



# Encoding strategies in self-initiated visual working memory

Hagit Magen<sup>1</sup> · Anat Berger-Mandelbaum<sup>1</sup>

Published online: 11 June 2018  
© Psychonomic Society, Inc. 2018

## Abstract

During a typical day, visual working memory (VWM) is recruited to temporarily maintain visual information. Although individuals often memorize external visual information provided to them, on many other occasions they memorize information they have constructed themselves. The latter aspect of memory, which we term *self-initiated WM*, is prevalent in everyday behavior but has largely been overlooked in the research literature. In the present study we employed a modified change detection task in which participants constructed the displays they memorized, by selecting three or four abstract shapes or real-world objects and placing them at three or four locations in a circular display of eight locations. Half of the trials included identical targets that participants could select. The results demonstrated consistent strategies across participants. To enhance memory performance, participants reported selecting abstract shapes they could verbalize, but they preferred real-world objects with distinct visual features. Furthermore, participants constructed structured memory displays, most frequently based on the Gestalt organization cue of symmetry, and to a lesser extent on cues of proximity and similarity. When identical items were selected, participants mostly placed them in close proximity, demonstrating the construction of configurations based on the interaction between several Gestalt cues. The present results are consistent with recent findings in VWM, showing that memory for visual displays based on Gestalt organization cues can benefit VWM, suggesting that individuals have access to metacognitive knowledge on the benefit of structure in VWM. More generally, this study demonstrates how individuals interact with the world by actively structuring their surroundings to enhance performance.

**Keywords** Visual working memory · Self-initiation · Gestalt · Grouping · Metacognition

Visual working memory (VWM), the visual component of the WM model (Baddeley & Hitch, 1974), is a limited-capacity system in charge of maintaining visual information for short periods of time (see Luck & Vogel, 2013, and Ma, Husain, & Bays, 2014, for reviews). Contrary to initial conceptualizations of a fixed limit, it is well-established that the limit of VWM varies considerably with the complexity of stimuli, demonstrating superior performance for structured displays (Brady, Konkle, & Alvarez, 2011; Orhan & Jacobs, 2014). Given that our natural surroundings are mostly structured, everyday memory for structured external information is therefore highly efficient. Nevertheless, memory for external or provided information is only one aspect of memory in every-

day settings. In many daily activities, memory is self-initiated rather than provided externally, meaning that individuals memorize information they have selected and structured themselves.

Self-initiated VWM may be recruited in different everyday scenarios. For instance, during a typical day, individuals often place objects such as keys, glasses, or a cup of coffee in different locations and retrieve them shortly afterward. Before performing complex tasks such as cooking, individual temporarily organize the tools or ingredients they will need and memorize the items and their locations, so they can easily retrieve the items during task performance. Other situations also involve the selection of the to-be-memorized visual information, such as when browsing the aisles of a store and memorizing items that are retrieved shortly after. Contrary to laboratory tasks in which individuals have no control over the information they memorize, in all these everyday scenarios individuals actively construct the information that they memorize. Thus, during ongoing behavior, individuals initiate the memory processes and select the

---

✉ Hagit Magen  
msmagen@mail.huji.ac.il

<sup>1</sup> The Hebrew University, P.O. Box 24026, Mount Scopus, 91240 Jerusalem, Israel

information they memorize, and therefore we have termed this aspect of memory *self-initiated WM*.

Although self-initiated WM in everyday behavior is a complex process that is performed continuously in dynamic surroundings, our initial investigation focused on basic processes, laying the ground for future studies that will capture its more complex nature. Our approach has been to follow the tradition of the WM literature by using modified well-known WM tasks. A recent study from our lab investigated the processes involved in spatial self-initiated WM (Magen & Emmanouil, 2018), and in the present study we explored the visual aspect of self-initiated WM using a modified change detection task.

Our basic assumption is that when individuals are in control of encoding, they construct memory displays in an attempt to maximize memory performance. Nevertheless, they may lack the necessary metacognitive knowledge about what constitutes efficient memory representations in VWM, leaving self-initiated VWM at a disadvantage. In a recent study from our lab on spatial self-initiated WM, we used a modified spatial span task in which participants constructed the spatial sequences they memorized (Magen & Emmanouil, 2018). The results demonstrated that participants constructed structured spatial sequences following careful planning. The reaction time (RT) for the first location in the sequence increased with set size and was higher than the RTs for subsequent locations, which demonstrated that participants planned the spatial sequences before executing their selections. Specifically, participants constructed sequences that, relative to random sequences, minimized the distances between successive locations, and followed simple and linear shapes. The spatial structures that participants constructed had been identified in the literature on provided spatial WM, as structures that benefited memory performance (e.g., Kemps, 2001; Parmentier, Elford, & Maybery, 2005). Thus, the participants in the self-initiated spatial span task revealed metacognitive knowledge on ideal spatial memory displays.

The present study takes a further step in understanding the processes and representations of self-initiated VWM, by using a modified change detection task in which participants constructed the memory displays they memorized. Our main assumption was that participants would attempt to maximize performance by constructing structured memory representations. To our knowledge, the basic representations of self-initiated VWM have not yet been studied, and therefore we review the literature on external provided VWM, from which we derive our predictions regarding the structure of VWM representations.

A growing number of studies in recent years have illustrated the benefits of organizing memory displays in VWM by Gestalt organization cues (e.g., Gao, Gao, Tang, Shui, & Shen, 2015; Peterson & Berryhill, 2013; van Lamsweerde, Beck, & Johnson, 2016). Gestalt organization cues (e.g., proximity,

similarity, symmetry, good continuation) represent perhaps the most fundamental structures in our environment (Wagemans et al., 2012; Wertheimer, 1923/2012). During perception, Gestalt organization cues allow independent elements to be grouped effortlessly into coherent objects, and figures to be distinguished from their background (Wagemans et al., 2012). Organized elements in the visual scene enjoy prioritized processing, as attention is biased toward them. The evidence suggests that perceptual organization enjoys prioritized processing even if one is not aware of the organized elements (Kimchi, 2009; Kimchi, Yeshurun, & Cohen-Savransky, 2007). The benefits of organization extend to postperceptual processes of VWM, since memory performance improves when items in the memory display are organized on the basis of one or more Gestalt organization cues (e.g., Gao et al., 2015; Peterson & Berryhill, 2013; van Lamsweerde et al., 2016).

Several studies have demonstrated that memory accuracy increased when items in the memory display (e.g., colors, shapes, faces) were grouped by Gestalt cues of proximity, similarity, or their combination (Jiang, Lee, Asaad & Remington, 2016; Lin & Luck, 2009; Morey, Cong, Zheng, Price, & Morey, 2015; Peterson & Berryhill, 2013; Peterson, Gözenman, Arciniega, & Berryhill, 2015; Quinlan & Cohen, 2012; van Lamsweerde et al., 2016; Woodman, Vecera, & Luck, 2003). For example, Peterson and Berryhill found superior memory for displays of color targets that were positioned next to each other, allowing grouping of the memory targets by proximity. Memory for displays in which one of the memory targets repeated, thereby allowing grouping by similarity, improved as well. The benefit was restricted to the grouped targets, and did not extend to ungrouped targets. Furthermore, grouping by Gestalt cues of proximity and similarity interacted, such that similarity benefited performance only when the similar items were placed next to each other (see, however, Morey et al., 2015; Peterson et al., 2015).

Connectedness influenced VWM, as well, since recall of connected targets was superior relative to memory of unconnected items (van Lamsweerde et al., 2016; Woodman et al., 2003; Xu, 2006). Organizational cues of collinearity, closure (Gao et al., 2015), and common fate (Luria & Vogel, 2014) also benefited VWM performance. Symmetry, an additional Gestalt organization cue, has been studied as well but to a lesser extent. Memory performance for visual patterns (Rossi-Arnaud, Pieroni & Baddeley, 2006; Rossi-Arnaud, Pieroni, Spataro, & Baddeley, 2012) or spatial configurations (Kemps, 2001) was enhanced when participants encoded and maintained symmetrical visual arrays.

Studies that have examined the role of structure in VWM most often have presented the memory targets simultaneously. However, when items are presented simultaneously, the

benefit due to organization may depend entirely on efficient perceptual processing. Gao et al. (2015) examined this issue by presenting items in the memory display sequentially and ensuring that grouping could only occur postperceptually within VWM. Their study revealed substantial benefits of grouping on memory performance, demonstrating that grouping exerted its impact postperceptually within VWM.

Although it is well-established that grouping can benefit WM performance, the underlying mechanisms have yet to be clarified. Information in structured displays is compressed into higher order configurations (i.e., chunks; Miller, 1956), reducing the amount of information that is encoded in memory and effectively increasing its capacity. Accordingly, neural studies (functional magnetic resonance imaging [fMRI] and electroencephalography [EEG]) have revealed reduced activity during the maintenance period of VWM tasks, when items in the memory display were grouped by one or more Gestalt cues (Gao et al., 2011; Luria & Vogel, 2014; Peterson et al., 2015; Xu & Chun, 2007). Thus, organization in VWM seems to benefit memory by reducing the required cognitive and neural resources during maintenance.

The main objective of the present study was to explore the structure of the memory displays that participants would construct when they were in control during encoding. The construction of efficient displays requires metacognitive knowledge on the benefit of structure in VWM, and in particular on the benefit of Gestalt organizational cues. We are unaware of any studies that have explored how individuals select and implement organizational cues to construct memory representations in VWM. Nevertheless, one study exploring VWM for provided information suggested that when several cues were available in the memory display, participants exerted top-down control and grouped items by the most beneficial organizational cue (van Lamsweerde et al., 2016). The authors presented participants with displays of colors and shapes, in which intradimension targets were grouped by similarity and cross-dimension color and shape targets were grouped by connection cues. Across participants, either one dimension or both dimensions were task relevant. The results demonstrated that participants strategically utilized only the cue that benefited performance. Participants grouped items on the basis of similarity when only one dimension was task-relevant, but grouped items on the basis of connection cues when both dimensions were task-relevant. Although the construction of self-initiated displays requires more elaborate processes, this study suggests that grouping in VWM is not automatic but can be under top-down control.

## The present study

In the present study we explored the metacognitive knowledge of participants regarding the beneficial effects

of structure on VWM, as we observed in spatial self-initiated WM (Magen & Emmanouil, 2018). More specifically, deriving our predictions from the literature on provided VWM, we anticipated that if participants hold that knowledge, they would construct structured memory representations based on Gestalt organizational cues. The study employed a modified change detection task in which participants were presented with eight visual targets and a circular display of eight locations. They first selected three or four targets (set sizes were chosen to be within the capacity of VWM), and then placed them in three or four locations of their choice. We also addressed the question of whether participants would consider grouping the memory targets by similarity, an organization cue that also benefits memory performance (e.g., Peterson & Berryhill, 2013). Therefore, on half of the trials one of the targets repeated, allowing participants to select and memorize two identical targets (following previous studies, we refer to this organization cue as *similarity*, although the targets were identical).

The task was self-paced, providing participants with enough time to familiarize themselves with the memory targets before they selected and placed them. Thus, we assumed that if participants constructed structured memory representations, they would do so in order to benefit maintenance processes more than encoding processes. At the end of the experiment, half of the participants filled out a questionnaire asking them to detail the strategies that guided them during encoding. Participants were also asked whether they employed similar memory strategies and processes in their everyday life.

The focus of the study was the structure of the representations that participants constructed, and the nature of the visual targets they selected. Nevertheless, we also analyzed RTs during encoding to explore the selection process in more depth. Specifically, following the results from the modified spatial span task (Magen & Emmanouil, 2018), we expected that RT for the first target and the first location that participants selected would demonstrate whether they planned the sequences of targets and locations before selecting them. Finally, we also report accuracy measures in the different conditions. To generalize our findings across different types of materials, we tested two change detection tasks, one in which participants memorized familiar and meaningful real-world objects, and another in which participants memorized unfamiliar and meaningless abstract shapes. We explored whether participants who encoded arbitrary stimuli would adopt the same strategies as participants who encoded meaningful stimuli—speculating, for example, that participants might associate structure more readily with real-world objects.

## Method

### Participants

Eighty students participated in a 1-h session. All reported having normal or corrected-to-normal vision. Participants provided informed consent before participating in the study for course credit or payment.

### Stimuli and design

Participants sat in a dimly lit room at a distance of 100 cm from a 17-in. CRT monitor and rested their head on a chin rest. The memory targets consisted of 200 pictures of real-world objects and 200 abstract shapes. We used a large stimulus set to allow participants to focus on the structure of the memory displays, without being biased toward frequently reoccurring memory targets. Nevertheless, we learned about the strategies participants used to select the visual targets from their responses to the strategy questionnaire (see below). The type of stimuli that participants memorized was manipulated between participants, in case different strategies might be utilized when real-world objects and abstract shapes were selected.

At the beginning of each trial, eight targets (abstract shapes or real-world objects), each measuring  $0.85^\circ \times 0.85^\circ$  of visual angle, were presented at the top part of the monitor (see Fig 1). Half of the trials consisted of eight unique targets, whereas on the remaining trials two of the eight targets were identical. The identical targets were placed in random locations in the target array. Another display of eight circles (each with a radius of  $0.31^\circ$  of visual angle) was presented below the target display. The circles were spaced equally on an imaginary circle with a radius of  $1.72^\circ$  of visual angle. The circular location display did not appear during the delay or retrieval phases, so that we could explore whether participants would construct structured

memory displays even when external location cues were absent following the encoding phase. The set size varied between three or four targets, and the set size manipulation was blocked. Participants were informed of the current set size at the beginning of each block.

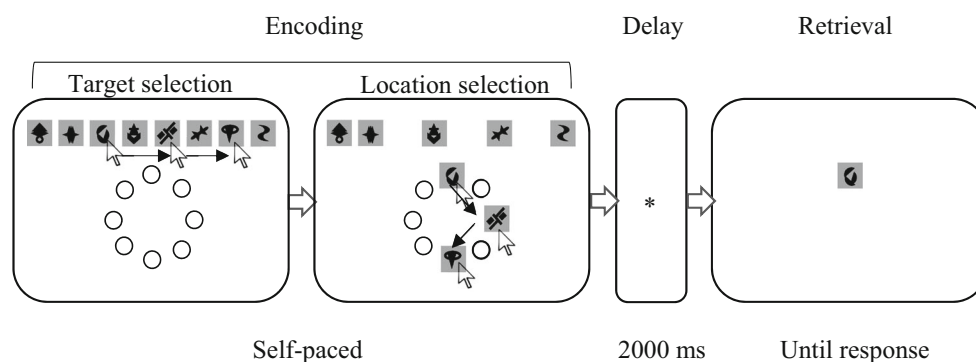
During retrieval, a single probe appeared in one of the selected locations. Participants decided whether the probe was the same as the target that they had placed in that location during encoding. Fifty percent of the trials were match, and the remaining trials were nonmatch. In match trials, the probe matched the target that had been placed in that location, whereas in nonmatch trials, the probe was one of the targets that had been placed in one of the other locations. To respond correctly, participants needed to memorize both the items and the locations in which they were placed.

We assumed that discrimination between the abstract shapes in the target array would be more difficult than discrimination between the real-world objects. Thus, we were concerned that participants might miss the repeated abstract shapes more often than the repeated real-world objects. Therefore, for each stimulus type, half of the participants were informed that two identical targets would appear on half of the trials.

Overall, set size and identity (whether two identical targets appeared in the target display) were manipulated within participants, whereas stimulus type and instructions were manipulated between participants.

### Strategy questionnaire

At the end of the experiment, half of the participants filled out a questionnaire regarding the different strategies they had used during encoding. First, participants were asked about the strategies they had used to select the visual targets and whether these strategies had benefited memory performance. Second, participants were asked whether they had noticed the identical



**Fig. 1.** Illustration of the change detection task used in this study. At the beginning of each trial, arrays of eight targets and eight locations were presented. Participants first selected three or four targets in sequence by clicking on them with the left key of the mouse, and then they placed the targets in the same order, by clicking on three or four locations (the

selection process is represented in the figure by arrows). Following a delay of 2,000 ms, participants were presented with a single probe in one of the locations that had been selected during encoding and decided whether the probe matched the target that they had placed in that location. The figure illustrates a match trial



targets in the visual target display. If they had noticed the identical targets, they were asked how often they had selected the two targets and whether a display with identical targets benefited memory performance. Third, participants were asked about the strategies they had used in placing the targets in the circular display and whether these strategies had benefited memory. Finally, participants were asked whether they employed similar memory strategies and processes in their everyday life.

## Procedure

Each trial began with the appearance of a fixation point for 500 ms, followed by the target and location displays. Participants first selected three or four targets sequentially and then placed them sequentially in three or four out of the eight possible locations. The selection of the targets and locations followed the same order, such that after all the targets had been selected, the first target that was selected was placed first at one of the eight locations. Participants could not go back to change their selections, and therefore were required to plan the sequences of visual targets and locations before their execution. This aspect of the design allowed us to obtain a measure of planning in terms of the structure of the memory display (i.e., planning memory representations based on metacognitive knowledge) and RTs. After the last location was selected, the items remained on the screen for an additional 500 ms and disappeared. A delay period of 2,000 ms, during which the fixation point appeared at the center of the screen, followed encoding and was followed in turn by the appearance of a single probe that occupied one of the selected locations. The probe remained on the screen until the participant had responded or until 2,000 ms had elapsed. If a response was not registered within 2,000 ms, the probe was replaced by the fixation point, until response. Participants responded to the probe on a keyboard, pressing the “1” key for the match condition and the “2” key for the nonmatch condition.

The experiment began with eight practice trials, followed by six experimental blocks, three for each set size. The number of trials in each block depended on the set size, in order to allow equal numbers of probes in each selected location. Therefore, each block consisted of 36 trials of set size 3 and 32 trials of set size 4. The blocks were interleaved, and block order was counterbalanced across participants. Accuracy was stressed in all of the conditions. An error message was presented on screen for 500 ms following an incorrect response. In any case, the intertrial interval was 1,500 ms. Half of the participants (ten in each condition of stimulus type and instruction) filled out the strategy questionnaire at the end of the experiment.

## Results

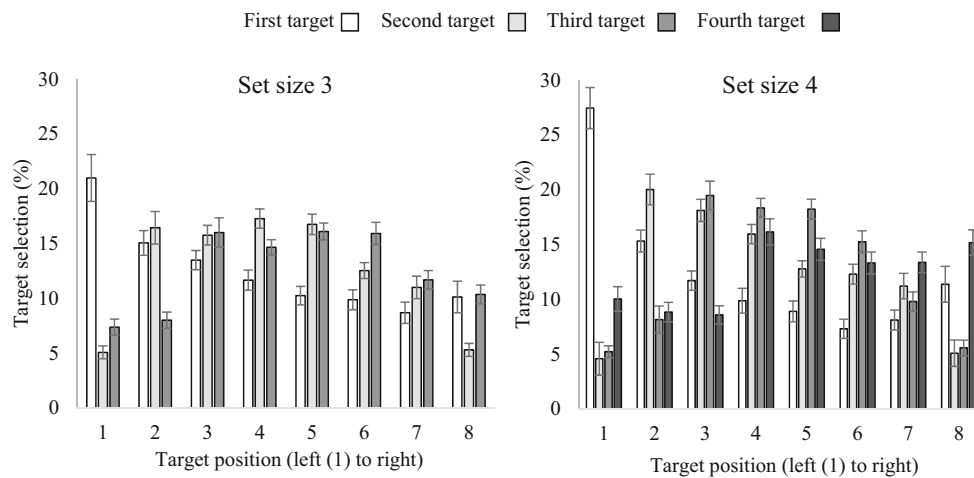
This section is divided into two main parts. The first focuses on encoding processes, on the targets and locations participants selected, and on encoding RTs. The second focuses on the responses to the strategy questionnaire. The accuracy data are presented in the [Appendix](#). Three types of trials were defined: *no-repetition* trials, in which only unique items were presented; *repetition-unselected* trials, in which two identical items were presented but only unique targets were selected; and *repetition-selected* trials, in which identical items were presented and the two identical items were selected.

## Encoding

**Target selection** Participants selected three or four visual targets in each trial from a horizontal display of eight visual targets. The targets themselves were drawn from a large pool of potential targets and most likely changed from trial to trial. Nevertheless, we examined whether participants were biased toward certain target positions within the target display. [Figure 2](#) depicts the percentages of visual targets that were selected at each of the eight target positions, as a function of serial position (the order in which the visual targets were selected). The results are presented separately for each set size (collapsed across the three repeated conditions, which yielded similar results).

As is illustrated in [Fig. 2](#), targets from the leftmost position were selected first most often, whereas the second target was selected most frequently from Positions 2–4. This pattern shows that participants tended to initiate their selections on the left side and to move rightward. Calculating the overall percentage of time that each target position was selected showed that the selections were somewhat biased toward the left and central targets. The overall percentages for each target position (left to right) were 11.23, 13.25, 15.14, 14.13, 14.11, 12.70, 10.54, and 8.90 for set size 3, and 11.81, 13.08, 14.47, 15.07, 13.61, 12.04, 10.6, and 9.29 for set size 4, respectively. For each set size, we performed a chi-square goodness-of-fit test on the number of selections in each target position against a uniform distribution. The Bonferroni correction for multiple comparisons was used to correct for the number of comparisons, with a corrected threshold of .025 (.05/2 comparisons). The results demonstrated significant deviations from a uniform distribution for set size 3,  $\chi^2(7) = 749.34$ ,  $p < .001$ , and set size 4,  $\chi^2(7) = 653.31$ ,  $p < .001$ . Thus, participants were spatially biased in their selections, although in their responses to the strategy questionnaire they reported that they had selected the visual targets on the basis of nonspatial features (see below).

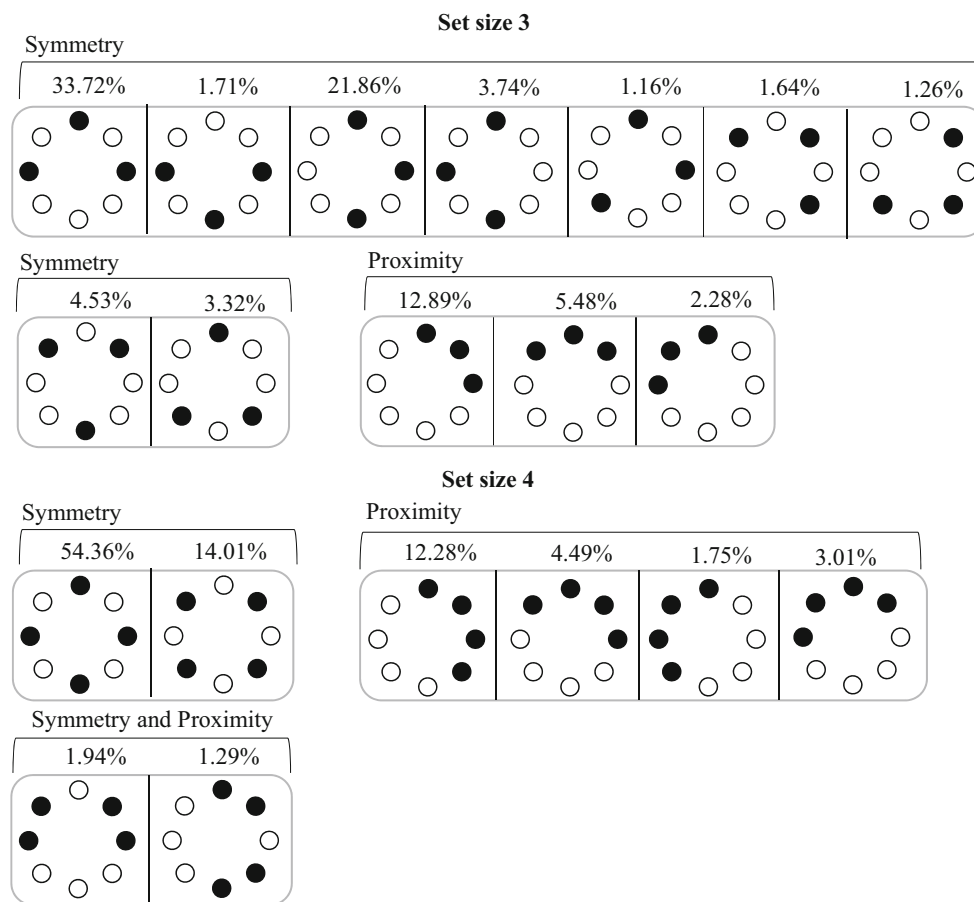
**Configuration construction** Inspection of the configurations that participants constructed revealed that across conditions,



**Fig. 2.** Percentages of visual targets that were selected at each target position, as a function of serial position (the order in which the visual targets were selected), presented separately for each set size. Error bars represent standard errors of the means

96% of the configurations were structured. Most of the participants based the configurations they constructed on symmetry, and a smaller number of participants based their

configurations on proximity (see Fig. 3 for depictions of the most common configurations). Unstructured configurations were defined if not all of the targets were placed in adjacent



**Fig. 3.** Configurations that participants constructed for each set size, based on symmetry and proximity. The number at the top of each configuration indicates the percentage of trials on which the configuration was selected (across all conditions). The figure includes all configurations that were selected more than 1% of the trials

(accounting for 93.59% of the configurations for set size 3, and 93.13% of the configurations for set size 4). Note that a small percentage of the configurations in set size 4 were based on a combination of symmetry and proximity (these configurations were defined as symmetry for the analysis)

locations (i.e., grouping by proximity) and if the configurations did not form a symmetrical shape. When two identical targets were present in the target array, participants selected both targets on less than 50% of the trials (see Table 1). We conducted a repeated measures analysis of variance (ANOVA) on the percentages of configurations that were based on similarity, with set size as a within-subjects factor and stimulus type and instructions (whether participants were informed of the presence of the identical targets) as between-subjects factors. The ANOVA revealed only a main effect of set size,  $F(1, 76) = 36.78, p < .001, \eta_p^2 = .326$ , reflecting the finding that participants selected more identical targets for set size 4. For all other effects,  $ps > .31, \eta_p^2s < .013$ . Thus, informing participants of the presence of the identical targets had no significant impact on whether these targets were selected.

For each of the participants, we calculated the percentages of trials in which their configurations were based on either symmetry or proximity cues. Participants were identified as having a dominant strategy if more than 75% of their choices were based on one cue. The somewhat lenient 75% threshold was chosen in order to account for initial learning and adjustments that may have occurred before participants committed to one specific strategy. The results were collapsed across the no-repetition and repeated-unselected conditions, which showed similar results. Table 2 shows the number of participants who had a dominant strategy, the number of participants who did not have a dominant strategy, and the number of participants who switched strategies throughout the experiment. The results are presented as a function of set size, identity (whether two identical targets were selected), and stimulus type (abstract shapes and real-world objects). Because not all participants selected the identical targets, the number of participants was lower in the repeated-selected condition.

**Table 1.** Percentages of trials on which the configurations were based on similarity cues (the repeated-selected trials), as a function of set size, stimulus type, and instructions (whether participants were informed that two identical targets would be present in the display)

Set Size	Instructions	Stimulus	Identical Items Selected (%) (SE)
3	Yes	Abstract shapes	38.90 (5.97)
		Real objects	43.47 (6.45)
	No	Abstract shapes	39.71 (5.67)
4	Yes	Abstract shapes	51.77 (7.05)
		Real objects	56.67 (6.65)
	No	Abstract shapes	50.31 (6.72)
		Real objects	47.71 (7.43)

SE = Standard error of the mean

As is demonstrated in Table 2, most participants had a dominant strategy. Moreover, in all conditions the averaged percentage of the trials that were based on the dominant strategy in these participants was over 94%. Thus, most participants consistently used the same organization cue throughout the task. Of the participants who did not have a dominant strategy, we identified a group of eight participants who had switched their strategy from the first to the third block of each set size (i.e., more than 75% of their choices were based on different strategies in the first and third blocks; see Table 2). Switching strategies was more frequent in the group of participants who encoded abstract shapes.

Chi-square tests were conducted in order to examine whether the type of stimuli that was encoded had an impact on the numbers of participants who adopted the different strategies. The analysis was conducted separately for each combination of set size and repetition conditions. We used the Bonferroni correction for multiple comparisons to correct for the number of comparisons (the threshold was set at .0125 [= .05/4 comparisons]). None of the comparisons were significant, even at an uncorrected .05 significance level: in set size 3,  $\chi^2(3) = 1.24$  and  $1.49, ps = .75$  and  $.68$ , for the nonidentical and identical conditions, respectively; in set size 4,  $\chi^2(3) = 6.79$  and  $5.56, ps = .08$  and  $.14$ , for the nonrepeated and repeated conditions, respectively. To verify our impressions that symmetry was more dominant than proximity as an organization cue, we ran a series of binomial tests for each combination of set size and repetition conditions (averaged across stimulus types), with a corrected threshold of .0125. All four tests showed that the proportion of participants with a dominant symmetry strategy exceeded the expected proportion of .5, all  $ps < .00001$  (two-tailed).

**Table 2.** Numbers of participants with and without dominant strategies of symmetry or proximity, as a function of set size, identity (whether identical targets were selected), and stimulus type

Set size	Stimulus	Symmetry n (%)	Proximity n (%)	Both n (%)	Switched n (%)
3	Abstract shapes	27 (67.5)	7 (17.5)	3 (7.5)	3 (7.5)
	Real objects	30 (75)	6 (15)	3 (7.5)	1 (2.5)
3 Identical	Abstract shapes	21 (58.3)	7 (19.4)	5 (13.9)	3 (8.3)
	Real objects	22 (71)	5 (16.1)	3 (9.7)	1 (3.2)
4	Abstract shapes	23 (57.5)	9 (22.5)	3 (7.5)	5 (12.5)
	Real objects	31 (77.5)	6 (15)	3 (7.5)	0 (0)
4 Identical	Abstract shapes	20 (54.1)	7 (18.9)	6 (16.2)	4 (10.8)
	Real objects	25 (75.8)	4 (12.1)	4 (12.1)	0 (0)

Note that set size and identity were manipulated within participants, therefore the same participants appear in these different conditions

Inspection of the configurations in which participants selected two identical targets revealed that the two identical targets were often placed in close proximity. We defined two placement conditions for the identical targets: near and far. In the symmetrical configurations, the near condition was defined when the two identical targets were placed at the smallest distance possible between two locations in the configuration. In configurations based on proximity, the near condition was defined when the two identical targets were placed in adjacent locations. The far condition was defined for all other placements. The results were analyzed separately for each configuration (symmetry or proximity) and set size (averaged across the two types of stimuli). When participants constructed configurations based on symmetry, the near condition accounted for 53% and 63% of the trials in set sizes 3 and 4, respectively. When configurations were based on proximity, the near condition accounted for 77% and 83% of the trials in set sizes 3 and 4, respectively.

Four one-sample *t* tests were conducted for each combination of configuration type and set size, against the value of 50%, to explore whether the near placements were more frequent (a corrected threshold of .0125 was used). The symmetrical configurations showed a nonsignificant effect for set size 3,  $t(54) = 0.49$ ,  $p = .63$ , Cohen's  $d = 0.07$ . The effect for set size 4 was significant at the uncorrected .05 threshold, but was not significant given the corrected threshold,  $t(54) = 2.52$ ,  $p = .015$ , Cohen's  $d = 0.34$  (all comparisons were two-tailed). Both *t* tests were significant when the configurations were based on proximity,  $t(32) = 4.60$ ,  $p < .001$ , Cohen's  $d = 0.80$ , and  $t(28) = 5.39$ ,  $p < .001$ , Cohen's  $d = 1.00$ , for set sizes 3 and 4, respectively. Note that participants had to match the sequence of selected targets to the sequence of selected locations in order to place the two identical targets together (see the [Method](#) section), which required more elaborate planning, suggesting that participants deemed this aspect of the configuration important.

**RT during encoding** Additional analyses focused on the time it took participants to select and place the memory targets in the different conditions. The RT for the first selected target was measured from the onset of the display until participants had clicked the first target. The RT for each subsequent target was measured with respect to the selection of the previous target. The RT for the first location was measured with respect to the selection of the last target, and the RTs for subsequent locations were measured relative to the selection of the previous location.

The first analysis focused on the selection RTs for the first target and the first location in the sequences, which we took as a measure of planning (i.e., the time it took participants to scan the display and decide which targets and locations to select). The second analysis explored the RT serial position effect in

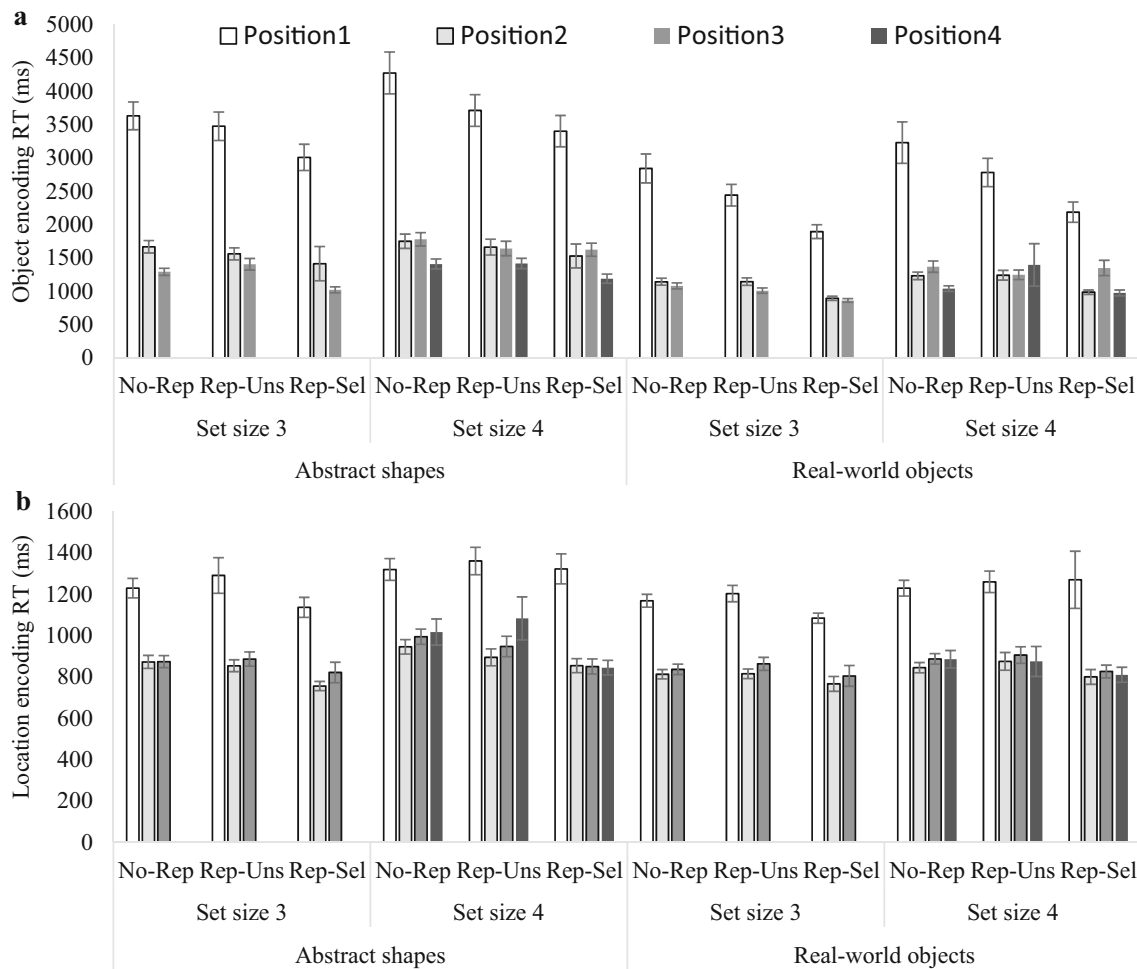
each sequence. One of the factors in the analyses was repetition (including the repetition-selected condition), and therefore the analyses included only the participants who had selected identical items in each of the conditions (the number of participants who were included in each analysis appears below). The RTs for all the conditions are shown in Fig. 4.

**RT for the first target** An ANOVA was conducted with set size and repetition as within-subjects factors and stimulus as a between-subjects factor ( $n = 65$ ). The ANOVA showed significant main effects of set size,  $F(1, 63) = 19.21$ ,  $p < .001$ ,  $\eta_p^2 = .234$ , and of stimulus,  $F(1, 63) = 12.11$ ,  $p < .001$ ,  $\eta_p^2 = .161$ . The interaction between these factors was not significant,  $F(1, 63) < 1$ . It took participants longer to select the first target when they were selecting abstract shapes and as set size increased. These effects suggest that participants planned the sequence of targets they selected before clicking on the first target. Naturally, this would take longer when the targets were perceptually more complex and when an additional target was selected.

The main effect of repetition was also significant,  $F(2, 126) = 31.12$ ,  $p < .001$ ,  $\eta_p^2 = .331$ . Repetition did not interact significantly with set size,  $F(2, 126) = 1.88$ ,  $p = .16$ ,  $\eta_p^2 = .029$ ; with stimulus,  $F(2, 126) < 1$ , or with set size and stimulus in a three-way interaction,  $F(2, 126) < 1$ . Additional planned contrasts showed that the RT was slowest in the no-repetition condition, when all the targets in the display were unique, and that it was significantly longer than the RT in the repetition-unselected condition, in which identical targets were presented but participants selected only unique targets,  $F(1, 63) = 16.84$ ,  $p < .001$ ,  $\eta_p^2 = .211$ . The RT in the latter condition was intermediate and was significantly longer than the RT in the repetition-selected condition, in which two identical targets were presented and participants selected both,  $F(1, 63) = 21.97$ ,  $p < .001$ ,  $\eta_p^2 = .259$ . This pattern suggests that participants did perceive the identical targets in the repetition-unselected condition, because they required less time to plan (i.e., to scan the target display) and initiate the target sequence. Yet, they did not select both targets in these trials.

**RT for the first location** A similar ANOVA ( $n = 65$ ) on location selection RTs revealed a nonsignificant main effect of set size,  $F(1, 63) = 3.64$ ,  $p = .06$ ,  $\eta_p^2 = .055$ , demonstrating that the RT was only slightly longer for set size 4. The main effect of stimulus was nonsignificant, as well,  $F(1, 63) = 3.13$ ,  $p = .08$ ,  $\eta_p^2 = .047$ , as was the interaction between these factors,  $F(1, 63) < 1$ . The main effect of repetition was significant,  $F(2, 126) = 8.10$ ,  $p < .001$ ,  $\eta_p^2 = .114$ , but its interactions with set size, with stimulus, and with set size and stimulus were all nonsignificant, all  $F_s(2, 126) < 1$ .





**Fig. 4.** Encoding reaction times (RTs, in milliseconds) in the different conditions as a function of serial position, for (a) memory targets and (b) locations. Note that the y-axes are scaled between the two figures to accommodate the RT data. No-Rep = no-repetition trials, in which only distinct targets were presented. Rep-Uns = repetition-unselected trials, in

which two identical items were presented but only unique items were selected. Rep-Sel = repetition-selected trials, in which identical items were both presented and selected. Error bars represent standard errors of the means

The first of two planned contrasts revealed similar RTs in the two conditions in which distinct items were selected (with a tendency for RTs to be longer in the repetition-unselected condition),  $F(1, 63) = 3.31, p = .07, \eta_p^2 = .050$ . The second contrast revealed significantly faster RTs when two identical items were presented and selected, relative to when identical items were presented but not selected,  $F(1, 63) = 9.91, p = .003, \eta_p^2 = .136$ . Thus, participants were faster to select targets and locations when the configurations included identical items.

**Serial position effects on RTs for set size 3 (n = 67)**

**Target selection** An ANOVA with serial position and repetition as within-subjects factors and stimulus as a between-subjects factor revealed significant main effects of serial position,  $F(2, 130) = 189.96, p < .001, \eta_p^2 = .745$ , and stimulus,  $F(1, 65) = 21.78, p < .001, \eta_p^2 = .251$ , as well as a significant

interaction between them,  $F(2, 130) = 6.90, p < .001, \eta_p^2 = .096$ . The RT decreased from the first to the subsequent targets, which further supports the assumption that participants planned the sequence of targets they would select before selecting the first target. This decrease was larger when abstract shapes were encoded, mostly due to the increased RT for the first item in the sequence. The main effect of repetition was also significant,  $F(2, 130) = 16.10, p < .001, \eta_p^2 = .199$ , as was its interaction with serial position,  $F(4, 260) = 8.46, p < .001, \eta_p^2 = .115$ . The interactions of repetition with stimulus and with stimulus and set size were nonsignificant,  $F(2, 130) < 1$  and  $F(4, 260) = 1.60, p = .18, \eta_p^2 = .024$ , respectively. Planned contrasts showed that the RT was slower in the no-repetition condition than in the repetition-unselected condition,  $F(1, 65) = 9.19, p = .003, \eta_p^2 = .124$ , and that the latter condition was significantly slower than the repetition-selected condition,  $F(1, 65) = 10.09, p = .002, \eta_p^2 = .134$ .

**Location selection** The analysis of RTs for the sequence of locations revealed a significant main effect of serial position,  $F(2, 130) = 175.08, p < .001, \eta_p^2 = .729$ , due to the increased RTs for the first location in the sequence, again supporting the assumption that participants planned the sequence of locations before placing the first target. The main effect of stimulus and its interaction with serial position were nonsignificant,  $F(1, 65) < 1$  and  $F(2, 130) < 1$ , respectively. The main effect of repetition was significant,  $F(2, 130) = 10.77, p < .001, \eta_p^2 = .142$ . The repetition effect did not interact significantly with serial position,  $F(4, 260) = 2.08, p = .08, \eta_p^2 = .031$ ; with stimulus,  $F(2, 130) < 1$ ; or with serial position and stimulus,  $F(4, 260) < 1$ . Planned contrasts showed similar RTs in the no-repetition and repetition-unselected conditions,  $F(1, 65) = 2.42, p = .12, \eta_p^2 = .036$ , and significantly faster RTs in the repeated-selected than in the repeated-unselected condition,  $F(1, 65) = 11.95, p < .001, \eta_p^2 = .155$ .

### Serial position effects on RTs for set size 4

The serial position effects for set size 4 ( $n = 69$ ) were similar to the effects for set size 3.

**Target selection** The main effect of serial position was significant,  $F(3, 201) = 138.21, p < .001, \eta_p^2 = .674$ , as was the main effect of stimulus,  $F(1, 67) = 12.82, p < .001, \eta_p^2 = .161$ , and the interaction between these factors,  $F(3, 201) = 138.21, p = .003, \eta_p^2 = .067$ . The RT decreased from the first to subsequent targets, and the decrease was steeper when abstract shapes were encoded. The repetition effect was also significant,  $F(2, 134) = 14.57, p < .001, \eta_p^2 = .179$ , as was its interaction with serial position,  $F(6, 402) = 7.64, p < .001, \eta_p^2 = .102$ . The repetition effect did not interact significantly with stimulus,  $F(2, 134) < 1$ , or with serial position and stimulus,  $F(6, 402) < 1$ .

Planned contrasts showed significant differences between the no-repetition and repetition-unselected conditions, due to faster RTs in the latter condition,  $F(1, 67) = 5.52, p = .02, \eta_p^2 = .076$ , as well as a significant interaction of this contrast with serial position,  $F(1, 67) = 6.62, p = .01, \eta_p^2 = .090$ . The repetition-selected condition was significantly faster than the repetition-unselected condition,  $F(1, 67) = 7.77, p = .007, \eta_p^2 = .104$ . This contrast interacted with serial position, as well,  $F(1, 67) = 5.86, p = .02, \eta_p^2 = .080$ .

**Location selection** An ANOVA revealed a significant main effect of serial position,  $F(3, 201) = 74.68, p < .001, \eta_p^2 = .527$ , reflecting an increased RT for the first location. The main effect of stimulus was nonsignificant,  $F(1, 67) = 1.88, p = .18, \eta_p^2 = .027$ , as was the interaction between serial position and stimulus,  $F(3, 201) < 1$ . The main effect of repetition was significant,  $F(2, 134) = 5.63, p = .005, \eta_p^2 = .077$ , but it did not interact significantly with serial position,  $F(6,$

$402) = 1.71, p = .12, \eta_p^2 = .025$ ; with stimulus,  $F(2, 134) < 1$ ; or with serial position and stimulus,  $F(6, 402) < 1$ . Additional contrasts showed a nonsignificant difference between the no-repetition and repetition-unselected conditions,  $F(1, 67) < 1$ , as well as significantly faster RTs in the repetition-selected than in the repetition-unselected condition,  $F(1, 67) = 6.16, p < .02, \eta_p^2 = .084$ .

### Retrieval

Accuracy was high across all the conditions, showing only a significant main effect of set size. The results are presented in the [Appendix](#).

### Strategy questionnaire

Half of the participants filled out a questionnaire at the end of the experiment regarding the strategies they had used to select the targets and locations during encoding. The strategies are presented in [Table 3](#). Participants who encoded abstract shapes reported using different strategies for selecting the visual targets than did the group of participants who encoded real-world object. In contrast, the two groups of participants reported using similar strategies for selecting the target locations. Regardless of the specific strategy, all participants reported that the strategies they used had improved their memory performance.

**Target selection** The participants who memorized abstract shapes reported that they had selected shapes that resembled familiar objects. The benefit of familiarity was explained as an opportunity to attach verbal labels to the abstract shapes. Several participants explicitly commented that verbal labels were easier to rehearse. Only two participants selected abstract shapes on the basis of visual features. Specifically, they selected shapes that were compatible with the locations at which the targets were placed (e.g., a shape with a unique feature at the top was placed at the top of the circular array). Of the participants who had memorized real-world objects that were readily verbalized, most reported that they had selected targets on the basis of distinct visual features, mostly color. Five participants mentioned that they had selected objects with distinct colors, whereas one participant selected objects with similar colors. Four of these participants reported that in addition to color, they selected objects that were easily verbalized. Three participants reported selecting objects that were semantically related.

All the participants noticed that two identical targets appeared in the target display. Thirty-seven of the participants mentioned that they had selected the identical targets most of time or whenever they noticed them. Only three participants noticed the identical targets but did not select them. Except for

**Table 3.** Numbers of participants who reported using each of the identified encoding strategies, as a function of stimulus type ( $n = 40$ )

Target Selection Strategies	Verbal Labels	Visual Features	Semantic Association
	<i>n</i>	<i>n</i>	<i>n</i>
Abstract shapes	18	2	0
Real-world objects	4	17	3
Location Selection Strategies	Spatial Separation	Fixed Locations	Familiar Shapes
	<i>n</i>	<i>n</i>	<i>n</i>
Abstract shapes	15	7	1
Real-world objects	13	16	1

Note that several participants reported using more than one strategy

these three participants, all participants replied that memorizing memory displays in which two targets were identical benefited memory performance.

**Location selection** Most of the participants reported that they had selected locations that were far apart in order to aid memory performance. Participants also reported that they had selected fixed locations throughout the experiment (11 participants reported that their strategy was to select fixed locations, without referring to the specific locations). Two participants expressed the locations that they had selected in terms of familiar shapes (e.g., triangle or a square). Thus, even participants who reported that they had selected and rehearsed targets on the basis of verbal labels also reported constructing structured spatial configurations.

Finally, when asked whether they had employed memory processes similar to those they used in everyday life, 27 of the 40 participants who filled out the questionnaire responded positively. These reports are compatible with the findings that participants in the task quickly implemented encoding strategies that were mostly consistent within and across participants, strategies that, as most of the participants reported, they implemented in their everyday lives to enhance memory performance.

## Discussion

In the present study we approached the question of the interaction between structure and VWM from a novel perspective, by exploring how individuals structure their own memory displays. Although it is prevalent in everyday behavior, this aspect of VWM, which we termed *self-initiated VWM*, is absent from the research literature. To explore the processes of self-initiated WM for visual displays, we used a modified change detection task in which participants selected the targets they were to memorize and then placed them in several locations in a circular array of eight locations. We predicted that participants would construct memory displays that would maximize accuracy. The results of the study demonstrated that the participants constructed structured representations based, most frequently, on the

Gestalt cue of symmetry and to a lesser extent on the Gestalt cues of proximity and similarity. Furthermore, when participants selected two identical items, they tended to place them in close proximity, demonstrating that participants constructed complex structures based on the interaction of different Gestalt organization cues. The same results were obtained with real-world objects and abstract shapes and were similar across the two set sizes that were tested.

Half of the participants filled out questionnaires at the end of the experiment. In their answers to the questionnaire, participants reported several strategies that they had used to select the visual targets. The participants who encoded abstract shapes selected shapes that resembled familiar objects that they could verbalize. The assumption that verbal labels are formed in VWM is widely accepted, since the participants in numerous VWM studies are asked to perform secondary tasks in order to block verbal labeling of the visual targets (e.g., articulatory suppression). Nevertheless, the specific role played by verbal labels in VWM, or whether such labels genuinely benefit memory performance, is still debated (Donkin, Nosofsky, Gold, & Shiffrin, 2015; Sense, Morey, Prince, Heathcote, & Morey, 2017; Souza & Skora, 2017). The participants in the present study reported that the verbal labels they had used benefited memory performance.

Contrary to the strategies used when abstract shapes were selected, participants who encoded real-world objects, which were readily verbalized, attempted to enhance memory performance by selecting objects on the basis of their distinct visual features, mostly color. Thus, the strategies that the different groups of participants employed when abstract shapes and real-world objects were encoded point to a complex interaction between visual and verbal codes in VWM, an interaction that may depend on the complexity of each code. Furthermore, although participants reported that they had selected the visual targets on the basis of nonspatial features, selections were spatially biased to targets in the left and central parts of the horizontal target display.

In terms of the structure of the memory configurations, in order to maximize accuracy, participants took isolated elements (i.e., the memory targets) and created Gestalt wholes to be held in memory. Creating the structured displays was time consuming, and the encoding RTs suggested that

participants planned before they executed the sequences of targets and locations they selected. The targets and locations were selected by participants at their own pace, and therefore it is most likely that participants had already encoded the memory targets when they placed them in the locations they chose themselves. Thus, we assume that the structured displays that participants constructed after careful planning were intended mostly to benefit maintenance processes rather than encoding. Taken together, the results of the present study suggest that participants have access to metacognitive knowledge about the benefit of structure in VWM encoding and maintenance. Moreover, the finding that naïve participants constructed structured configurations based on Gestalt organization cues suggests that self-initiated VWM can provide a glimpse into the structure of what constitute natural and efficient memory representations in VWM, and can inform models of memory for externally provided information, as well.

Although structure can take many forms, participants were rather consistent in their preference for Gestalt organization cues, and specifically for symmetry, perceiving it as the most efficient cue for enhancing memory performance. Relative to other organization cues, such as proximity and similarity, symmetry has been studied much less in connection with VWM. The studies that have explored the role of symmetry in VWM have revealed that, similar to the other cues, symmetry was efficient in enhancing memory performance for visual patterns or spatial configurations (Kemps, 2001; Rossi-Arnaud et al., 2006; Rossi-Arnaud et al., 2012). However, in these studies the symmetrical patterns were probed during test as wholes, contrary to the design of the present study, in which participants were probed on single probes.

Symmetry is considered a highly salient cue in perception, and symmetry detection has been shown to be fast and efficient, especially when symmetry along the vertical axis is detected (Wagemans, 1997; Wagemans et al., 2012). Nevertheless, there are indications that symmetry is not an efficient perceptual grouping cue, and its role in perceptual organization is unclear (e.g., Machilsen, Pauwels, & Wagemans, 2009; Pomerantz & Portillo, 2011; Wagemans et al., 2012). Machilsen et al., for instance, found a small but significant advantage of symmetry in the perceptual grouping of contour elements. To explain the apparent contrast with the high saliency (and efficient detection) of symmetry in shapes and objects, they concluded that symmetry may be a salient characteristic of percepts that are already established, but not a strong cue to establish a group or figure–ground organization in the first place. Proximity, on the other hand, is considered a highly efficient organization cue in perception, and to be more efficient than symmetry when the two cues are compared directly (e.g., Elder & Goldberg, 2002; Kubovy & Van den Berg, 2008; Pomerantz & Portillo, 2011). Moreover, recently Kimchi, Yeshurun, Spehar, and Pirkner (2016) demonstrated that grouping by symmetry failed to capture stimulus-driven

visual attention automatically, as opposed to collinearity and closure, the other grouping cues tested.

Despite the less effective role of symmetry in perceptual grouping, the participants in the present study preferred symmetry over proximity when they constructed their own memory displays. Perhaps its high saliency (Wagemans, 1997; Wagemans et al., 2012) prompted participants to use symmetry to construct their memory displays. Alternatively, it is also possible that the demands of the VWM task led to the selection of symmetry over proximity. It is well-established that targets in VWM are maintained along with their spatial locations (Gratton, 1998; Jiang, Olson, & Chun, 2000; Treisman & Zhang, 2006). Furthermore, to respond correctly in the present task, participants were required to maintain object–location conjunctions. Symmetry, contrary to proximity, allowed participants the benefit of structure while at the same time keeping the memorized targets spatially separated, in order to reduce the overlap between them. Indeed, in their responses to the strategy questionnaire, most participants emphasized the spatial separation between the different elements in the memory display as an organization cue that benefited memory performance. Thus, grouping the memory targets by symmetry may be more beneficial than grouping by proximity in visual change detection tasks.

Given the participants emphasized the spatial separation between items in the constructed configurations, an alternative explanation of the results would suggest that participants adopted an un-proximity strategy (i.e., attempting to maximize the distance between items) rather than a symmetry strategy. Given the design of the circular array, the attempt to maximize the distance between items led to the incidental construction of symmetrical configurations. This explanation, however, is less likely given that only several symmetrical configurations dominated the participants selections. A strict unproximity strategy should have yielded a more uniform distribution of a wider range of configurations. Moreover, for set size 3, the distance between items was not maximized in the majority of the symmetrical configurations (which accounted for 64% of all the set size 3 configurations). Thus, these data support our conclusions that participants actively selected structured symmetrical configurations.

In addition to symmetry and proximity, another Gestalt cue that was used by participants to construct their memory displays was similarity. Participants did not use similarity as often as they could, although in their responses to the strategy questionnaire they reported that they selected the identical targets frequently, and acknowledged the similarity beneficial effect on memory. These reports are somewhat at odds with the findings that target selection RT was faster in the repetition-unselected condition relative to the no-repetition condition, which suggested that participants noticed the identical targets but selected unique items. On the other hand, the identical targets were selected more frequently in set size 4 when the task load



increased, suggesting that participants were indeed aware of the additional benefit of similarity. The finding that participants did not use similarity more frequently is also puzzling given that encoding was overall faster when two identical targets were selected. One potential resolution of these conflicting results may be found in studies that examined the interaction between two or more Gestalt organization cues.

Previous studies in perception and VWM have shown that the effects of combining two or more Gestalt grouping cues on behavior may be additive or interactive. For example, the effects of proximity and similarity were additive in perception, such that the total effect of grouping by proximity and similarity was equal to the sum of the independent effects of each grouping cue (Kubovy & Van den Berg, 2008). Thus, one plausible explanation for the less frequent use of similarity as a grouping cue in the present study suggests that participants did not consider the additional effect of similarity necessary, given the overall task difficulty, although they reported that they used it often. Furthermore, Gestalt cues have also been shown to interfere with each other. For instance, Peterson et al. (2015) found enhanced VWM processing when items in the memory display were grouped on the basis of similarity. However, when similarity and connectedness were both present in the display, the two cues interfered with each other. Perhaps in the context of the present study, similarity interfered to some degree with the cues of symmetry and proximity, and therefore the benefit of similarity was not strong enough to compensate for this interference. Participants' reports that they used similarity often may suggest that they did not have access to the metacognitive knowledge on the beneficial or interfering effects of several Gestalt cues. These suggestions are speculative at present, and will be a subject for future investigations.

When identical items were selected, they were often placed close to each other. Peterson and Berryhill (2013) demonstrated that similarity benefited VWM only when the similar items appeared in close proximity, a finding that is compatible with the memory displays that participants constructed in the present study. This pattern, however, was not replicated in two other studies (Morey et al., 2015; Peterson et al., 2015). Thus, the interaction between similarity and proximity in VWM (for both self-initiated and provided information) should be clarified further in future studies.

Accuracy in the task was high, almost at ceiling in all conditions (see the [Appendix](#)). The only significant effect on accuracy was that of set size. Considering that encoding was self-paced suggests that participants efficiently adjusted encoding RT to achieve almost perfect performance, whether participants maintained real-world objects or the more demanding abstract shapes. The increase in encoding RT in the abstract shapes task eliminated the expected difference in accuracy between the two conditions. Interestingly, although RT was also adjusted when set size increased (i.e., RT was longer

in set size 4 than set size 3), the accuracy difference between the two set sizes was not eliminated. One likely interpretation of this pattern of results is that target complexity provides a challenge for encoding processes that can be eliminated when encoding RT is adjusted, whereas set size provides a challenge for encoding but also for maintenance and possibly retrieval processes.

The self-initiated aspect of WM has been mostly absent from the WM literature. In the present study we examined the strategies and representations of VWM, whereas in a recent study we explored the representations formed in spatial WM (Magen & Emmanouil, 2018). In both studies participants constructed structured representations that they tended to initiate on the left side. However, although the participants in the present study, who memorized visual information, based their configurations on symmetry, in spatial self-initiated WM participants preferred proximity, constructing spatial sequences with short distances between locations in the sequence. In both types of self-initiated WM, participants constructed simple memory displays that nevertheless, entailed careful planning. It would be interesting in future studies to use eye-tracking to follow the encoding processes, and especially the planning phase before the participants select the first element in the sequence. Finally, in the spatial self-initiated WM task, we compared participants accuracy in the self-initiated task, with accuracy in a non-self-initiated task in which participants memorized spatial sequences provided to them. The results demonstrated that self-initiation benefited memory performance in comparison with provided structured and unstructured spatial configurations. Due to the ceiling effects on accuracy in the present task, this aspect of performance was not explored here and will be a subject for future investigations.

## Conclusions

As part of their everyday ongoing behavior, individuals often memorize information on their own initiative. Although it is prevalent in everyday contexts, this aspect of memory has been largely absent from the literature on VWM. The present study illuminates the underlying processes of self-initiated VWM when participants construct the visual displays they memorize. Most notably, participants planned and constructed configurations based on Gestalt organization cues of symmetry, proximity, and similarity, revealing that they have access to metacognitive knowledge of the benefit of Gestalt organization cues in VWM. Furthermore, participants employed several strategies while selecting the visual targets themselves, including attaching verbal labels to abstract shapes and selecting real-world objects on the basis of salient visual features such as color. Seventy percent of the participants reported that they used similar memory strategies and processes in their everyday life, demonstrating the importance of studying self-initiated VWM. More generally, self-initiated WM captures the complexity of everyday memory



functions and explores how individuals interact with the world to enhance memory performance. To function efficiently, individuals must consider the state of their current and future surroundings and the strengths and weaknesses of their cognitive system. This study opens new questions and avenues for research on VWM, emphasizing self-initiation, metacognition, and planning when maintaining visual information for short periods of time.

**Author note** We thank Johan Wagemans, Lore Goetschalckx, and two anonymous reviewers for their helpful comments on an earlier version of the manuscript.

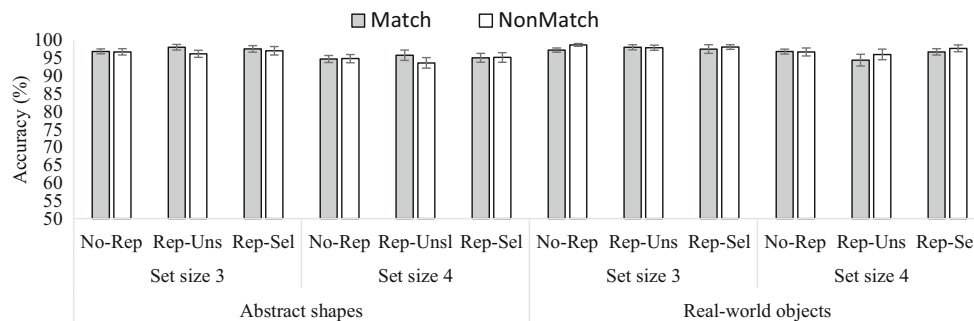
## Appendix: Accuracy measures

For each set size, accuracy was collapsed over probes matched to items in the different serial positions, to obtain an accuracy measure of set size (see Fig. 5). An ANOVA explored the impact of the different conditions on accuracy, with set size, probe (match or nonmatch), and repetition as within-subjects factors, and stimulus as a between-subjects factor (see Table 4 for statistics,  $n = 62$ ). The only significant effect was a main effect of set size, and a marginally significant interaction between probe and stimulus type, reflecting the observation that accuracy was slightly higher in the match condition when participants memorized abstract shapes, but was slightly higher in the nonmatch condition when participants memorized real world objects. Thus, overall accuracy was almost at ceiling.

**Table 4.** Univariate inferential statistics for the analysis of accuracy

Variable	$F$	$\eta_p^2$
Set Size	21.94 <sup>a</sup>	.268 <sup>***</sup>
Probe	<1 <sup>a</sup>	.004
Repetition	2.06 <sup>b</sup>	.033
Stimulus	1.30 <sup>a</sup>	.021
Set Size × Probe	<1 <sup>a</sup>	.000
Set Size × Repetition	1.87 <sup>b</sup>	.030
Set Size × Stimulus	<1 <sup>a</sup>	.013
Probe × Repetition	1.50 <sup>b</sup>	.024
Probe × Stimulus	4.06 <sup>a</sup>	.063 <sup>§</sup>
Repetition × Stimulus	<1 <sup>b</sup>	.013
Set Size × Probe × Stimulus	<1 <sup>a</sup>	.002
Set Size × Repetition × Stimulus	<1 <sup>b</sup>	.013
Probe × Repetition × Stimulus	<1 <sup>b</sup>	.015
Set Size × Probe × Repetition	<1 <sup>b</sup>	.015
Set Size × Probe × Repetition × Stimulus	1.09 <sup>b</sup>	.018

<sup>a</sup>  $df = 1, 60$ . <sup>b</sup>  $df = 2, 120$ . <sup>§</sup>  $p = .05$ . <sup>\*\*\*</sup>  $p < .001$



**Fig. 5.** Accuracy (%) in the different conditions. No-Rep = no-repetition trials in which only distinct targets were presented. Rep-Uns = repetition-unselected, trials in which two identical items were presented but were

not selected. Rep-Sel = repetition-selected, trials in which identical items were presented and selected. Error bars represent standard errors of the means

## References

- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York, NY: Academic Press.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 4:1–34. <https://doi.org/10.1167/11.5.4>
- Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review*, *22*, 170–178.
- Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, *2*, 324–353.
- Gao, Z., Gao, Q., Tang, N., Shui, R., & Shen, M. (2015). Organization principles in visual working memory: Evidence from sequential stimulus display. *Cognition*, *146*, 277–288.
- Gao, Z., Xu, X., Chen, Z., Yin, J., Shen, M., & Shui, R. (2011). Contralateral delay activity tracks object identity information in visual short-term memory. *Brain Research*, *1406*, 30–42.
- Gratton, G. (1998). The contralateral organization of visual memory: A theoretical concept and a research tool. *Psychophysiology*, *35*, 638–647. <https://doi.org/10.1017/S0048577298980453>
- Jiang, Y. V., Lee, H. J., Asaad, A., & Remington, R. (2016). Similarity effects in visual working memory. *Psychonomic Bulletin & Review*, *23*, 476–482.
- Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 683–702. <https://doi.org/10.1037/0278-7393.26.3.683>
- Kemps, E. (2001). Complexity effects in visuo-spatial working memory: Implications for the role of long-term memory. *Memory*, *9*, 13–27.
- Kimchi, R. (2009). Perceptual organization and visual attention. *Progress in Brain Research*, *176*, 15–33.
- Kimchi, R., Yeshurun, Y., & Cohen-Savransky, A. (2007). Automatic, stimulus-driven attentional capture by objecthood. *Psychonomic Bulletin & Review*, *14*, 166–172. <https://doi.org/10.3758/BF03194045>
- Kimchi, R., Yeshurun, Y., Spehar, B., & Pirkner, Y. (2016). Perceptual organization, visual attention, and objecthood. *Vision Research*, *126*, 34–51. <https://doi.org/10.1016/j.visres.2015.07.008>
- Kubovy, M., & Van den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review*, *115*, 131–154. <https://doi.org/10.1037/0033-295X.115.1.131>
- Lin, P.-H., & Luck, S. J. (2009). The influence of similarity on visual working memory representations. *Visual Cognition*, *17*, 356–372. <https://doi.org/10.1080/13506280701766313>
- Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*, 391–400. <https://doi.org/10.1016/j.tics.2013.06.006>
- Luria, R., & Vogel, E. K. (2014). Come together, right now: Dynamic overwriting of an object's history through common fate. *Journal of Cognitive Neuroscience*, *26*, 1819–1826.
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*, 347–356.
- Machilsen, B., Pauwels, M., & Wagemans, J. (2009). The role of vertical mirror symmetry in visual shape detection. *Journal of Vision*, *9*(12), 11:1–11. <https://doi.org/10.1167/9.12.11>
- Magen, H., & Emmanouil, T. A. (2018). Working memory for self-initiated and provided spatial configurations. *Quarterly Journal of Experimental Psychology*. <https://doi.org/10.1177/1747021817739808>
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.
- Morey, C. C., Cong, Y., Zheng, Y., Price, M., & Morey, R. D. (2015). The color-sharing bonus: Roles of perceptual organization and attentive processes in visual working memory. *Archives of Scientific Psychology*, *3*, 18–29.
- Orhan, A. E., & Jacobs, R. A. (2014). Toward ecologically realistic theories in visual short-term memory research. *Attention, Perception, & Psychophysics*, *76*, 2158–2170.
- Parmentier, F. B. R., Elford, G., & Maybery, M. (2005). Transitional information in spatial serial memory: Path characteristics affect recall performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 412–427. <https://doi.org/10.1037/0278-7393.31.3.412>
- Peterson, D. J., & Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychonomic Bulletin & Review*, *20*, 1282–1289.
- Peterson, D. J., Gözenman, F., Arciniega, H., & Berryhill, M. E. (2015). Contralateral delay activity tracks the influence of Gestalt grouping principles on active visual working memory representations. *Attention, Perception, & Psychophysics*, *77*, 2270–2283. <https://doi.org/10.3758/s13414-015-0929-y>
- Pomerantz, J. R., & Portillo, M. C. (2011). Grouping and emergent features in vision: Toward a theory of basic Gestalts. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1331–1349.
- Quinlan, P. T., & Cohen, D. J. (2012). Grouping and binding in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1432–1438.
- Rossi-Arnaud, C., Pieroni, L., & Baddeley, A. D. (2006). Symmetry and binding in visuo-spatial working memory. *Neuroscience*, *139*, 393–400.
- Rossi-Arnaud, C., Pieroni, L., Spataro, P., & Baddeley, A. (2012). Working memory and individual differences in the encoding of vertical, horizontal and diagonal symmetry. *Acta Psychologica*, *141*, 122–132.
- Sense, F., Morey, C. C., Prince, M., Heathcote, A., & Morey, R. D. (2017). Opportunity for verbalization does not improve visual change detection performance: A state-trace analysis. *Behavior Research Methods*, *49*, 853–862. <https://doi.org/10.3758/s13428-016-0741-1>
- Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, *166*, 277–297.
- Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition*, *34*, 1704–1719. <https://doi.org/10.3758/BF03195932>
- van Lamsweerde, A. E., Beck, M. R., & Johnson, J. S. (2016). Visual working memory organization is subject to top-down control. *Psychonomic Bulletin & Review*, *23*, 1181–1189. <https://doi.org/10.3758/s13423-015-0976-3>
- Wagemans, J. (1997). Characteristics and models of human symmetry detection. *Trends in Cognitive Sciences*, *1*, 346–352.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, *138*, 1172–1217. <https://doi.org/10.1037/a0029333>
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt: II. *Psychologische Forschung*, *4*, 301–350. <https://doi.org/10.1007/BF00410640> [Translated in 2012 as “Investigations on Gestalt principles.” In L. Spillmann (Ed.), *Max Wertheimer: On perceived motion and figural organization* (pp. 127–182). Cambridge, MA: MIT Press.]

- Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, *10*, 80–87. <https://doi.org/10.3758/BF03196470>
- Xu, Y. (2006). Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness. *Perception & Psychophysics*, *68*, 815–828. <https://doi.org/10.3758/BF03193704>
- Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences*, *104*, 18766–18771. <https://doi.org/10.1073/pnas.0705618104>