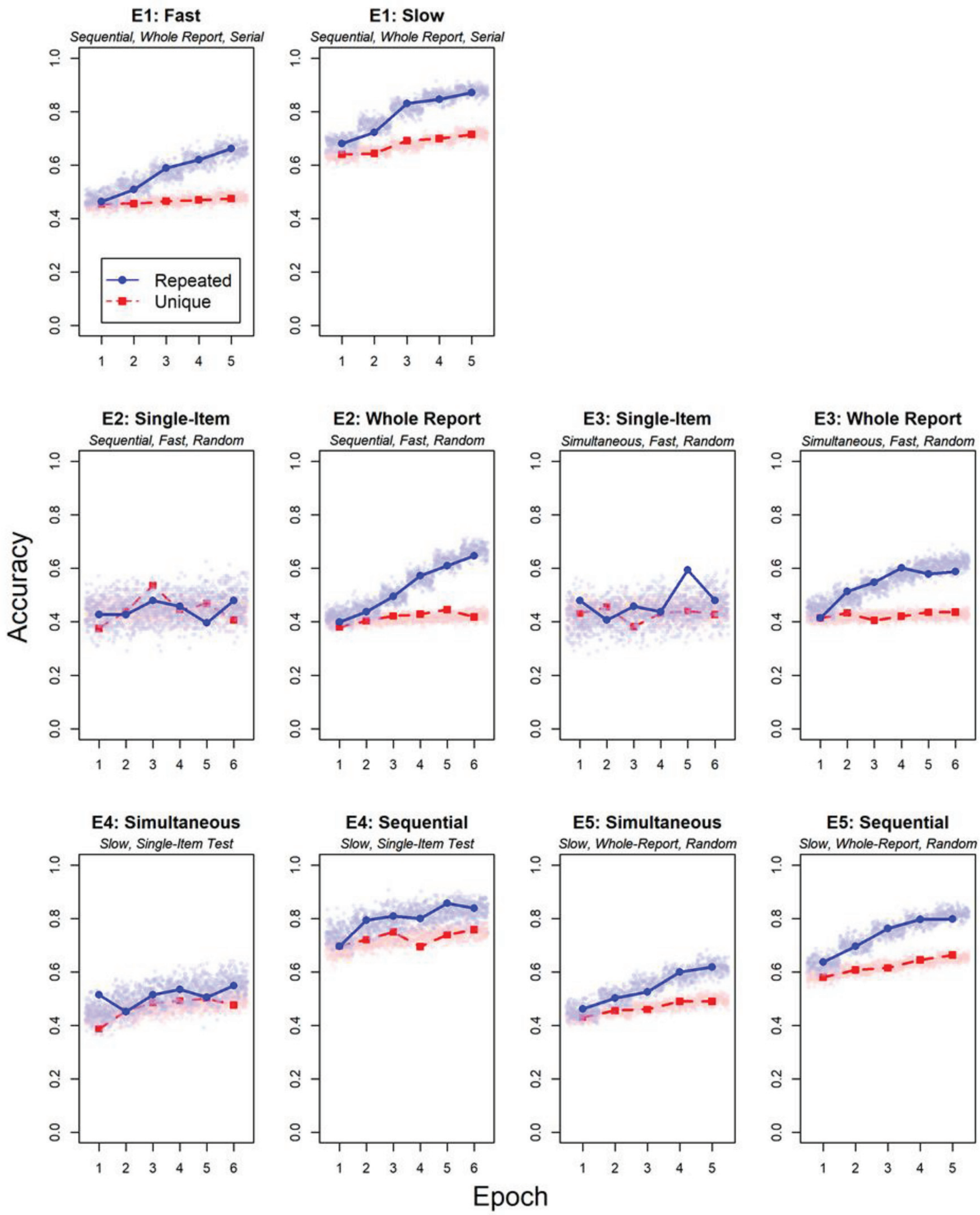
Discussion

Across all five experiments—which implemented 10 different combinations of the variables: (a) presentation mode, (b) time in working memory, and (c) number of elements recalled—we observed that the task parameter most critical to promote learning of the visuospatial arrays was how many elements were tested over the short-term. In all six conditions that involved whole-report, credible learning of the repeated array was observed. This was observed irrespective of the recall order being the same on every trial (serial recall procedure in Experiment 1) or varying randomly across trials (Experiments 2, 3, and 5). Hence, the learning of the repeated arrays cannot be explained by assuming that participants learned a motor sequence. Additionally, learning did not depend on the other task parameters varied, namely presentation mode or timing. In contrast to the credible learning observed in the whole-report conditions, three of the four conditions in which single-item tests were implemented showed no credible learning. Learning was only observed when the single-item test was combined with sequential presentation and a slow presentation rate.

It is worth noting that simultaneous condition of Experiment 3 is the one that most closely resembles the traditional visuospatial working memory tasks: simultaneous array presentation for 200

Figure 2

Accuracy (Proportion Correct) for the Repeated and Unique Arrays as a Function of Epoch in Each Condition of Experiments 1–5



Note. E = experiment. The solid lines and dots show average proportion correct for the observed data. The small clouds of overlaid dots present samples from the posterior-predictive distribution of the model fitted to the data of each experiment (i.e., posterior predictive check), jittered slightly along the x-axis for visibility. Each epoch represents the average of four miniblocks (consisting of four repeated and 12 unique arrays). See the online article for the color version of this figure.

ms, followed by a retention interval of 1,000 ms, and a single-item test. In this set-up, we replicated the absence of learning previously reported in visuospatial tasks. Increasing the retention interval in this set-up (simultaneous condition of Experiment 4) also did not produce learning.

The results of these five experiments suggest that learning of the repeated array is favored when all encoded items are recalled. In this series of experiments, however, we only contrasted single-item to whole-array report. If indeed the number of elements retrieved is critical to promote learning, we should observe that the learning rate is proportional to the number of elements tested. Experiment 6 was designed to test this prediction.

Experiment 6

Experiment 6 was designed to examine the effect of the number of elements tested on the degree of visuospatial learning in a more fine-grained way. Participants were exposed to three conditions that differed regarding the number of items tested: one, three, or six. The one-test and six-tests conditions replicate the single-item and whole-report conditions of Experiment 3. The three-tests condition involved test of a random half of the items in the memory array. If testing promotes learning, learning in the three-test condition should be intermediary between the one-test and six-tests conditions.

Method

Participants, Design, and Procedure

We collected a new sample of 30 participants from the University of Zurich (same inclusion criterion as described for Experiments 1–5). Data was collected by an experienced research assistant in the laboratory setting, that is, in individual booths and desktop computers, with monetary reimbursement for participation (15 CHF).

The general experimental task in Experiment 6 was identical to the one implemented in Experiment 3, that is, the memoranda were presented simultaneously for 200 ms followed by a 1,000-ms retention interval. We selected the task set-up of Experiment 3 because it implements parameters most similar to the ones used in visuospatial working memory tasks (i.e., simultaneous, brief presentation of the memoranda followed by a short retention interval).

The only difference between Experiment 6 and Experiment 3 pertains to the test phase. We implemented three within-subjects conditions in which we varied how many items were tested: (a) the *one-test condition* was exactly as the single-test condition of Experiment 3 in which a single, random element of the array was tested on each trial; (b) the *six-tests condition* was exactly as the whole-report condition implemented in Experiment 3, that is, all six studied items were tested in random order; finally, (c) the *three-tests condition* consisted of the random test of three out of the six studied items. Each condition was completed in one large block in which one new repeated array was implemented for learning, following the same schedule as in Experiments 1–5 (every fourth trial was a repetition). Participants completed 96 trials in each condition (24 repetitions of Hebb array). The order of the conditions was counterbalanced across participants (i.e., six possible different orders which were replicated 5 times).

Data Analysis

We submitted the data of Experiment 6 to a similar regression model as in the previous experiments, with exception that we fixed the alpha parameter to 1,[1]entering miniblock (aka block, standardized) and array-type (contrasted coded as unique = –1; repeated = 1) as predictors of the proportion of correctly recalled colors. We took two approaches: (a) we fitted the model to the data of each condition separately. This allowed us to estimate whether there was evidence for a credible learning of the repeated array in each condition when considered alone. In this model, the main parameter of interest is the Array-Type × Block interaction. (b) we fitted the model to the data of two conditions at a time (1 vs. 3; 3 vs. 6; 1 vs. 6), entering the number of tested items as a third predictor. Of interest in this model is the main effect of number of tested items, the two-way interaction between Block × Array-Type, and the three-way interaction between Block × Array-Type × n Tested. This allowed us to assess whether there was credible evidence for a difference in learning rate in the pairwise comparison of the three within-subject conditions.

We fitted the models with three chains of 5,000 iterations, from which the first 1,000 were discarded as burn-in. All models converged (i.e., R-hat for all parameters was below 1.05).

Results

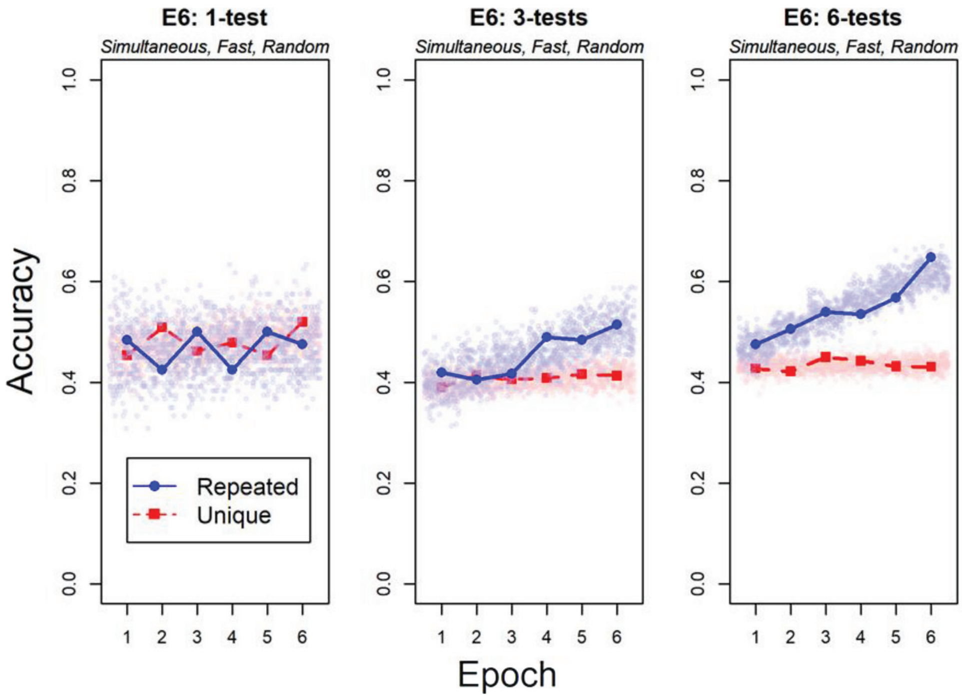
Figure 3 presents the proportion of correct responses for the repeated and unique arrays over the epochs (i.e., the average of four miniblocks) in each condition of Experiment 6. Table 2 presents the posterior of the parameters of interest to our research questions for the models fitted to the data of each condition separately (Table 2, Column B) and for the pairwise contrast between conditions (Table 2, Column C). The posterior of all estimated parameters is available at the online supplemental materials (https://osf.io/65bmd/).

Figure 3 shows that there was no credible learning when a single element of the memory array was tested (one-test condition), but learning was credible when multiple items were tested (three-tests and six-tests). As shown in Table 2, Column B, the estimate of the learning effect (Block × Array-Type interaction) was larger in the whole-array test condition (six-tests) than when only half of the array was tested (three-tests). The pairwise comparison of the conditions (Table 2, Column C) revealed that the three-way interaction between Block × Array-Type × n Tested was only credible for the contrast between the one-test versus six-tests conditions, replicating Experiment 3. In the pairwise contrasts the three-tests condition did not credibly differ from both the one-test and from the six-tests condition.

[2] Fitting of the exact same nonlinear model implemented across Experiments 1–5 did not converge when used in Experiments 6, 7, and 8, particularly when considering the pairwise contrast between conditions. For the fitting of the data of individual conditions separately, the estimated values for critical the Block × Array-Type interaction were very similar to the one obtained with the model in which we dropped the alpha parameter (see online supplemental materials).

Figure 3

Accuracy (Proportion Correct) for the Repeated and Unique Arrays as a Function of Epoch in Each Within-Subject Condition of Experiment 6



Note. The solid lines and dots show the average proportion correct for the observed data. The small clouds of overlaid dots present samples from the posterior-predictive distribution of the model fitted to the data of each condition separately (i.e., posterior predictive check), jittered slightly along the x-axis for visibility. Each epoch represents the average of four miniblocks (consisting of four repeated and 12 unique arrays). See the online article for the color version of this figure.

It is worth pointing out that when the three-tests condition was contrasted with the one-test condition, the two-way interaction between Block × Array-Type was not credible (see Table 2, Column C), but this interaction was credible when the three-tests condition was compared to the six-tests condition. These results further support the conclusion that learning in the three-tests condition was intermediary between the one- and six-tests conditions.

Discussion

In Experiment 6, we varied the number of elements tested in a more fine-grained matter than in Experiments 1–5. Our goal was to assess whether testing only a subset of the array would produce learning. When only half of the array was tested, participants showed some evidence of learning. The estimated learning effect was, however, smaller than in the condition in which all items were tested. Because learning in the three-tests condition was intermediary between the one-test and the six-tests conditions, we could not credibly differentiate the three-tests learning effect from both other conditions. Overall these results are consistent with the hypothesis that learning is proportional to the number of elements tested.

Experiment 7

In Experiments 1 to 6, we implemented a discrete color-recall procedure to test memory. Previous studies of long-term learning with the visuospatial-array task, however, used recognition tests that presented a single-probe (Fukuda & Vogel, 2019; Olson & Jiang, 2004) or the whole-array (Logie et al., 2009; Olson & Jiang, 2004; Shimi & Logie, 2019) at the test phase, and for which participants were required to indicate whether or not it contained a change. So far, it is unclear if our results replicate with the change-detection task.

Accordingly, the goal of Experiment 7 was to assess the extent in which testing multiple elements can promote learning when a recognition test is used instead of a recall test. We compared the effect of the number of tests in two versions of the change-detection paradigm: single-probe versus whole-array. Olson and Jiang (2004) failed to observe learning of the repeated arrays irrespective of using a single-probe or whole-array to test memory. This leads to the expectation that presenting a single-probe or the whole-array at test will make no difference when a single element is being tested, replicating our single-report conditions in Experiments 1–6. Nevertheless, it is unclear whether testing all elements of the array would boost learning in change-detection tasks akin to what we observed for our discrete recall test, and whether the type of probe array would matter. When the probe-display presents all items (whole-array probe), participants are represented with an entire array containing a change in relation to the encoded array in 50% of the trials. When the probe has six colors, participants may also encode the probe colors to working memory. Because the

Table 2

Overview of the Conditions Implemented in Experiments 6, 7, and 8 (Column A), and the Main Parameter Estimates (M and 95% HDI) for the Models Fitted to the Data of Each Condition Separately (Column B), and to Data of Two Conditions at a Time (Column C)

A. Task parameters					B. Condition	C. Pairwise condition contrast			
Exp.	Cond	Presentation	Timing	Test	Array × Block	Contrast	Condition	Array × Block	3-way
6	1	Simultaneous	Fast	*1-item test*	−.01 [−.10, .07]	*1 vs. 3*	**.14 [.04, .24]**	.03 [−.02, .08]	−.04 [−.09, .01]
	3	Simultaneous	Fast	*3-items test*	**.06 [.01, .11]**	*3 vs. 6*	**.10 [.04, .16]**	**.08 [.05, .11]**	.02 [−.01, .05]
	6	Simultaneous	Fast	*6-items test*	**.10 [.07, .14]**	*1 vs. 6*	−.04 [−.13, .06]	.05 [−.00, .09]	**.06 [.01, .11]**
7	S1	Simultaneous	Fast	*Single-probe/one-test*	.01 [−.05, .07]	*S1 vs. W1*	**−.20 [−.30, −.09]**	.01 [−.03, .05]	−.01 [−.05, .03]
	W1	Simultaneous	Fast	*Whole-array/one-test*	.00 [−.05, .06]	*S1 vs. S6*	**−.13 [−.24, −.01]**	−.03 [−.07, .02]	.03 [−.01, .07]
	S6	Simultaneous	Fast	*Single-probe/six-tests*	.00 [−.03, .03]	*S6 vs. W6*	−.01 [−.10, .08]	.01 [−.02, .03]	.00 [−.02, .03]
	W6	Simultaneous	Fast	*Whole-array/six-tests*	.01 [−.02, .04]	*W1 vs. W6*	.06 [−.06, 18]	−.00 [−.04, .04]	.01 [−.03, .05]
8		Simultaneous	Fast	Whole-report (six-tests)	**.09 [.06, .12]**				

Note. HDI = highest density interval. In (A), conditions contrasted in each experiment are printed in italics. In (B) and (C), parameters that were credibly different from 0 are presented in bold. The condition predictor reflects the effect of *n* tested in Experiment 6, and the effect of probe array in Experiment 7.

probe has some different colors compared to the memory array, but it is similar to it otherwise, this may create substantial interference. In a single-probe display, interference is arguably more limited because participants can focus on one element at a time. Given these considerations, and the potential benefit of generalizing our conclusions across different task set-ups, we tested four between-subjects conditions in Experiment 7 that crossed the type of change-detection array (single-probe vs. whole-array) and the number of elements tested in each trial (one-test vs. six-tests).

Due to the COVID-19 pandemic, we had to move our experiment to an online format and collect data on Prolific. Online data-collection, however, is limited by potential uncontrollable distractions at home as well as the risk of connection failures, and therefore we had to strive to keep our experiment brief. We decided therefore to implement our manipulations across four between-subjects conditions, with participants being randomly assigned to: (a) single-probe/one-test; (b) single-probe/six-tests; (c) whole-array/one-test; (d) whole-array/six-tests condition. Assigning participants to a single condition allowed us to keep our experiment lasting between 20 and 30 min.

Method

Participants

We recruited participants online via the Prolific platform. Participants were English speakers, in the age range from 18 to 35 years, and with no self-reported color blindness. The study lasted about 20 min, and participants were compensated with three British Pounds. Participants were randomly assigned to one of four groups that differed regarding the test procedure: (a) single-probe/one-test (S1), (b) single-probe/six-tests (S6), (c) whole-array/one-test (W1), and (d) whole-array/six-tests (W6). We obtained valid data of a total of N = 192 participants, of which n = 29 were excluded due to low performance (i.e., performance one standard-deviation below their group mean). Results do not differ if we include all participants in the analysis. This left us with a final sample size of N = 169. Table 3 shows the distribution of participants over the four different groups as well as the number of exclusions in each group.

Design and Procedure

This experiment was programmed using lab.js (Henninger et al., 2020) which provides an online builder (https://labjs.felixhenninger.com/) for programming experiments. The adaptation of the task for running in lab.js, which uses HTML and Java-script, required some small modifications. First, the size of the grid was slightly increased from 6 × 6 to 7 × 7. This was implemented to increase the potential pool of different arrays in the task. Second, the size of the squares was reduced to 40 pixels. We expected participants to have smaller screens (laptops) and hence we resized the squares and our grid to compensate for that. Participants completed the experiment within their own computers at an unknow viewing distance.

The design of Experiment 7 was equivalent to the one implemented in Experiment 3 and 6: The memoranda were presented simultaneously for 200 ms, followed by a 1,000-ms retention interval, and finally the test phase (see Figure 4). The test phase is what sets this experiment apart from the ones previously described. Here we replaced the discrete color-recall task by a

Table 3
Number of Participants in Each of the Four Groups in Experiment 7

Type of array	1 test	6 tests
Single-probe	$n = 40$ (7)	$n = 46$ (8)
Whole-array	$n = 35$ (7)	$n = 42$ (7)

Note. Numbers in parentheses indicate excluded participants due to low performance.

recognition test requiring only a same or change response: Participants were required to indicate if the probed color was the same (right-arrow keypress) or changed (left-arrow keypress) compared to the one that appeared in the same location in the memory array.

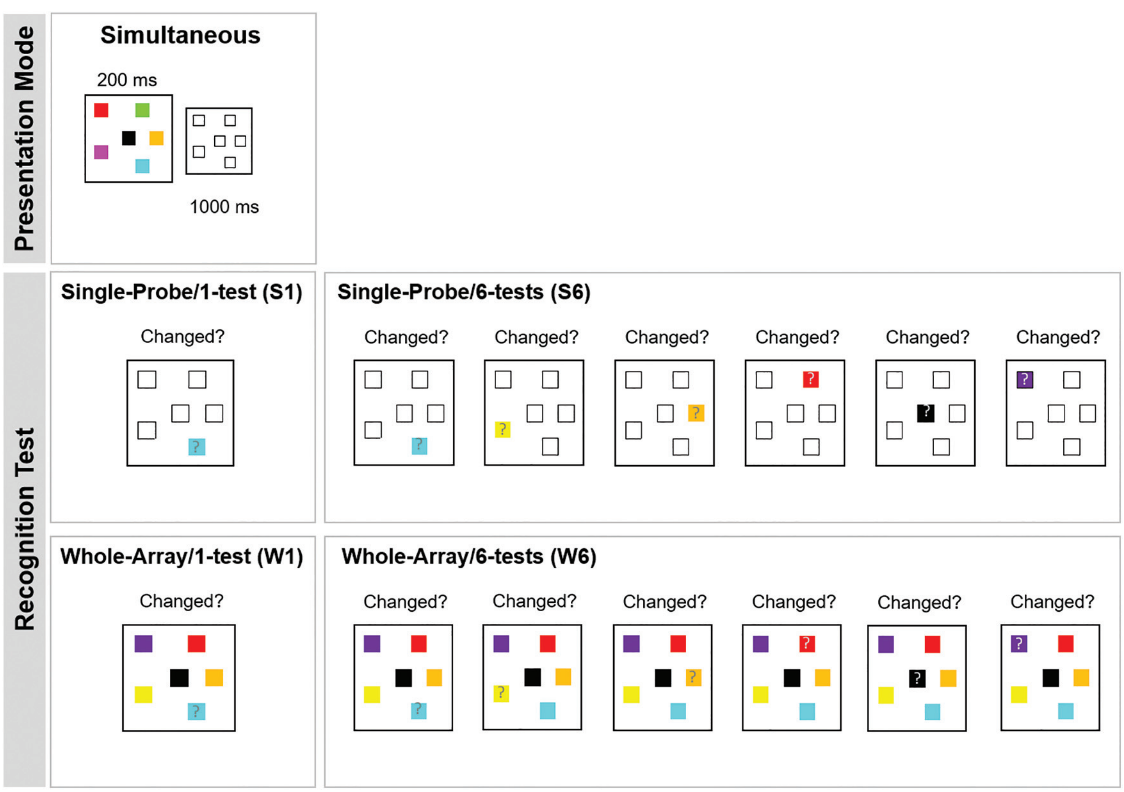
The four experimental groups differed regarding two variables (see Figure 4). The first one was the probe array configuration in the test phase: this could consist either of (a) a local, single-probe recognition test; or (b) a whole-array recognition test. The second variable was the number of elements tested: one or six items.

In the single-probe test, the placeholders marking all memory locations remained onscreen. Then a single location was filled with a color, and a gray question-mark appeared inside this square. Participants were instructed to judge if this probed color matched the one presented in the same location in the memory array. In 50% of the trials, the probed color was a match. In the remaining 50% of the trials, the probed color was either a new color (25% trials) or a color previously presented in another location in that array (25% trials).

In the whole-array test, all memory locations were filled with colors and appeared simultaneously onscreen. Although all locations were filled with color simultaneously, we asked participants to make decisions regarding each color individually. Participants were instructed to judge whether the colored square with a question mark inside matched the one previously presented in that location. Overall, there was a 50% chance that each color in the whole-array matched the previously presented color. If the color was a mismatch, it could either be a new color (25% chance) or a color from another changed location in the array (25% chance). Color repetitions were not allowed within the test array (as it was the case for the memory array).

Different from our implementation, the whole-array probe is more typically implemented without a local decision of change—that is, participants have to indicate if any change occurred. Yet, both of these versions have been used previously in studies showing lack of learning of visuospatial arrays: Olson and Jiang (2004) used a whole-array display with a local decision; Logie et al. (2009) and Shimi and Logie (2019) used whole-array displays with a global decision. We decided to use the local decision because it enabled us to compare single versus multiple tests.

Figure 4
Illustration of Memory Array and the Four Types of Recognition Tests Implemented Across the Four Between-Subjects Conditions in Experiment 7



Note: The background in the actual task was gray. The question mark was presented in gray color inside of the probed item. See the online article for the color version of this figure.

Regarding the number of tests, there either was one single element tested from each array (one-test groups), or all elements were tested in random order (six-tests groups). For the single-probe/six-tests group, there was a 50% chance that each color tested was a match or mismatch. The same constrains as described for the generation of the whole-array test were implemented in selecting the six probe-colors. For the whole-array/six-tests group, the whole-array remained visible during the test, and the question mark moved to each location in random order until all items were probed. Note that the colors remained the same in the test array from one probe to the next (i.e., only the question mark moved places). There was no time-limit to respond to each probe. After completing the test, an intertrial interval of 1,000 ms followed. Then participants pressed the spacebar to self-initiate a new trial. Unlike the previous experiments, participants did not receive feedback regarding the accuracy of their responses. The implementation of the repeated and unique arrays was exactly as described for the previous experiments (i.e., a repetition every fourth trial).

Participants in the six-tests groups completed 24 miniblocks, including 24 repetitions of the repeated array, as in the previous experiments. In the one-test groups, we doubled the number of trials to keep the overall experiment duration similar across the four groups (48 miniblocks). This also gave us the opportunity to assess whether more exposition to the Hebb (repeated) array would improve learning when a single element was tested.

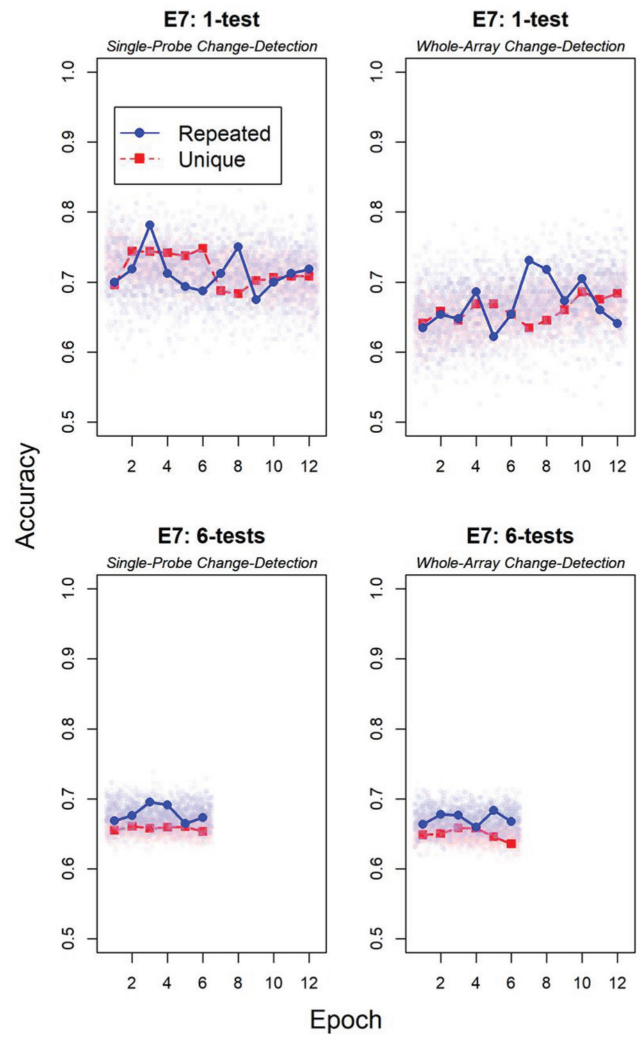
Results

We fitted the same model used in Experiment 6 to the data of Experiment 7. We fitted the data of each condition separately to estimate the Array-Type × Block interaction (three chains; 5,000 iterations each with 1,000 regarded as burn-in). We also performed condition contrasts by assessing the impact of type of probe array while keeping the number of tests constant (S1 vs. W1; S6 vs. W6) and the impact of the number of tests when keeping the type of probe array constant (S1 vs. S6; W1 vs. W6). For the latter contrasts, we considered only data of the first 24 miniblocks to equate the number of repetitions across the one- and six-tests groups. We contrast coded the conditions (single-probe = –1, whole-array = +1; and one-test = –1, six-tests = +1).

Figure 5 presents the results of Experiment 7. Overall accuracy was around 65% in most conditions, except for somewhat higher accuracy (ca. 70%) in single-probe/one-test condition. Hence, participants were performing well above guessing level (which is .5 in a two-choice test) but below ceiling. Table 2, Column B presents the posterior of the Array-Type × Block interaction when we modeled the data of each condition separately. The posterior of all parameters is presented in the online supplemental materials. There was no evidence of learning in any of the four conditions: The mean of the posteriors was close to 0 and extended symmetrically around this value. We also contrasted conditions to each other. The only credible effect of condition was found when contrasting the remaining conditions to the single-probe/one-test (S1) condition because this condition produced better performance overall. This result indicates that the single-probe/one-test condition reduced the amount of interference by showing only one color at test, thereby producing better performance.

Figure 5

Accuracy (Proportion Correct) for the Repeated and Unique Arrays as a Function of Epoch in Each Between-Subject Condition of Experiment 7



Note. The solid lines and dots show the average proportion correct for the observed data. The small clouds of overlaid dots present samples from the posterior-predictive distribution of the model fitted to the data of each condition separately (i.e., posterior predictive check), jittered slightly along the x-axis for visibility. See the online article for the color version of this figure.

We also explored whether there was any effect of repetition on reaction time (RT). We selected only correct RTs (excluding 33.3% of the responses), and then applied two criteria to trim the RTs. First, we took an absolute cut-off: We removed RTs faster than 200 ms and slower than 6,000 ms. This resulted in the removal of 1.3% of the correct RTs. Then we computed the mean RT and *SD* per participant and design cell (consisting of the combination of array-type and epoch), and excluded RTs more than 2.5 x *SD* above the mean. This resulted in the further exclusion of 2.66% of the remaining RTs. There was no evidence of a difference in RTs across epochs or array-type (see online supplemental materials). Hence, it is unlikely that speed-accuracy tradeoffs could explain the lack of an effect of repetition in Experiment 7.

Discussion

Change-detection tasks are commonly used in visuospatial working memory experiments. They require a decision regarding whether or not the probe matches the item in the memory array. This task has been used in previous studies to test for learning of repeated arrays of visuospatial information (Fukuda & Vogel, 2019; Logie et al., 2009; Olson & Jiang, 2004). Here we implemented four variants of the change-detection task and we failed to observe learning even when all items of the memory array were tested. Hence, the beneficial effect of the number of tests is only observed in recall tasks. We conclude that it is not only the number of tests that limits learning of visuospatial information: Recognition tests are also not conducive to promote learning.

Experiment 7 also serves as important point of replication of previous findings in the literature. Studies that failed to observe learning of repeated visuospatial arrays have used a range of different methodologies that tend to depart from traditional implementations of the Hebb paradigm (Fukuda & Vogel, 2019; Logie et al., 2009; Olson & Jiang, 2004; Shimi & Logie, 2019). Here we implemented two versions of the change-detection task (single-probe and whole-array) and replicated the absence of learning of the repeated arrays in a more standard Hebb set-up. This shows that this null finding is robust across different task implementations.

Experiment 8

Experiment 7 showed no credible evidence of learning of the repeated array irrespective of the type of probe array or the number of elements tested. This suggests that recognition tests do not favor the creation or use of the long-term memory traces of the repeated array. There were, however, still some differences between Experiment 7 and the previous ones we reported that could explain their discrepant results. First, Experiment 7 was conducted in an online set-up whereas all other experiments were conducted in the laboratory. Second, participants did not receive feedback for the responses in the change-detection task but they received visual feedback of the correctly recalled colors in Experiments 1–6. To assess if any of these variables could explain the lack of learning we observed with the change-detection procedure, Experiment 8 replicated Experiment 7 substituting the change-detection test for the whole-report test that always produced learning in our previous experiments.

Method

Participants, Design, and Procedure

Experiment 8 replicated the set-up of Experiment 7: Participants (N = 49) were recruited online via the Prolific platform using the same selection criteria and payment schedule as in Experiment 7. Individuals that took part in Experiment 7 were not allowed to take part in this study. The task set-up was the same as in the single-probe/six-tests group with the only difference that the recognition decision was replaced by the color-reproduction test used previously (i.e., a Rubik's cube with nine colors appeared at the tested location). As in Experiment 7, participants did not receive any feedback regarding the accuracy of the responses. There were 24 miniblocks of four trials, with the first three trials consisting of unique arrays, and the fourth of the repeated array.

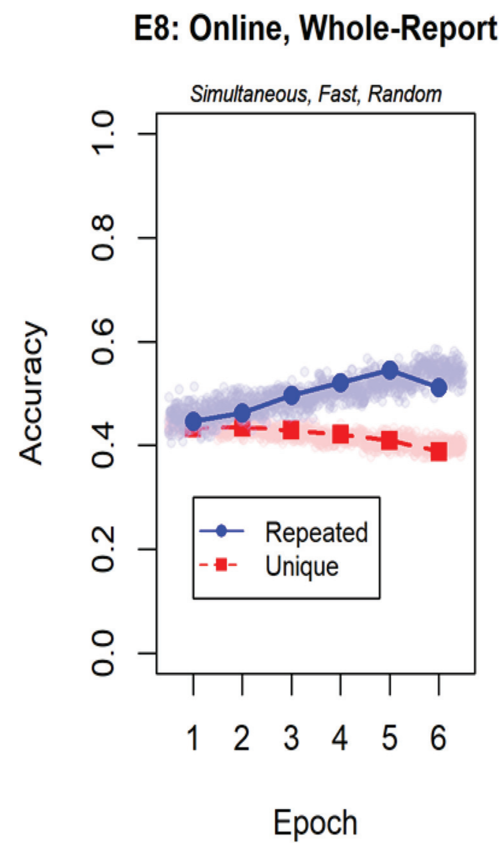
Results and Discussion

Figure 6 presents the proportion of correct responses over epoch for the unique and repeated arrays in Experiment 8. Table 2 presents the posterior of the Array-Type × Block interaction. Both the visual inspection of Figure 6 and the posterior of the interaction term between array-type and block indicate that there was credible learning when all items were tested with a discrete recall procedure even if this experiment was run online and even in the absence of feedback for the correctly recalled colors. Hence, our failure to observe credible learning in Experiment 7 cannot be attributed to these two features.

In fact, performance in Experiment 8 is very similar to the one observed in our other similar experiments that were run in the laboratory, namely Experiment 3 and Experiment 6: Participants could recall ca. 40% of the correct colors (2.4 colors on average).

Figure 6

Accuracy (Proportion Correct) for the Repeated and Unique Arrays as a Function of Epoch in Experiment 8



Note. The solid lines and dots show the average proportion correct for the observed data. The small clouds of overlaid dots present samples from the posterior-predictive distribution of the model fitted to the data of each condition separately (i.e., posterior predictive check), jittered slightly along the x-axis for visibility. See the online article for the color version of this figure.

The only difference is a slight decline of performance over blocks (but note that the main effect of block was not credible—see online supplemental materials). This suggests that the feedback could help in sustaining motivation as the experiment progresses. Yet, it does not seem critical to promote the learning of the repeated arrays.

General Discussion

The Hebb effect has been taken as evidence that repetition promotes learning: long-term traces of repeated lists gradually gain strength, improving immediate serial recall. The generality of this effect has been challenged by studies using visuospatial arrays that failed to observe learning even after dozens of array repetitions. Here we charted the conditions that facilitate or hinder visuospatial learning. We systematically considered variables related to the three main memory processes (i.e., encoding, maintenance, and retrieval) to assess what exactly influences visuospatial learning: (a) to assess differences in encoding, we manipulated presentation mode; (b) to assess differences in maintenance, we varied parameters related to how long information had to be maintained in working memory prior to the test; and (c) to consider retrieval mechanisms, we varied the number of elements tested and the test procedure (i.e., recall vs. recognition). We will summarize our findings and discuss their implications next.

Encoding: Assessing the Role of Presentation Mode

Visuospatial working memory tasks typically present all of the memoranda simultaneously for encoding, whereas typical Hebb tasks present the memoranda sequentially. Our first hypothesis was that processes affecting the encoding of the memoranda could potentially explain the absence of long-term learning of repeated arrays in previous studies. Sequential presentation allows each item to be attended without competition, which may facilitate its consolidation in memory (Ricker & Cowan, 2014), creating a stable trace that could be more easily retrieved later. With simultaneous presentation, attention could be unevenly distributed among the memoranda. If what is attended varies from one presentation to the next, repetitions might go undetected, explaining the lack of learning in previous visuospatial working memory tasks that used simultaneous arrays. To assess this possibility, we examined whether learning occurred for sequential (Experiments 1, 2, 4, and 5), and simultaneous arrays (Experiment 3, 4, 5, 6, 7, and 8). Our results indicate that simultaneous presentation per se did not hinder learning: the simultaneous conditions of Experiments 3 and 5 showed credible learning when all memoranda were tested. The same result was replicated in Experiments 6 and 8. Yet it is worth noting that sequential presentation may favor learning if items are presented slowly even when a single element is tested (as observed in Experiment 4).

These results challenge the hypothesis that sequential presentation is critical to the formation of long-term memory traces. Hence, it is unlikely that changes in encoding strategy due to presentation mode could explain the lack of learning previously reported. Simultaneous presentation has been taken as key feature to favor encoding of visual information in a visual format, given that sequential presentation could enable the recoding of the visual information into a verbal format (Frick, 1985; Sense et al., 2016). Therefore, our findings also render it unlikely that recoding of the presented elements into a verbal format is required to observe learning of the repeated visuospatial arrays.

Maintenance: Assessing the Role of Time in Working Memory

The next critical difference between Hebb tasks and visuospatial working memory tasks refers to the presentation duration and retention period. In Hebb studies, information is usually presented sequentially and slowly (e.g., 1 s per item), such that memoranda have to be maintained in working memory for longer periods. In contrast, in visuospatial working memory tasks, the memoranda are presented briefly (a few hundred milliseconds), and memory is tested after a short interval (ca. 1 s). This means that information is encoded and maintained for brief periods of time. Longer time in working memory affords more opportunities for the contribution of maintenance processes which could favor the transfer of information from working memory to long-term memory, as proposed by gateway models of the interplay between working memory and long-term memory (Atkinson & Shiffrin, 1968; Fukuda & Vogel, 2019).

To evaluate the contribution of time in working memory, half of the conditions implemented across Experiments 1–5 involved slow presentation (sequential arrays) or a long retention interval (for simultaneous arrays), requiring longer working memory maintenance (five times longer than in the fast condition). The slower presentation rate produced better performance overall, replicating studies showing a beneficial effect of time for working memory (Hartshorne & Makovski, 2019; Mızrak & Oberauer, 2021; Ricker & Hardman, 2017; Souza et al., 2020; Souza & Oberauer, 2017, 2018, 2020; Tan & Ward, 2008). This benefit is evident in the comparison of performance in the two within-subject conditions of Experiment 1, but also by contrasting performance in Experiments 2 and the sequential conditions of Experiments 4 and 5, and between performance in Experiment 3 and the simultaneous conditions of Experiments 4 and 5.

Critical to the hypothesis that time in working memory favors long-term learning, in Experiment 1 there was evidence for stronger learning with a slower presentation rate. Furthermore, in Experiment 4, the combination of a slower presentation rate and the sequential presentation of the memoranda produced learning even when a single element was tested—although single-item testing was one of the key features preventing learning of the repeated visuospatial arrays. One can also note from descriptive comparisons of the Block × Array-Type interactions in Tables 1 and 2, that learning was stronger in conditions with sequential and slow presentation of the memoranda in general (see Experiments 2, 4, and 5) compared with all other experiment with fast presentation. Hence, longer consolidation time for each memory item seems to favor learning. That said, time in working memory was not a major factor constraining learning: Learning was also observed in the fast conditions where time was only 20% of that implemented in the slow conditions. Our results therefore suggest that although more time in working memory increases the rate of learning, fast presentation rate does not prevent learning of visuospatial information.

So, how was time used for improving learning? In the verbal domain, previous studies found little support for the assumption that maintenance processes such as rehearsal and refreshing contribute to the Hebb effect. Oberauer and Meyer (2009) compared Hebb learning in conditions that implemented immediate serial recall and when a delay of 9 s was imposed between list encoding and recall. The assumption was that participants would attempt rehearsal of the memoranda during the delay, and if rehearsal boosts learning, then better learning would occur in the delay condition. That was not the case. This result also dovetails with studies showing that Hebb learning is not affected by articulatory suppression, which makes rehearsal virtually impossible (Hitch et al., 2009; Page et al., 2006; but see Sjöblom & Hughes, 2020). Another candidate process that could be used to facilitate the transfer of information from working memory to long-term memory is known as attentional refreshing, which consists of the use of attention to strengthen representations in working memory (Camos et al., 2018). One way to disrupt refreshing is by imposing an attentionally-demanding distractor task in between encoding of the memoranda, as commonly done in complex span tasks. Oberauer et al. (2015) compared learning of Hebb lists in simple span (aka the traditional Hebb task) and complex span trials. Although performance was generally lower in complex span than simple span, learning rate was similar between tasks (see also Araya et al., 2021). This result casts doubts regarding the use of attention-based maintenance processes to strengthen list representation in long-term memory. A final candidate is consolidation. Short-term consolidation involves the attention-based stabilization of the trace of just-presented information (Ricker et al., 2018). Because consolidation only operates on the last encoded information, it only benefits memory when a delay is imposed directly after information is encoded. In the present series of experiments, we observed benefits of time for learning of repeated arrays only in conditions in which time was added after encoding of each memory item—that is, only in sequential conditions. Hence, it is possible that stronger consolidation of the information indeed promoted learning. Further studies will be needed to more clearly pinpoint how exactly time may boost long-term learning of repeated arrays.

Retrieval: Number of Elements Retrieved and Testing Procedure

The last mechanism we considered was retrieval, and how differences in the testing procedure implemented in visuospatial tasks may contribute to the lack of learning previously observed. The most common means of assessing visuospatial working memory has been with the use of change-detection tasks: A probe stimulus is presented and participants have simply to indicate whether or not it matches the memorized content. To avoid issues with output interference, it has been common to test only a single element of each memory array. This procedure differs in two regards from traditional Hebb paradigms. First, it uses a recognition test instead of recall and, second, a single response is required.

To start tackling these differences, we selected a discrete color recall procedure and used it to assess the role of the number of elements tested. Across Experiments 1–8, we implemented five conditions with single-item tests. In four of them, learning was not credible (Experiments 2, 3, 4, and 6). Only one condition produced learning: when the memoranda were presented sequentially and slowly (Experiment 4). In contrast, the eight conditions that implemented recall of all items (Experiments 1, 2, 3, 5, 6, and 8) produced credible learning irrespective of variations in presentation mode and time in working memory. Experiment 6 further substantiated this finding: we replicated Experiment 3 varying the number of tests in a more fine-grained matter. We again observed learning when more than one item was tested (three or six elements tested), yet learning was weaker in the three-tests condition. Our results suggest that, with a discrete color reproduction test, learning is hindered when only a single-element is tested in most situations, but occurs robustly when all memoranda are tested.

Do our results generalize to a recognition procedure? In Experiment 7, we assessed whether increasing the number of elements tested would promote learning when memory was assessed with a recognition test. With recognition, increasing the number of elements tested did not credibly produce learning of the repeated arrays. The lack of learning was not due to low power: pooling the data of groups exposed to a single-probe and whole-array display (total N = 88) did not indicate evidence of learning (see contrast S6 vs. W6 in Table 2). Hence simply requiring participants to recognize the presented information is not conducive to Hebb learning of visuospatial arrays.

Altogether our results suggest that the critical features to unlock learning of repeated visuospatial information lies at the procedure implemented at the retrieval stage: namely (a) how many elements are tested and (b) whether participants need to recall the correct element instead of simply recognizing it. Long-term memory research has pointed to the importance of retrieval in forming stable traces (Roediger & Butler, 2011). Our findings demonstrate that the number of elements tested and the type of test is crucial also for visuospatial Hebb learning. Our results are in line with a processing view of the link between working memory and long-term memory (Craik & Lockhart, 1972).

One reason whole-array recall promotes learning could be that it favors the build-up of stronger traces of the array as an integrated configuration in long-term memory. When a single element is tested, the tested color-position association is likely strengthened, but this is not enough to generate strong learning: In the intervening unique arrays, new color-location recombinations create interference, impeding learning over trials. By contrast, unified representations of entire array configurations are unique, and these traces gain strength for repeated arrays, outcompeting traces of the unique arrays. This conclusion converges with insights from studies implying that individual item-position associations are not sufficient to explain the Hebb effect (Cumming et al., 2003; Fastame et al., 2005; Saito et al., 2020).

The role of recall for Hebb learning has been the focus of previous research on the verbal domain. Cohen and Johansson (1967) observed that learning of lists of nine digits only occurred when participants recalled the lists over the short-term but not when the lists were encoded but not recalled. Oberauer and Meyer (2009) observed learning of a repeated list of nine digits over 10 repetitions in which the list was never recalled. Yet, Hebb learning of not-recalled items was smaller than when the lists were recalled after each presentation. These results suggest that recall also contributes to the Hebb effect in the verbal domain. Overall, it is difficult to make firm comparisons between studies on the verbal and visual domain given that we are not aware of studies with verbal materials that used single-item tests, or recognition procedures. Future studies may need to assess whether the testing procedure also affects learning of verbal materials.

We can only speculate on the reasons why recognition is less suited to promote learning than recall. Several studies have suggested that people encode information differently when they expect a recall test versus a recognition test, leading to specific processing modes (Carey & Lockhart, 1973; Duncan & Murdock, 2000; Schmidt, 1983; Tversky, 1973). Accordingly, it is possible that the processing mode induced by the expectation of a recognition test may not be conducive for the creation of long-term memory traces of the memoranda. Recall has been assumed to involve a more effortful search through memory. This could be due to recall tests requiring a stronger reactivation of the cues associated with the memory trace, whereas during the recognition test the cue is already given, namely by the probe itself. The stronger cue-based memory search induced by the repeated requirements of the whole-array recall test may create the ideal conditions for the strengthening of the traces of the retrieved representations, thereby promoting learning. An alternative interpretation would be that people do learn the repeated arrays in change detection, but the recognition test is not sensitive to the acquired knowledge. Goecke and Oberauer (2021) tested for this possibility. They trained participants on three arrays of colored dots (labeled A, B, and C) until memory performance was very high (>80% correct). Afterward, participants did a change-detection task in which unique arrays were intermixed with the trained arrays and no forewarning was provided regarding the presentation of old arrays. Performance for the trained arrays was superior to performance of nontrained arrays. This indicates that if people build long-term memory representations of the memory arrays, they use them to boost their performance on change-detection tasks. In sum, their results together with ours suggest that change-detection tasks hinder the learning of the repeated arrays and not the display of the formed knowledge.

Alternative Explanations: Output Interference

One possibility to explain the evidence of learning of the repeated array in the whole-report conditions is that recall of all items require participants to retrieve information in conditions of output interference, that is, interference from preceding retrieval attempts of other items. When a single item is tested, retrieval occurs in the absence of output interference. Accumulation of output interference might be a precondition to observe learning: Only under this condition participants would bother to retrieve long-term memory representations to support their working memory performance. If this was the case, then it would indicate that participants were learning the repeated arrays in the single-test conditions, but retrieval of this information was not favored by the testing procedure. To assess this possibility, we reanalyzed data of our experiments considering only the first tested item in the whole-report experiments. Given that the first response is free from output interference, finding learning in this analysis would rule out the hypothesis that output interference explains our findings. Figures showing the data of each experiment are presented in the online supplemental materials.[3] Table 4 presents the evidence for better performance for repeated versus unique arrays in each whole-report condition in Experiments 1–8 when considering only the first output position, and aggregating across all trials.

The data of Experiment 1 is not very informative for this question because it implements a serial recall test and hence the first recalled item was always the first presented item, which led to near-ceiling performance (ca. 90% correct) in all conditions. Considering Experiments 2–8, we found general support for better performance for the repeated arrays even when considering data of only the first output position in four conditions (see Experiments 2, 6, and 8). In the remaining three conditions (Experiments 3 and 5) evidence was ambiguous. None of the experiments provided evidence against better performance for the repeated arrays. Overall, our results are inconsistent with the assumption that the lack of output interference in the single-item test prevents the manifestation of learning. Our results are more in line with the assumption that learning of the repeated array itself is hindered in conditions in which only a single element is tested.

Why Did Previous Studies Not Observe Learning of Visuospatial Arrays?

Our investigation was motivated by several studies that found no learning over many repetitions of visuospatial arrays. Our results help to explain why. First, all previous experiments not observing learning used a change-detection test, and the present Experiment 7 confirms that there is no Hebb repetition learning when visuospatial arrays are tested through change detection. Second, as is typical for change detection, each trial involved only a single memory test. Here we show that there is little, if any, Hebb learning when there is only a single test of working memory for each array.

Olson and Jiang (2004) suggested that the lack of learning of the repeated visuospatial arrays in their study may be due to two factors: (a) The precision of information in visuospatial working memory may be higher than in visual long-term memory, which may maintain only categorical information. Hence, information in long-term memory would be perceived as less useful than the one available in working memory. (b) Learning in the traditional Hebb paradigm testing verbal materials may rely on verbal traces, and hence cannot be generalized to visuospatial arrays. Here we demonstrate that these two hypotheses are lacking: Visuospatial long-term memory representations can be used to boost working memory performance provided that all items are recalled, and we observe this type of learning even in conditions assumed to favor the encoding of pure visual representations, namely simultaneous presentation conditions with brief onscreen time and a short retention interval (Sense et al., 2016). Instead, we showed that the lack of learning in their study is more likely explained by the type of test used, namely recognition.

The two experiments that did observe learning of repeated arrays at a rate comparable to the standard Hebb effect (Logie et al., 2009, Experiment 3; Shimi & Logie, 2019, Experiment 2) both tested working memory with a recall test. Shimi and Logie (2019) tested all items, thereby creating the conditions that we identified to be optimal for long-term learning. Logie et al. (2009) had people recall only one item of each array and nevertheless found substantial learning, different from our findings. One reason for that could be that in Experiment 3 of Logie et al. (2009), the same array was repeated on every trial. This probably made the repetition more obvious than in our experiments, thereby facilitating

[3] Data of Experiment 4 was not included because it only contained single-item tests.

Table 4
Evidence of Learning of the Repeated Arrays When Considering Only the First Tested Position Across the Whole-Report Conditions of All Reported Experiments

Exp.	Presentation	Timing	Whole-report test	BF_{10}
1	Sequential	Fast	Serial recall	0.33
1	Sequential	Slow	Serial recall	0.42
2	Sequential	Fast	Random	**6.29**
3	Simultaneous	Fast	Random	2.05
5	Simultaneous	Slow	Random	0.90
5	Sequential	Slow	Random	1.93
6	Simultaneous	Fast	Random (3 tests)	**18.7**
6	Simultaneous	Fast	Random	**2,850**
8	Simultaneous	Fast	Random	**1,862**

Note. Substantial evidence (BF >3) is printed in bold.

learning even with a single test. Another possibility is that learning was facilitated in their experiment because they had people recall the tested item as a verbal description of its features. A possibility to be explored in further studies is that the combination of a visual representation created at encoding with a verbal representation created at test boosts long-term learning.

The Mechanisms of Hebb Learning

What can we learn about learning from our results as well as previous findings? Here we sketch a tentative mechanism of how working memory contributes to long-term learning. Experiencing multiple repetitions of the same memory set can lead to the acquisition of two forms of long-term memory. One is a collection of separate representations of each encounter, as is the default assumption in exemplar models of memory. Separate representations of each event are necessary for episodic memory, which enables us to remember specific events. The other form of long-term memory is a single representation of the repeated memory set that is strengthened with every repetition. This kind of representation is suited for semantic memory, which represents what is common among multiple similar events, abstracting away their specifics (McClelland et al., 1995). In line with the proposal that the Hebb effect for verbal lists is a model for word-form learning (Szmalec et al., 2009)—a form of knowledge—we assume that the Hebb effect reflects the acquisition of semantic memory.

We assume that every time information is encoded into working memory, a weak trace of that information is laid down in episodic long-term memory as well. It could form the seed of a semantic-memory representation if further encounters with the same information further strengthen this representation, rather than just creating a new episodic representation. At this point we borrow an assumption from REM.3 (Ensor et al., 2021), a version of the REM (retrieving efficiently from memory) model (Shiffrin & Steyvers, 1997): If a repeatedly encoded stimulus is recognized as having been encoded before, then the memory trace of that earlier encounter is strengthened. In the absence of recognition, however, a new episodic memory trace is created. Forming separate traces of the same stimulus can be beneficial for tests of episodic memory because it creates multiple chances of accessing at least one trace of that stimulus (Ensor et al., 2021), but it is not a good basis for the acquisition of knowledge, and therefore not sufficient for generating Hebb repetition learning.

Evidence for this set of assumptions comes from the finding that people recognize repeated arrays in an explicit recognition test at the end of the experiment, even when their change-detection performance did not improve at all through the repetitions (Fukuda & Vogel, 2019; Olson et al., 2005; Olson & Jiang, 2004). Evidently, some episodic-memory trace was created for the arrays, but it was not sufficient to cause Hebb learning.

Why then does testing memory through recall, and in particular through tests of multiple array items, enable Hebb learning? We consider two—not mutually exclusive—hypotheses. Both build on the assumption that retrieval of information from long-term memory is controlled by a flexible gate that is opened only when the person believes that long-term memory information is likely to be helpful for the current task (Oberauer, 2009). This assumption is supported by previous work on visual working memory, which suggests that people draw on episodic memory representations to help in a working memory test only if they find the information in working memory insufficient (Oberauer et al., 2017). This is more likely to be the case the harder the test—or at least, the harder the person believes the test to be. We argue that people are more likely to find the information in working memory insufficient in a recall test than in change detection, and that likelihood increases with tests of multiple items for two reasons: Each attempt to recall an item from working memory is a new chance of experiencing failure. In addition, output interference makes each additional recall attempt in a trial harder and harder. As a consequence, people are more likely to try to access their episodic memory of the current array in a recall test than in a change-detection test, and more so the more items are tested.

Trying to draw on an episodic-memory representation of the current array has two consequences. The first is that it creates an opportunity for practicing retrieval of that representation, which facilitates its subsequent retrieval. This makes it more likely that the episodic representation will be retrieved when the same array is repeated a few trials later, and hence, increases the chance of recognition of the repeated array. The second consequence is that an attempt to retrieve the episodic-memory trace of the current array makes it more likely that an episodic-memory trace of a previous occurrence of the same array is retrieved, thereby enabling recognition of the repeated array. Both causal routes increase the probability of recognizing a repeated array as such, and thereby increase the chance of cumulatively strengthening its representation in long-term memory, starting the trajectory toward the formation of semantic memory.

Our proposal leads to the following prediction that could be tested in future work: Other manipulations that make access to representations in working memory more difficult—without at the same time impairing encoding into episodic memory—should also improve Hebb repetition learning. Examples of such manipulations could be a longer retention interval (Pertzov et al., 2017) or a distractor task requiring central attention in the retention interval (Makovski et al., 2006; Souza & Oberauer, 2017).

Conclusion

Across eight experiments, we observed that people can learn repeatedly encountered visuospatial arrays when they are required to recall all learned information. Characteristics of how working memory is tested appear to be the strongest determinants of long-term memory. We propose that more difficult tests lead people to try more often to draw on episodic long-term memory in a working memory task, which creates opportunities for retrieval practice and increases the probability of recognizing earlier occurrences of a repeated set of items. This, in turn, makes cumulative learning of repeated sets more likely.

Context of Research

Our research program examines the capacity limits of working memory and how long-term memory can be used to alleviate the burden on this capacity. We are looking at different paradigms assessing the interplay of these two forms of memory, including the use of strategies such as elaboration and chunking, as well as Hebb learning. We were intrigued by the lack of learning of repeated visuospatial arrays and how it defied the generalizability of the classic Hebb-repetition effect. In one group meeting, we listed all differences between studies showing the typical Hebb effect, and studies in the visuospatial working memory literature not finding any learning, and we made individual prior bets on which factors were most likely to explain their divergent results. We chose the within-subject manipulations of Experiments 1–5 to maximize their expected information gain (variables with pooled priors $\geq$ .5). Experiments 6–8 were designed in response to criticisms raised at discussion of the results of these initial experiments. We continue to be intrigued by the conditions that favor visuospatial learning. For example, we are currently investigating whether single-item tests prevent learning, or just prevent people from using it (Musfeld et al., 2021). We also started investigating the role of awareness of the repetition for Hebb learning (Musfeld et al., 2022). The present work contributes one piece to the puzzle that permits to understand how information flows from working memory to long-term memory and vice-versa. Yet, mapping exactly how they interact remains a challenge for further empirical and computational work.

References

Adam, K. C. S., & Vogel, E. K. (2017). Confident failures: Lapses of working memory reveal a metacognitive blind spot. *Attention, Perception & Psychophysics*, *79*(5), 1506–1523. https://doi.org/10.3758/s13414-017-1331-8

Araya, C., Oberauer, K., & Saito, S. (2021). The Hebb repetition effect in complex span tasks: Evidence for a shared learning mechanism with simple span tasks. *Memory & Cognition*. Advance online publication. https://doi.org/10.3758/s13421-021-01261-3

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. 90–195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., van Ravenzwaaij, D., White, C. N., De Boeck, P., & Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2607–2612. https://doi.org/10.1073/pnas.1708285114

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Bürkner, P.-C. (2017). brms: An R Package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Camos, V., Johnson, M., Loaiza, V., Portrat, S., Souza, A., & Vergauwe, E. (2018). What is attentional refreshing in working memory? *Annals of the New York Academy of Sciences*, *1424*(1), 19–32. https://doi.org/10.1111/nyas.13616

Carey, S. T., & Lockhart, R. S. (1973). Encoding differences in recognition and recall. *Memory & Cognition*, *1*(3), 297–300. https://doi.org/10.3758/BF03198112

Cohen, R. L., & Johansson, B. S. (1967). The activity trace in immediate memory: A re-evaluation. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 139–143. https://doi.org/10.1016/S0022-5371(67)80064-1

Couture, M., & Tremblay, S. (2006). Exploring the characteristics of the visuospatial Hebb repetition effect. *Memory & Cognition*, *34*(8), 1720–1729. https://doi.org/10.3758/BF03195933

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*(2), 159–180. https://doi.org/10.1037/0096-3445.125.2.159

Cumming, N., Page, M., & Norris, D. (2003). Testing a positional model of the Hebb effect. *Memory*, *11*(1), 43–63. https://doi.org/10.1080/741938175

Cunningham, T. F., Healy, A. F., & Williams, D. M. (1984). Effects of repetition on short-term retention of order information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 575–597. https://doi.org/10.1037/0278-7393.10.4.575

Duncan, M., & Murdock, B. (2000). Recognition and recall with precuing and postcuing. *Journal of Memory and Language*, *42*(3), 301–313. https://doi.org/10.1006/jmla.1999.2690

Ensor, T. M., Surprenant, A. M., & Neath, I. (2021). Modeling list-strength and spacing effects using version 3 of the retrieving effectively from memory (REM.3) model and its superimposition-of-similar-images assumption. *Behavior Research Methods*, *53*(1), 4–21. https://doi.org/10.3758/s13428-019-01324-z

Fastame, M. C., Flude, B., & Hitch, G. J. (2005). How is the serial order of a verbal sequence coded? Some comparisons between models. *Memory*, *13*(3–4), 247–258. https://doi.org/10.1080/09658210344000314

Frick, R. W. (1985). Testing visual short-term memory: Simultaneous versus sequential presentations. *Memory & Cognition*, *13*(4), 346–356. https://doi.org/10.3758/BF03202502

Fukuda, K., & Vogel, E. K. (2019). Visual short-term memory capacity predicts the "bandwidth" of visual long-term memory encoding. *Memory & Cognition*, *47*(8), 1481–1497. https://doi.org/10.3758/s13421-019-00954-0

Gagnon, S., Bédard, M.-J., & Turcotte, J. (2005). The effect of old age on supra-span learning of visuo-spatial sequences under incidental and intentional encoding instructions. *Brain and Cognition*, *59*(3), 225–235. https://doi.org/10.1016/j.bandc.2005.07.001

Goecke, B., & Oberauer, K. (2021). Is long-term memory used in a visuospatial change-detection paradigm? *Psychonomic Bulletin & Review*, *28*(6), 1972–1981. https://doi.org/10.3758/s13423-021-01951-8

Hartshorne, J. K., & Makovski, T. (2019). The effect of working memory maintenance on long-term memory. *Memory & Cognition*, *47*(4), 749–763. https://doi.org/10.3758/s13421-019-00908-6

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning* (pp. 37–46). Blackwell.

Henninger, F., Shevchenko, Y., Mertens, U., Kieslich, P. J., & Hilbig, B. E. (2020). *lab.js: A free, open, online experiment builder*. Zenodo. https://doi.org/10.5281/zenodo.3767907

Hitch, G. J., Flude, B., & Burgess, N. (2009). Slave to the rhythm: Experimental tests of a model for verbal short-term memory and long-term sequence learning. *Journal of Memory and Language*, *61*(1), 97–111. https://doi.org/10.1016/j.jml.2009.02.004

Horton, N., Hay, D. C., & Smyth, M. M. (2008). Hebb repetition effects in visual memory: The roles of verbal rehearsal and distinctiveness. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *61*(12), 1769–1777. https://doi.org/10.1080/17470210802168674

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*(6), 685–701. https://doi.org/10.1016/0749-596X(91)90032-F

Johnson, A. J., & Miles, C. (2019a). Visual Hebb repetition effects: The role of psychological distinctiveness revisited. *Frontiers in Psychology*, *10*, 17. https://doi.org/10.3389/fpsyg.2019.00017

Johnson, A. J., & Miles, C. (2019b). Visual Hebb repetition effects survive changes to both output order and concurrent articulation. *Journal of Cognitive Psychology*, *31*(3), 1–9. https://doi.org/10.1080/20445911.2019.1586715

Johnson, A. J., Dygacz, A., & Miles, C. (2017). Hebb repetition effects for non-verbal visual sequences: Determinants of sequence acquisition. *Memory*, *25*(9), 1279–1293. https://doi.org/10.1080/09658211.2017.1293692

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*(2), 573–603. https://doi.org/10.1037/a0029146

Loaiza, V. M., & McCabe, D. P. (2012). Temporal-contextual processing in working memory: Evidence from delayed cued recall and delayed free recall tests. *Memory & Cognition*, *40*(2), 191–203. https://doi.org/10.3758/s13421-011-0148-2

Logie, R. H., Brockmole, J. R., & Vandenbroucke, A. R. E. (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

Makovski, T., Shim, W. M., & Jiang, Y. V. (2006). Interference from filled delays on visual change detection. *Journal of Vision*, *6*(12), 1459–1470. https://doi.org/10.1167/6.12.11

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

Mızrak, E., & Oberauer, K. (2021). What is time good for in working memory? *Psychological Science*, *32*(8), 1325–1337. https://doi.org/10.1177/0956797621996659

Musfeld, P., Souza, A. S., & Oberauer, K. (2021, November). *Testing working memory boosts long-term learning of visual arrays [Poster presentation]. 62nd Annual Meeting of the Psychonomic Society, Virtual.*

Musfeld, P., Souza, A. S., & Oberauer, K. (2022, March). *Chicken or egg? The role of awareness in long-term learning by repeated exposure [Talk]. 64th TeaP Conference, Virtual. University of Cologne.*

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation*, *51*, 45–100. https://doi.org/10.1016/S0079-7421(09)51002-X

Oberauer, K., & Meyer, N. (2009). The contributions of encoding, retention, and recall to the Hebb effect. *Memory*, *17*(7), 774–781. https://doi.org/10.1080/09658210903107861

Oberauer, K., Awh, E., & Sutterer, D. W. (2017). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *43*(1), 1–22. https://doi.org/10.1037/xlm0000302

Oberauer, K., Jones, T., & Lewandowsky, S. (2015). The Hebb repetition effect in simple and complex memory span. *Memory & Cognition*, *43*(6), 852–865. https://doi.org/10.3758/s13421-015-0512-8

Olson, I. R., & Jiang, Y. (2004). Visual short-term memory is not improved by training. *Memory & 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

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1536), 3737–3753. https://doi.org/10.1098/rstb.2009.0173).

Page, M. P. A., Cumming, N., Norris, D., Hitch, G. J., & McNeil, A. M. (2006). Repetition learning in the immediate serial recall of visual and auditory materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 716–733. https://doi.org/10.1037/0278-7393.32.4.716

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 427–442. https://doi.org/10.1163/156856897X00366

Pertzov, Y., Manohar, S., & Husain, M. (2017). Rapid forgetting results from competition over time between items in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 528–536. https://doi.org/10.1037/xlm0000328

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Ricker, T. J., & Cowan, N. (2014). Differences between presentation methods in working memory procedures: A matter of working memory consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 417–428. https://doi.org/10.1037/a0034301

Ricker, T. J., & Hardman, K. O. (2017). The nature of short-term consolidation in visual working memory. *Journal of Experimental Psychology: General*, *146*(11), 1551–1573. https://doi.org/10.1037/xge0000346

Ricker, T. J., Nieuwenstein, M. R., Bayliss, D. M., & Barrouillet, P. (2018). Working memory consolidation: Insights from studies on attention and working memory. *Annals of the New York Academy of Sciences*, *1424*(1), 8–18. https://doi.org/10.1111/nyas.13633

Roediger, H. L. III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Saito, S., Nakayama, M., & Tanida, Y. (2020). Verbal working memory, long-term knowledge, and statistical learning. *Current Directions in Psychological Science*, *29*(4), 340–345. https://doi.org/10.1177/0963721420920383

Schmidt, S. R., (1983). The effects of recall and recognition test expectancies on the retention of prose. *Memory & Cognition*, *11*(2), 172–180. https://doi.org/10.3758/BF03213472

Sense, F., Morey, C. C., Prince, M., Heathcote, A., & Morey, R. D. (2016). Opportunity for verbalization does not improve visual change detection performance: A state-trace analysis. *Behavior Research Methods*, *49*(3), 853–862. https://doi.org/10.3758/s13428-016-0741-1

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/BF03209391

Shimi, A., & Logie, R. H. (2019). Feature binding in short-term memory and long-term learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *72*(6), 1387–1400. https://doi.org/10.1177/1747021818807718

Sjöblom, A., & Hughes, R. W. (2020). The articulatory determinants of verbal sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(10), 1977–1997. https://doi.org/10.1037/xlm0000926

Souza, A. S., & Oberauer, K. (2017). Time to process information in working memory improves episodic memory. *Journal of Memory and Language*, *96*, 155–167. https://doi.org/10.1016/j.jml.2017.07.002

Souza, A. S., & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall? *Cognitive Psychology*, *107*, 1–21. https://doi.org/10.1016/j.cogpsych.2018.09.002

Souza, A. S., & Oberauer, K. (2020). No evidence that articulatory rehearsal improves complex span performance. *Journal of Cognition*, *3*(1), 11. https://doi.org/10.5334/joc.103

Souza, A. S., Czoschke, S., & Lange, E. B. (2020). Gaze-based and attention-based rehearsal in spatial working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(5), 980–1003. https://doi.org/10.1037/xlm0000771

Sukegawa, M., Ueda, Y., & Saito, S. (2019). The effects of Hebb repetition learning and temporal grouping in immediate serial recall of spatial location. *Memory & Cognition*, *47*(4), 643–657. https://doi.org/10.3758/s13421-019-00921-9

Sutterer, D. W., & Awh, E. (2016). Retrieval practice enhances the accessibility but not the quality of memory. *Psychonomic Bulletin & Review*, *23*(3), 831–841. https://doi.org/10.3758/s13423-015-0937-x

Szmalec, A., Duyck, W., Vandierendonck, A., Mata, A. B., & Page, M. P. A. (2009). The Hebb repetition effect as a laboratory analogue of novel word learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *62*(3), 435–443. https://doi.org/10.1080/17470210802386375

Szmalec, A., Page, M. P. A., & Duyck, W. (2012). The development of long-term lexical representations through Hebb repetition learning. *Journal of Memory and Language*, *67*(3), 342–354. https://doi.org/10.1016/j.jml.2012.07.001

Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review*, *15*(3), 535–542. https://doi.org/10.3758/PBR.15.3.535

Tremblay, S., & Saint-Aubin, J. (2009). Evidence of anticipatory eye movements in the spatial Hebb repetition effect: Insights for modeling sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1256–1265. https://doi.org/10.1037/a0016566

Tversky, B. (1973). Encoding processes in recognition and recall. *Cognitive Psychology*, *5*(3), 275–287. https://doi.org/10.1016/0010-0285(73)90037-6

Wood, K., & Simons, D. J. (2017). Reconciling change blindness with long-term memory for objects. *Attention, Perception & Psychophysics*, *79*(2), 438–448. https://doi.org/10.3758/s13414-016-1240-2

Xie, W., & Zhang, W. (2017). Familiarity increases the number of remembered Pokémon in visual short-term memory. *Memory & Cognition*, *45*(4), 677–689. https://doi.org/10.3758/s13421-016-0679-7

Zanto, T. P., Clapp, W. C., Rubens, M. T., Karlsson, J., & Gazzaley, A. (2016). Expectations of task demands dissociate working memory and long-term memory systems. *Cerebral Cortex*, *26*(3), 1176–1186. https://doi.org/10.1093/cercor/bhu307

Received January 22, 2021
Revision received March 8, 2022
Accepted March 17, 2022 ■