

# Representations in mental imagery and working memory: Evidence from different types of visual masks

Gregoire Borst · Giorgio Ganis ·  
William L. Thompson · Stephen M. Kosslyn

© Psychonomic Society, Inc. 2011

**Abstract** Although few studies have systematically investigated the relationship between visual mental imagery and visual working memory, work on the effects of passive visual interference has generally demonstrated a dissociation between the two functions. In four experiments, we investigated a possible commonality between the two functions: We asked whether both rely on depictive representations. Participants judged the visual properties of letters using visual mental images or pictures of unfamiliar letters stored in short-term memory. Participants performed both tasks with two different types of interference: sequences of unstructured visual masks (consisting of randomly changing white and black dots) or sequences of structured visual masks (consisting of fragments of letters). The structured visual noise contained elements of depictive representations (i.e., shape fragments arrayed in space), and hence should interfere with stored depictive representations; the unstructured visual noise did not contain such elements, and thus should not interfere as

much with such stored representations. Participants did in fact make more errors in both tasks with sequences of structured visual masks. Various controls converged in demonstrating that in both tasks participants used representations that depicted the shapes of the letters. These findings not only constrain theories of visual mental imagery and visual working memory, but also have direct implications for why some studies have failed to find that dynamic visual noise interferes with visual working memory.

**Keywords** Visual mental imagery · Visual working memory · Dynamic visual noise · Short-term memory

Visual mental imagery plays a role in a wide range of everyday activities—such as navigating to a store, remembering a grocery list, and packing groceries into the trunk of the car—and is important more generally in such cognitive functions as learning (e.g., Paivio, 1971), memory (e.g., Schacter, 1996), and reasoning (e.g., Kosslyn 1983). Visual mental imagery (MI) typically occurs “when a representation of the type created during the initial phase of perception is present but the stimulus is not actually being perceived; such representations preserve the perceptible properties of the stimulus and ultimately give rise to the subjective experience of perception” (Kosslyn, Thompson, & Ganis, 2006). Many of the functions of imagery, especially its role in reasoning, echo functions that have been attributed to working memory (WM; Baddeley, 1986). However, relatively little research has attempted to pinpoint the ways in which visuospatial imagery and visuospatial working memory are the same or different. In the present experiments, we investigated whether visual MI and visual WM rely on representations that share the same format.

---

G. Borst (✉)  
GINDEV - CNRS (UMR 6232),  
Université Paris Descartes, Sorbonne Paris Cité,  
46 rue Saint Jacques,  
75005 Paris, France  
e-mail: gregoire.borst@parisdescartes.fr

G. Ganis  
School of Psychology, University of Plymouth,  
Plymouth, UK

W. L. Thompson  
Department of Psychology, Harvard University,  
Cambridge, MA, USA

S. M. Kosslyn  
Center for Advanced Study in the Behavioral Sciences and  
Department of Psychology, Stanford University,  
Stanford, CA, USA

In the model originally proposed by Baddeley and his colleagues (Baddeley, 1986; Baddeley & Hitch, 1974), WM includes three distinct components: a phonological loop (which maintains auditory representations of verbal and auditory information), a visuospatial sketchpad (which maintains representations of visual and spatial information), and a central executive that uses representations stored in these two “slave systems” in complex cognitive tasks, such as reasoning and learning. Logie (1995, 2003; see also Logie & van der Meulen, 2009) further articulated the architecture of WM by suggesting that perceptual information accesses previously stored knowledge, and relatively abstract representations are then fed into a passive visual store (i.e., a “visual cache”) and rehearsed in a spatial active store (i.e., an “inner scribe”). According to this view, the visual cache serves as a visual short-term memory (VSTM) by holding the product of initial perceptual input. According to Pearson (2001), information maintained in this visual store is not itself a visual mental image, but rather can be used to create visual mental images within a visual buffer similar to the one described in Kosslyn’s model (1994; see also Kosslyn et al. 2006). In Logie’s (2003) view, MI and WM rely on partially distinct structures—the generation and the manipulation of visual mental images rely on executive processes, not on the visual cache. Recently, Quinn (2008) proposed an alternative distinction between the visual buffer and the visual cache: The buffer supports depictive representations and receives direct visual inputs, and irrelevant visual inputs would interfere with its content, whereas the cache is insensitive to direct perceptual interference and maintains previously interpreted materials.

In order to shed light on the nature of the visuospatial WM system, most studies have relied on observing the effect of a secondary task (i.e., an interference task) on the performance of a primary task (i.e., the WM task). Critically, passive interference tasks—such as the presentation of irrelevant visual information—have been used to infer the nature of the representations maintained in the visual cache. For example, in the dynamic visual noise (DVN) technique, participants are asked to watch an 80×80 grid of black and white dots that randomly change from black to white, or vice versa, to create a flickering effect. In a series of experiments, Quinn and McConnell (1996, 1999; McConnell & Quinn, 2000) and others (e.g., Andrade, Kemps, Werniers, May, & Szmalec, 2002) have demonstrated that DVN disrupts the memorization of words using the imagery-based peg-word mnemonic technique, whereas irrelevant speech does not disrupt such learning—and that the opposite is true for words memorized by rote rehearsal. In addition, Dean, Dewhurst, and Whittaker (2008) reported that DVN disrupts memo-

rization of colored matrices, which were not easily encoded verbally or spatially (and, hence, presumably were memorized by using visual MI).

However, Quinn and McConnell (2006) showed that although DVN interferes with encoding and recall of words learned with the peg-word mnemonic technique, it does not interfere during the maintenance phase. In addition, Andrade et al. (2002) and others (e.g., Avons & Sestieri, 2005; Zimmer & Speiser, 2002; Zimmer, Speiser, & Seidler, 2003) found no evidence that DVN interferes with short-term recognition or recall of Chinese characters (Andrade et al. 2002, Exp.5) or with VSTM of matrix patterns (Avons & Sestieri, 2005). These findings have led some to argue that DVN interferes selectively with visual MI but has no effect on VSTM—which in turn led these researchers to postulate a dissociation between the visual cache, which supports VSTM, and the visual buffer, which supports the generation of visual mental images (for discussion, see, e.g., Logie & van der Meulen, 2009; van der Meulen, Logie, & Della Sala, 2009). Hence, visual MI and WM could involve different sets of mental representations and processes.

In the present study, we examined whether MI and WM rely (at least in part) on representations that share the same format. In order to claim that two information-processing systems—such as MI and WM—share common cognitive processes, a prerequisite is to demonstrate that the two systems rely on representations in the same format. A growing body of evidence has indicated that visual MI relies on depictive representations (as does visual perception; see Kosslyn et al. 2006). A depictive representation is defined as one in which (a) each part of the representation specifies a part of the corresponding object and (b) the distances between the different parts in the representation preserve the corresponding distances between the parts of the object (e.g., Kosslyn, 1994; Kosslyn et al. 2006). Thus, by definition, information-processing systems that use depictive representations—such as visual MI and visual perception—will be disrupted to a larger extent by structured visual input that depicts information than by unstructured visual input. Structured visual input can consist of fragments of shapes that are positioned in specific parts of space, which include the crucial elements of depictive representations. In fact, Turvey (1973) demonstrated that the process of visually recognizing objects—which relies on depictive representations—is more impaired by backward visual masking using structured visual patterns (i.e., visual stimuli consisting of target stimulus fragments) than by masking using unstructured visual patterns (i.e., visual stimuli consisting of random black and white dots).

We reasoned that if depictive representations are used in both visual WM and visual MI, both types of tasks should be more impaired by irrelevant visual input composed

of structured visual patterns than they would be by unstructured visual patterns (see the description of pattern types in the [Experiment 1 Method](#) section). On the other hand, if WM does not rely on the retention of depictive representations, we would expect no more interference from structured visual patterns than from unstructured visual patterns—as opposed to the relative amounts of interference expected in the MI task.

In short, the rationale for using DVN as an interference task rests on the fact that DVN triggers a series of events—starting at the retina—that produces representations in the brain. Thus, each DVN display will produce visual representations supported by cortical activation in the visual system. In addition, as noted above, structured DVN includes the key characteristics of a depictive representation—and hence should interfere with depictive representations of other stimuli. Specifically, structured DVN has elements included in depictions of letters, and these elements are arrayed in space. Thus, we hypothesized that if participants rely on depictive representations of the letters in the WM and MI tasks, structured DVN incorporating parts that resemble those of the stimulus letters and that occur in different positions in space should produce more interference (more errors and/or longer response times) than does unstructured DVN, which does not have such characteristics.

## Experiment 1

In this experiment, participants judged the figural properties of letters on the basis either of visual mental images of these letters (MI task) or of letters briefly displayed visually (WM task). In the WM task, in order to limit the activation of previously stored visual knowledge and to prevent generation of mental images, we used letters from the Hebrew or Cyrillic alphabets (the participants did not know these alphabets). Participants performed both tasks during two interference conditions: unstructured DVN (similar to that used by Quinn & McConnell, 1996) and structured DVN (visual patterns with fragments resembling pieces of the letters that were being evaluated). We also included a control condition with no dynamic visual interference (i.e., a uniform gray background).

### Method

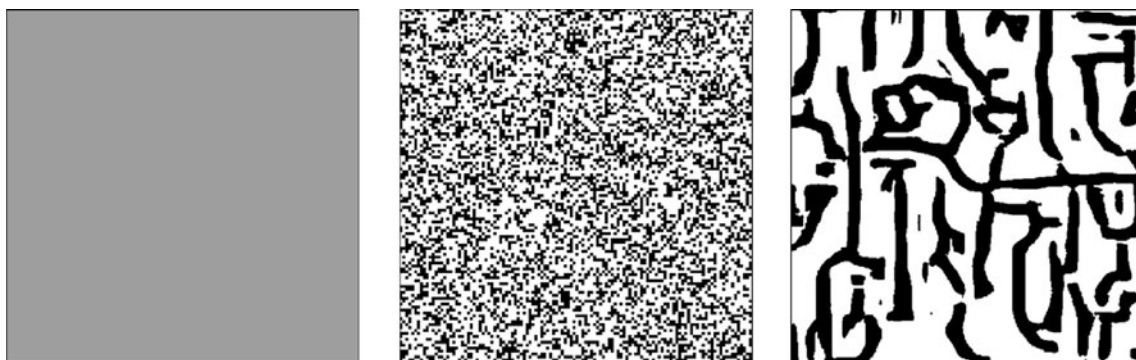
**Participants** We recruited 60 volunteers with normal or corrected-to-normal vision from Harvard University and the local community (34 females and 26 males, with an average age of 22.1 years; 11 of the participants were left-handed). Data from 3 additional participants were not analyzed because they performed at chance levels—hence, it was not clear whether they actually tried to

perform the task. Participants received either a cash payment or course credit. All participants provided written consent and were tested in accordance with national and international norms governing the use of human research participants. The research was approved by the Harvard University Faculty of Arts and Sciences Committee on the Use of Human Subjects.

**Materials** Stimuli were presented on a 19-in. Apple monitor (1,680×1,050 pixels resolution, and refresh rate of 75 Hz) using PsyScope X software running under Mac OS X. All stimuli were presented on a uniform dark background throughout the entire experiment. The stimuli were 240-point saturated black uppercase letters on a 480×480 pixel white background. In the MI task, letters were from the Roman alphabet. In the WM task, letters were from the Hebrew and Cyrillic alphabets, to guard against the participants being so familiar with the stimuli that they could generate mental images of them. We assessed each participant’s knowledge of Hebrew and Cyrillic letters at the outset of the study and did not test anyone who knew either alphabet. In addition, for the practice trials, we created nine stimuli in which the digits 1–9 were displayed (with the same properties as the letter stimuli).

For the MI task, we also prepared 300-ms audio files of the spoken name of each letter and number. Participants judged the letters in terms of four visual properties—namely, whether the letters had any curved lines, diagonal lines, an enclosed space, or a symmetrical form. As was shown by Thompson, Kosslyn, Hoffman, and van der Kooij (2008), curved and diagonal lines are explicit visual properties—stored as such in long-term memory—whereas enclosed space and symmetrical form are implicit properties—not included explicitly in the internal representation of the letter. We created one audio file for each property by using abbreviated words—respectively “curve” for *curved line*, “diag” for *diagonal line*, “close” for *enclosed space*, and “sym” for *symmetrical form*. Each audio file was approximately 200 ms in duration. We chose the 26 letters selected from the Hebrew and Cyrillic alphabets in order to equate as much as possible the occurrence of the four visual properties with respect to the occurrence for the 26 Roman alphabet letters. The four properties were present as follows, in the Roman letters and the Hebrew and Cyrillic letters, respectively: 42% versus 55% for curved lines, 38% versus 35% for diagonal lines, 27% versus 38% for enclosed spaces, and 61% versus 45% for symmetrical forms (all  $t_s > .10$ ).

In order to produce the two DVN conditions, structured versus unstructured, we first created 20 different black-and-white images (480×480 pixels; see Fig. 1). In the structured DVN condition, the patterns preserved the short- and mid-



**Fig. 1** Example of visual masks used in the control (*left*), unstructured dynamic visual noise (*middle*), and structured dynamic visual noise (*right*) conditions

range statistical properties of the letters to be judged (Portilla & Simoncelli, 2000). Each frame was created by using the Portilla and Simoncelli texture analysis/synthesis code. A dense array of characters of the type and font used in the study (different for each frame) was employed as input texture. The main parameters used were four spatial scales, four orientations, and a  $9 \times 9$  spatial neighborhood (Portilla & Simoncelli, 2000). Twenty iterations were used for the synthesis loop.

We created different noise patterns for the two tasks, because different letters were used (Roman vs. Hebrew and Cyrillic letters). In the unstructured DVN condition, the patterns were designed by randomly creating a pattern of black and white pixels. Each pattern contained 37.5% black pixels, in order to equate luminance, density, and contrast between these patterns and the ones created in the structured condition. We produced DVN by creating sequences of images as AVI movies (20 ms per frame). In order to avoid habituation to the DVN, in each condition we created eight different sequences of patterns. In addition, for the no-interference control condition, we created a uniform  $480 \times 480$  pixels gray image (RGB 159, 159, 159) that matched the luminance of the images created in the two DVN conditions.

**Procedure** Participants sat approximately 60 cm from the computer screen. All participants performed both the WM and the MI tasks.

In the MI task, participants began by memorizing the appearances of the 26 Roman alphabet letters and the 9 digits. All 35 characters were presented twice in a pseudorandom order (i.e., all characters appeared once before any appeared a second time), for a total of 70 learning trials. On each trial, a picture of the character was presented for 3 s, accompanied by its spoken name, followed by the presentation of structured DVN for 1 s (in order to eliminate any afterimage of the character); participants then visualized the character exactly as it appeared on the screen. Finally, the character reappeared,

and participants were asked to study it and to correct their visual mental image, making it as accurate as possible.

Before the experimental tasks, we gave participants definitions of the four visual properties, and then we asked them to decide whether abstract symbols possessed each of the properties (using the stimuli from Thompson et al. 2008). Following this, we presented 16 trials that included the four audio files of the property names and asked the participants to associate each of the audio files with the corresponding property.

In the MI task, on each trial, the name of a letter was presented aurally, and after 1.5 s—to allow time to generate the visual mental image of the corresponding letter—participants heard the name of one of the properties. In the WM task, a letter was briefly presented visually (25 ms), and after 1.5 s the name of one of the properties was presented aurally. In both tasks, participants decided as quickly and accurately as possible whether the letter possessed the property. Participants used their dominant hand to respond, pressing the “b” key to indicate that the property was present or the “n” key to indicate that it was not. We recorded both the response times (RTs), starting when the property’s audio file stopped, and the nature of the response.

Participants performed each task in three conditions: control, structured DVN, and unstructured DVN. We counterbalanced the order of the MI and WM tasks over participants. In the MI task, DVN, or the gray background in the control condition, started 500 ms before the auditory presentation of the letter’s name and stopped when the participant pressed either response button. This procedure ensured that participants would not simply memorize verbally the name of the letter and the probed property and wait for the interference to stop before generating a mental image of the letter. In the WM task, DVN or a gray background was presented for 1.5 s, starting 10 ms after the offset of the letter and ending when the property name was presented. We used unfamiliar letters as stimuli in the WM task in order to

discourage participants from generating mental images of them.

In each task, participants performed 168 trials. Participants were tested on 56 letter/property pairs in each of the three conditions, but the three conditions were intermixed. For each of the 56 letter/property pairs, all letters were presented once before any appeared a second time, each letter was presented either two or three times, each property name was presented exactly 14 times (for seven of the presentations, the letter possessed the property, and for seven it did not), and DVN movies—structured and unstructured—were randomly associated with the pairs and appeared once before being presented a second time. The presentation of the pairs was randomized, except that no more than three consecutive trials could appear with the same correct response, the same condition, the same letter, or the same property. In both tasks, participants performed 16 practice trials, in which digits were presented as stimuli, prior to the actual experimental trials.

Finally, participants completed a debriefing questionnaire at the end of each task, to ensure that they did not infer the purpose of the experiment and that they had followed the instructions at least 75% of the time. Participants whose data were analyzed reported having followed the instructions on an average of more than 95% of the trials.

## Results

For each participant in each condition in each task, we averaged the RTs on correct trials and computed the number of errors. Preliminary analyses revealed no effect of gender, no effect of the order of the tasks, and no interaction between these factors on RTs or error rates (ERs). Thus, we pooled the data over these variables and do not address them in the following analyses.

In the subsequent analyses of the results, all factors were within-participants (so that we computed repeated measures ANOVAs), and when we compared two means, we computed paired-samples one-tailed *t* tests in accordance with our hypotheses; all  $\alpha$  levels for *t* tests were adjusted with a Bonferroni correction.

We first performed a 2 (MI vs. WM)  $\times$  3 (control vs. unstructured DVN vs. structured DVN)  $\times$  2 (implicit vs. explicit visual properties) ANOVA on the ERs, which revealed that the participants made more errors in the WM task,  $F(1, 59) = 13.35$ ,  $p < .001$ ,  $\eta_p^2 = .19$ ; that they made different numbers of errors in the different interference conditions,  $F(2, 118) = 39.36$ ,  $p < .0001$ ,  $\eta_p^2 = .40$ ; and that the two effects interacted,  $F(2, 118) = 20.79$ ,  $p < .0001$ ,  $\eta_p^2 = .26$ ; the three-way interaction and the main effect of the type of visual property to judge, however, failed to reach significance,  $F_s < 1$ . We next considered RTs, and again we found that the tasks

differed,  $F(1, 59) = 41.96$ ,  $p < .0001$ ,  $\eta_p^2 = .42$ , and that interference conditions tended to have different effects,  $F(2, 118) = 2.81$ ,  $p = .06$ , but now we found no hint of an interaction between the two variables,  $F < 1$ .

As is shown in Table 1, in the MI task, participants committed significantly more errors on structured DVN trials ( $M = 8.5\%$ ) than on unstructured DVN trials ( $M = 6.4\%$ ),  $t(59) = 2.94$ ,  $p < .01$ ,  $d = 0.36$ . The same was true in the WM task, where participants committed more errors on the structured DVN trials ( $M = 14.5\%$ ) than on the unstructured DVN trials ( $M = 8.7\%$ ),  $t(59) = 7.26$ ,  $p < .0001$ ,  $d = 0.64$  (see Table 1). The interaction showed that the difference between structured and unstructured DVN was greater in the WM task than in the MI task,  $F(1, 59) = 13.71$ ,  $p < .0001$ ,  $\eta_p^2 = .19$ . In addition, in the MI task, the participants made fewer errors in the control condition ( $M = 6.8\%$ ) than in the structured DVN condition,  $t(59) = 2.45$ ,  $p < .05$ ,  $d = 0.29$ , but made a number of errors comparable to the control condition in the unstructured DVN condition,  $t < 1$ . In contrast, in the WM task, error rates were lower in the control condition ( $M = 7.3\%$ ) than in either the structured or the unstructured DVN condition: respectively,  $t(59) = 8.99$ ,  $p < .0005$ ,  $d = 0.84$ , and  $t(59) = 2.5$ ,  $p < .05$ ,  $d = 0.20$ .

In each task, we found no significant difference in the time taken to judge the properties of the letters in the two DVN conditions ( $t_s < 1$ ). Thus, the effect of the interference conditions on the ERs could not be attributed to a speed–accuracy trade-off.

In the WM task, each letter was repeated an average of six times. Hence, the appearance of these letters might have become familiar through the course of the task, allowing participants to generate mental images to perform the WM task. We reasoned that if participants came to generate mental images of the letters in the WM task during the later trials, the pattern of interference should differ between the first and the second half of the trials. To evaluate this possibility, we conducted a 2 (first half of the trials vs. second half of the trials)  $\times$  3 (control vs. unstructured DVN vs. structured DVN) repeated measures ANOVA. The results revealed that participants committed the comparable numbers of errors in the first and the second half of trials,  $F < 1$ , and that the pattern of interference did not differ in the first and the second half of trials, as witnessed by the lack of an interaction,  $F < 1$ . However, participants committed different numbers of errors in the three conditions,  $F(1, 118) = 60.16$ ,  $\eta_p^2 = .51$  (see Table 2).

We also analyzed separately the ERs and RTs for the two types of properties—explicit (i.e., curved and diagonal line) and implicit (i.e., symmetrical form and enclosed space)—in both tasks during the two types of interference. The key results is easily summarized: In no case did we find an interaction between the type of property and the type of interference,  $F < 1$  in all cases.

**Table 1** Experiment 1: Response times (RTs, in milliseconds) and error rates (ERs, percentages) in the mental imagery (MI) task and the working memory (WM) task, in three interference conditions (control, unstructured DVN, and structured DVN)

	RTs			ERs		
	Control	Unstructured DVN	Structured DVN	Control	Unstructured DVN	Structured DVN
MI task	748 (256)	763 (271)	765 (274)	6.8 (5.03)	6.4 (4.77)	8.5 (6.58)
WM task	588 (166)	588 (152)	604 (164)	7.3 (6.85)	8.7 (7.65)	14.5 (9.88)

Standard deviations appear in parentheses

## Discussion

The participants performed the MI task and the WM task less accurately when structured patterns were presented than when unstructured patterns were presented. This result suggests that depictive representations were processed in both tasks. We used unfamiliar letters in the WM condition of this experiment in order to ensure that participants could not base their judgments on visual mental images. However, each letter was presented an average of six times, and thus participants might have become familiar enough with the letters to be able to generate mental images of them. However, we note that (a) participants reported in a debriefing questionnaire that they followed the instructions on more than 95% of the trials by not generating mental images of the unfamiliar letters, and that (b) the patterns of interference were comparable in the first and the second halves of the WM task. In light of these findings, we are confident that the participants did not rely on mental images to perform the WM task.

In the MI task, structured DVN presumably interfered more than did unstructured DVN because visual mental images were generated in a visual buffer—implemented in early visual brain areas—that depicts information (see Kosslyn et al. 2006). In these brain areas, the spatial layout of the surface of an object is represented by the spatial layout of the patterns of activation on the cortex (e.g., Tootell, Hadjikhani, Mendola, Marrett, & Dale, 1998). Thus, because the representations of structured DVN overlapped more with the representations of the letters (i.e., visual

mental images of the letters) than did the representations of unstructured DVN, more interference occurred.

In the WM task, structured DVN presumably disrupted retention of the figural properties of Hebrew and Cyrillic letters more than did unstructured DVN because visual information was held in a visual cache implemented in topographically organized brain structures. In fact, according to Logie and van der Meulen (2009), short-term retention of visual information takes place in a visual cache implemented in the posterior parietal lobe. Given that portions of the posterior parietal lobe are topographically organized (Sereno, Pitzalis, & Martinez, 2001), they would depict information—which explains why structured DVN produced more interference than the unstructured condition in the WM task.

One could argue that the stronger interference produced by structured DVN simply reflects differences between the two types of patterns. Structured patterns might have contained more semantic information, been more complex (i.e., denser), or been more dynamic, and thus could have captured attention more effectively or evoked more eye movements. In fact, there is evidence that interference effects can be modulated by the degree of semantic information contained in irrelevant visual input (e.g., Logie, 1995), by the density of the DVN (McConnell & Quinn, 2004), and by exogenous visual attention and eye movements (e.g., Pearson & Sahraie, 2003).

However, none of these three alternative explanations can account for our results. First, the patterns in the structured DVN stimuli were not composed of parts that conveyed any meaning. The only requirement for those visual patterns was that parts of the patterns resembled fragments of letters. None of the stimuli could be recognized as a letter, nor as a fragment of a specific letter. Alternatively, one could argue that interference could have occurred at a semantic level in the MI condition because figural properties of the letters were explicitly encoded in its internal representation. In this account, structured DVN produced interference because participants encoded figural properties (such as curved lines) of the structured visual patterns that overlapped

**Table 2** Experiment 1: Error rates (as percentages) in the three interference conditions (control, unstructured DVN, and structured DVN) in the first and second halves of trials of the working memory task

	Control	Unstructured DVN	Structured DVN
Trials 1–84	7.1 (6.73)	8.6 (7.75)	14.3 (9.65)
Trials 85–168	7.5 (7.88)	8.8 (9.5)	14.6 (12.49)

Standard deviations appear in parentheses

with the description of the figural properties explicitly encoded in the internal representations of the letter in long-term memory (LTM). However, this semantic account cannot explain interference when participants judged implicit figural properties (such as the symmetrical form) of the letters. In fact, implicit figural properties—by definition—are not explicitly included in internal representations of the letters, and thus one would necessarily need to compute additional properties of the letter to determine whether it possesses such implicit figural properties (see Thompson et al. 2008). Given that we found no hint that the type of figural property modulates interference (with  $F < 1$  for the interaction between explicit and implicit properties and type of interference), differences in the semantic information conveyed by the two DVN conditions are unlikely to have produced the pattern of interference observed. Second, we equated density and luminosity in the structured and unstructured DVN conditions, which ensured that our results do not arise from such differences. Third, in both DVN conditions, the succession of visual patterns created a sense of movement; thus, there is no reason to think that eye movements or exogenous attention differed between the two DVN conditions.

In addition, the pattern of interference reported might be partially attributable to methodological differences between the MI and the WM conditions—although these differences were absolutely necessary to ensure that mental images were not generated in the WM task. For example, well-known Roman letters were used as stimuli in the MI task, but unfamiliar Hebrew and Cyrillic letters were used in the WM task. Therefore, visual mental images were generated on the basis of information stored in LTM, whereas LTM did not contribute to performance in the WM task. However, the fact that representations were created on the basis of information stored in LTM in the MI task and on the basis of direct visual input in the WM task does not affect their format, which is the issue being investigated in the present study. In fact, depictive representations can be generated on the basis of information stored in LTM, as in MI, or on the basis of direct visual inputs, in visual perception (see Kosslyn et al. 2006). An additional methodological difference between the two tasks is the timing of the DVN stimuli. DVN lasted until participants provided a response in the MI task, but it stopped after 1.5 s in the WM task—at the time the query was presented. This timing difference was critical in order (a) to prevent participants from delaying the generation of the visual mental images until after the presentation of the query in the MI task and (b) to ensure that we only interfered with the retention of

information in WM and not with the recall of this information in the WM task.

One could also argue that the WM task we designed required participants to retain iconic images rather than to store the representations of the letters in VSTM per se—which in turn might explain why structured visual masks produced more interference. However, this is unlikely, given that the 1.5-s retention interval we used exceeds the capacity (~100 ms) of a purely sensory store such as iconic memory (e.g., Phillips & Christie, 1977a, 1977b; Sperling, 1963). The effect of the structured visual masks could also stem from interference with the consolidation of information in WM. Indeed, the retention of visual information in WM is impaired when objects are presented in the same locations as to-be-memorized stimuli within 400 ms (e.g., Sun, Zimmer, & Fu, 2011; Vogel, Woodman, & Luck, 2006). According to the consolidation account, structured DVN would produce more interference than unstructured DVN because visual information in the masks would be integrated erroneously with the information in the letters (see Sun et al. 2011). We address these issues in [Experiment 3](#).

Finally, the different effects of unstructured DVN on performance in the two tasks were unexpected. Unstructured DVN interfered with performance in the WM task, whereas it had no effect (relative to the control condition) in the MI task. Two different accounts could explain why retention of the figural properties of letters in the WM task was affected by unstructured DVN. On the one hand, unstructured visual patterns might have masked the representations of the letters, which disrupted their being entered into the visual cache. In fact, studies have demonstrated that visual patterns that do not resemble the visual stimuli can produce backward masking (e.g., Delord, 1998). On the other hand, some memory tasks are affected by unstructured DVN. In fact, DVN has been shown to interfere with the recognition of the precise size of a circle (McConnell & Quinn, 2004), of the precise font of a letter (Darling, Della Sala, & Logie, 2007, 2009), and of the precise color of a dot (Dent, 2010). Thus, consistent with the idea that DVN selectively affects visual WM when precise information needs to be maintained, interference might have occurred in the WM task because precise figural information needed to be retained in order to perform the task.

How do we explain why unstructured DVN did not interfere with the MI task, relative to the control condition? Given that (a) high-resolution visual mental images rely on early visual areas (Kosslyn, Pascual-Leone, Felician, Camposano, Keenan, Thompson and Alpert 1999; see Kosslyn & Thompson, 2003, for a review) and that (b) patterns of activation in these brain areas are transient (in

order to prevent visual persistence after eye movements), we expect any visual input to interfere with a representation in the visual buffer. Thus, this result may have occurred because participants had to perform the MI task while looking at a visual stimulus. [Experiment 2](#) was an attempt to test this hypothesis.

## Experiment 2

In this experiment, we asked a new group of participants to perform the same MI task as in [Experiment 1](#), but under two conditions: the control condition of [Experiment 1](#) (i.e., participants look at a uniform gray background on the computer screen) and a blindfolded condition. We reasoned that if any visual input (such as the presentation of a gray background) interferes with MI (because mental images rely on early visual areas in which activation is disrupted by any visual input), participants should perform more poorly when using MI to judge the visual properties of letters while looking at a uniform gray background than with their eyes closed.

### Method

**Participants** A group of 24 right-handed participants from Harvard University and the local community volunteered to take part in this experiment (13 females and 11 males, with an average age of 22.6 years). None of them had taken part in [Experiment 1](#). All participants had normal or corrected-to-normal vision and received payment or course credit for their participation. The research was approved by the Harvard University Faculty of Arts and Sciences Committee on the Use of Human Subjects. All participants provided written consent and were tested in accordance with national and international norms governing the use of human research participants.

**Materials** The materials were the same as those used in [Experiment 1](#).

**Procedure** Before the MI task, as in [Experiment 1](#), participants learned to generate mental images of 26 Roman letters, to identify the four visual properties, and to identify the abbreviated audio versions of the property names. Experimental trials were blocked and were identical to those in the MI task of [Experiment 1](#), except that in the blindfolded condition a “beep” was presented 500 ms before the name of the letter, to signal the beginning of a new trial. Participants performed 56 experimental trials (i.e., 56 pairs of letter/property) in each experimental condition (looking

at a gray field vs. blindfolded), for a total of 112 trials, under exactly the same constraints as in [Experiment 1](#). The order of the two conditions was counterbalanced across participants. Before each experimental condition, participants performed 16 practice trials.

As in [Experiment 1](#), participants reported having followed the instructions on an average of more than 95% of the trials.

### Results

As in [Experiment 1](#), in each condition we separately averaged RTs (on correct trials) and ERs. Preliminary analyses revealed no effect of the order of the conditions or of the gender of the participants, as well as no interaction between order and gender. Thus, we pooled over these variables in subsequent analyses.

As predicted, participants made more errors in the MI task when they looked at a uniform gray background on the computer screen ( $M = 9.4\%$ ) than when they were blindfolded ( $M = 7.2\%$ ),  $t(23) = 2.51$ ,  $p < .01$ ,  $d = 0.52$  (see [Table 3](#)). However, they took the same amount of time to make judgments in the eyes-open ( $M = 822$  ms) and in the blindfolded condition ( $M = 823$  ms),  $t < 1$ .

### Discussion

The results confirmed our hypothesis: the simple fact of having a visual input during our visual MI task is sufficient to produce interference. Thus, the apparent lack of an interference effect of the unstructured DVN in the MI task of [Experiment 1](#) actually arose from the fact that the control condition produced visual interference.

However, at first glance the results of the two experiments might seem inconsistent: Participants were faster and more accurate in the control condition of [Experiment 1](#) (i.e., presentation of a gray background) than in the blindfolded condition of [Experiment 2](#). Thus, one could argue that not all visual inputs interfere with MI. We have two responses to this concern. First, different groups of people participated in the two experiments, which precludes a direct comparison of ERs and RTs

**Table 3** [Experiment 2](#): Response times (RTs, in milliseconds) and error rates (ERs, percentages) in the mental imagery (MI) task in the eyes-open and blindfolded conditions

	RTs		ERs	
	Eyes Opened	Blindfolded	Eyes Opened	Blindfolded
MI task	822 (248)	823 (280)	9.4 (4.71)	7.2 (3.71)

Standard deviations appear in parentheses



across them. Participants in [Experiment 2](#) probably, on average, simply performed the MI task less efficiently than did the participants in [Experiment 1](#). Second, some of the behavioral differences might reflect differences in the procedures used in [Experiment 1](#) and [2](#). In [Experiment 1](#), we presented the experimental conditions intermixed, whereas in [Experiment 2](#), because participants had to be blindfolded, we blocked the experimental conditions. Taken together, despite the differences in performance in the two experiments, the results converge in showing that low-level perceptual processes can be disrupted during visual MI merely by looking at a uniform gray background.

One might also ask whether a uniform gray background can disrupt the retention of information in WM. We reasoned that the retention of visual information should not be disrupted by a uniform gray background in WM, if the visual cache is implemented in the posterior parietal lobe: In contrast to early visual areas, activation in the posterior parietal lobes is not disrupted by all visual inputs. [Experiment 3](#) tested this hypothesis.

### Experiment 3

In [Experiment 3](#), a new group of participants performed a variant of the WM task that we had administered in [Experiment 1](#). This task was the same as the previous one, except that (a) participants performed the task in an additional condition in which no visual input occurred, (b) the gray background and the two DVN conditions started 500 ms—not 10 ms—after the offset of the letter, and (c) to create a no-visual-input condition, participants performed the task in complete darkness. We presented one of the three interference conditions (gray background, unstructured DVN, structured DVN) 500 ms after the offset of the letter in order to ensure that interference occurred with representations already consolidated in WM—not with the encoding or the consolidation of these representations in the visual cache (see Sun et al. 2011; Vogel et al. 2006).

If the visual cache is not sensitive to all kinds of visual inputs (such as the presentation of a gray background), participants' performance should not be disrupted more by looking at a uniform gray background than by no visual input. In addition, if the retention of the figural properties of letters in WM was affected by unstructured DVN because unstructured visual patterns masked the representations of the letters, which in turn disrupted their being entered into the visual cache, we would expect to see no effect of unstructured DVN on participants' performance when the representations are already consolidated in the visual cache (such as in this experiment). Finally, if representations held

in WM are depictive, structured DVN should produce more interference than any of the other three conditions.

### Method

*Participants* A group of 17 psychology students from Paris Descartes University with normal or corrected-to-normal vision participated in this experiment (10 females and 7 males, with an average age of 19.3 years; 3 of the participants were left-handed). None of them had taken part in [Experiment 1](#) or [2](#). Participants received course credit for their participation. Data from 1 additional participant were not analyzed because he performed at a chance level—hence, it was not clear whether he actually tried to perform the task. All participants provided written consent and were tested in accordance with national and international norms governing the use of human research participants.

*Materials* The materials were the same as those used in the WM task of [Experiment 1](#).

*Procedure* Participants were seated in front of a computer screen in complete darkness. The contrast of the computer screen was adjusted to prevent residual afterimages of the stimuli displayed. As in [Experiment 1](#), before the WM task we asked participants to learn to identify the four visual properties and to identify the abbreviated audio versions of the property names. Participants performed 56 experimental trials (i.e., 56 letter/property pairs) in each experimental condition (no visual input vs. looking at a gray field vs. unstructured DVN vs. structured DVN), for a total of 224 trials. Experimental trials were presented randomly, except that the same letters could not appear once before appearing twice, and the same visual property or the same condition could not appear more than three times in a row. Before the first experimental trial, participants performed 16 practice trials.

As in [Experiments 1](#) and [2](#), a debriefing questionnaire revealed that participants followed the instructions on an average of more than 95% of the trials.

### Results

As in [Experiments 1](#) and [2](#), in each condition we separately averaged RTs (on correct trials) and ERs. Preliminary analyses revealed no effect of the gender of the participants. Thus, we pooled over this variable in subsequent analyses. All  $\alpha$  levels for  $t$  tests were adjusted with a Bonferroni correction.

A one-way repeated measures ANOVA revealed a significant effect of the experimental condition on the number of errors committed in the WM task,  $F(3, 48) =$

7.66,  $p < .0001$ ,  $\eta_p^2 = .32$  (see Table 4). Participants committed more errors in the structured DVN condition ( $M = 12.8\%$ ) than in all three other conditions: respectively,  $M = 8.2\%$  in the unstructured DVN condition,  $t(16) = 3.58$ ,  $p < .01$ ,  $d = 0.72$ ;  $M = 6.3\%$  in the gray background condition,  $t(16) = 3.02$ ,  $p < .025$ ,  $d = 1.02$ ; and  $M = 5.5\%$  in the no-visual-input condition,  $t(16) = 4.38$ ,  $p < .0001$ ,  $d = 1.12$ . Looking at a gray background did not produce more interference than occurred in the no-visual-input condition,  $t < 1$ . Finally, participants were as accurate in the unstructured DVN condition as in the gray background condition and in the no-visual-input condition: respectively,  $t(16) = 1.06$ ,  $p = .91$ , and  $t(16) = 1.77$ ,  $p = .27$ .

We found no effect of the different interference conditions on the RTs, as revealed by a one-way repeated measures ANOVA,  $F(3, 48) = 1.32$ ,  $p = .28$ .

## Discussion

The results of this experiment shed light on three critical issues raised in Experiment 1: First, given that the gray background produced as much interference as the no-visual-input condition, we have evidence that not all visual inputs interfere with representations held in WM. Clearly, MI and WM are not one and the same. Second, even though they are not identical, MI and WM appear to rely on representations in a depictive format: As in Experiment 1, structured DVN interfered with performance in the WM task more than did unstructured DVN. Finally, unstructured DVN did not interfere with the retention of information in WM when presented 500 ms after the offset of the letters—after representations were already encoded in WM (Sun et al. 2011; Vogel et al. 2006). Thus, unstructured DVN probably interfered with performance in the WM task in Experiment 1 because it disrupted the representations of the letters, which in turn disrupted their being encoded into the visual cache.

Although the results of the three experiments converge in showing that visual MI and visual WM operate on representations that share the same format (i.e., depictive representations), one could argue—consistent with a unitary account of WM (e.g., Cowan, 2005)—that the pattern of interference reported in all three experiments simply

reflects differences in the attentional demands of the different interference conditions. According to the unitary account of WM, structured DVN produced a greater disruption of participants' performance in the MI and the WM task because looking at structured visual masks recruits more attentional resources than does looking at randomly changing black and white dots (i.e., unstructured DVN). Differences in attentional demands could also explain why looking at a gray background impaired participants' performance in the MI task relative to a condition in which the participants were blindfolded. We investigated the issue of attentional demands in Experiment 4.

## Experiment 4

If the results reported in the previous experiments simply reflect differences in attentional demands, we would expect participants' performance to be disrupted more by the more demanding interference condition—regardless of the nature of the primary task performed (e.g., visual or verbal). In this experiment, a new group of participants performed a verbal WM task in the four experimental conditions administered in Experiment 3 (no visual input, gray background, unstructured DVN, and structured DVN) and in an articulatory suppression condition. Participants memorized series of five digits (from 1 to 9) presented aurally. After a 5-s retention interval, they recalled the five digits in the order that they had been presented. Interference conditions were presented during the retention interval.

We reasoned that if the attentional demands of structured DVN exceed those of unstructured DVN, structured DVN should disrupt participants' performance in a verbal WM task to a greater extent than would unstructured DVN. Similarly, looking at a gray background should disrupt recall more than would a condition with no visual input. We included the articulatory suppression condition—a typical verbal interference task—to provide evidence that we had enough statistical power to detect significant differences between the different interference conditions, if we found that the four interference conditions used in previous experiments had no effect on participants' performance in the verbal WM task.

**Table 4** Experiment 3: Response times (RTs, in milliseconds) and error rates (ERs, percentages) in the working memory task in four interference conditions (no visual input, gray background, unstructured DVN, structured DVN)

RTs				ERs			
No Visual Input	Gray Background	Unstructured DVN	Structured DVN	No Visual Input	Gray Background	Unstructured DVN	Structured DVN
1,057 (219)	1,107 (231)	1,084 (161)	1,054 (187)	5.5 (5.2)	6.3 (4.8)	8.2 (4.8)	12.8 (7.7)

Standard deviations appear in parentheses

## Method

**Participants** We recruited 14 volunteers with normal or corrected-to-normal vision from Paris Descartes University (8 females and 6 males, with an average age of 20.2 years; 2 of the participants were left-handed). Participants received course credit. All participants provided written consent and were tested in accordance with national and international norms governing the use of human research participants.

**Materials** Stimuli were presented on a 19-in. Apple monitor (1,680×1,050 pixel resolution and refresh rate of 75 Hz) using PsyScope X software running under Mac OS X. We prepared 200-ms audio files of the spoken names of nine digits (1 to 9) for the verbal WM task. The gray background and the two DVN conditions were identical to those used in the previous experiments.

**Procedure** Participants sat in complete darkness approximately 60 cm from the computer screen. On each trial, a series of five digits was presented aurally at a rate of one every 500 ms. Participants were instructed to maintain their gaze on a fixation point displayed at the center of the computer screen during the presentation of the digits. Three hundred ms after the presentation of the last digit of the series, a “beep” indicated the beginning of the retention interval. After 5 s, a second “beep” signaled participants to recall the series of digits in the order in which they had been presented. Participants wrote the digits on a response sheet, which comprised five boxes arranged horizontally for each series. In the no-visual-input condition, the screen remained black during the retention interval. On trials in which participants were asked to perform articulatory suppression during the retention interval, “articulatory suppression” was displayed at the center of the screen for 1.5 s before the presentation of the first digit.

Participants received 6 series in each of the interference conditions (for a total of 30 series) in a random order—except that no more than two series under the same interference condition occurred in a row. Before the first experimental trials, participants performed five practice trials, one in each of the experimental conditions.

## Results

We separately averaged ERs in each of the five conditions. All  $\alpha$  levels for  $t$  tests were adjusted with a Bonferroni correction.

A one-way repeated measures ANOVA revealed that the different interference conditions produced different

numbers of errors,  $F(4, 52) = 18.18, p < .0001, \eta_p^2 = .58$  (see Table 5). Participants committed more errors in the articulatory suppression condition ( $M = 44.1\%$ ) than in, respectively, the no-visual-input condition ( $M = 10.7\%$ ),  $t(13) = -4.27, p < .005, d = 1.56$ ; the gray background condition ( $M = 10.3\%$ ),  $t(13) = -4.79, p < .0001, d = 1.64$ ; the unstructured DVN condition ( $M = 11.9\%$ ),  $t(13) = -4.68, p < .0001, d = 1.56$ ; and the structured DVN condition ( $M = 8.3\%$ ),  $t(13) = -4.98, p < .0001, d = 1.72$ . However, error rates were comparable in all four visual interference conditions,  $ps > .51$ .

## Discussion

As expected, articulatory suppression interfered with the retention of digits. However, the four interference conditions used in previous experiments had comparable effects on participants’ performance in the verbal WM task. The lack of difference among the four conditions used in the previous experiments is unlikely to reflect a lack of statistical power, given that we found a significant difference between the error rates in these interference conditions and in the articulatory suppression condition. However, one could argue that the effect of the articulatory suppression condition is so massive, compared to the other four conditions, that the statistical power needed to detect that difference is less than would be needed to detect the differences among the four interference conditions used in Experiments 1, 2, and 3. We note that—if anything—participants committed fewer errors in the structured DVN condition than in any of the three other conditions in the verbal WM task, even though the structured condition had systematically produced more interference than any other interference condition in the visual MI and WM tasks.

The results from this verbal WM task suggest that the different interference conditions used in Experiments 1, 2, and 3 do not differ in their attentional demands. Therefore, the patterns of interference reported in the present study cannot be explained by a general difference in the attentional demands of those interference conditions.

**Table 5** Experiment 4: Error rates (as percentages) in the verbal working memory task in an articulatory suppression condition and four visual interference conditions (no visual input, gray background, unstructured DVN, structured DVN)

Articulatory Suppression	No Visual Input	Gray Background	Unstructured DVN	Structured DVN
44.1 (27.4)	10.7 (15.5)	10.3 (10.2)	11.9 (10.2)	8.3 (10.8)

Standard deviations appear in parentheses

## General discussion

Taken together, the results demonstrate that representations processed in visual MI and in visual WM are interfered with more strongly by structured visual patterns (which have elements that are shared with objects in mental images) than by unstructured visual patterns (which have few, if any, such elements); the structured visual patterns we used included fragments of the stimuli, laid out in space.

We have argued that these results implicate the use of depictive representations. One alternative is that participants stored the letters as descriptions of features (such as straight lines and curved lines), and generated the same types of descriptions of the structured DVN. If so, one might then expect that the two sorts of descriptions would interfere with each other. However, we have direct evidence against this view: The types of interference had comparable effects when participants evaluated explicit properties (such as the presence of curved lines) and implicit properties (such as symmetry). By definition, implicit properties are not noted explicitly in the representation, but are only derived when needed (see Thompson et al. 2008). If a descriptive representation had been used, participants should have required more time and made more errors with the implicit properties than with the explicit properties in the structured DVN condition (and we should not have found such a difference in the unstructured DVD condition). However, we found no hint of such an interaction—which is as would be expected if the participants used depictive representations “to read” the information needed to evaluate the query.

Thus, these results provide evidence that representations used in both visual MI and visual WM depict information. Although visual MI and visual WM have different functions, our findings suggest that these two functions rely on representations that share the same format.

In addition, our findings shed light on some aspects of the debate about the degree of overlap between MI and WM. By providing evidence that MI and WM rely (at least in part) on representations in the same format, we have satisfied a necessary requirement for any theory that assumes a degree of overlap between them. In our view, MI and WM are information-processing systems, and by definition, information-processing systems take an input and produce an output. Critically, inputs and outputs are representations that convey information only because specific processes are available. From this point of view, if two systems rely at least in part on the same type of representations, they also must rely at least in part on common cognitive processes; certain processes are only suited for certain types of representations—for instance, one can “mentally rotate” an object in an image, but not an

auditory representation of a word. However, clearly some processes are not shared in MI and WM. For example, there is no need for generation processes in WM, whereas such processes are at the core of the MI system.

Even if MI and WM rely (at least in part) on depictive representations, which are processed in comparable ways, it is possible that these depictive representations occur in different parts of the brain and play different roles. At least in the monkey brain, some dozen topographically organized areas have been identified (e.g., Felleman & Van Essen, 1991; cf. Sereno et al. 1995). However, most topographically organized brain areas are driven primarily by bottom-up input. For example, in humans visual MI and visual perception share many of the same brain areas—as much as 92% of the same voxels are activated in common during fMRI—precisely because they rely on representations of the same format (Ganis, Thompson, & Kosslyn, 2004). But the earliest visual areas are activated much more strongly during perception than during visual MI. Relatively few topographically organized areas are activated to comparable degrees in perception and imagery, and it is these areas that are the strongest candidates for WM structures. For example, neuroimaging studies of WM have implicated high-level areas in the posterior parietal lobes and frontal lobes (e.g., Jonides, Smith, Koeppel, Awh, Minoshima and Mintun 1993; Postle & D’Esposito, 1999; Rowe, Toni, Josephs, Frackowiak, & Passingham, 2000; Smith, Jonides, Koeppel, Awh, Schumacher and Minoshima 1995), which are also often activated during MI (Kosslyn et al., 2006). This view is consistent with a recent review of the literature showing that perception, visuospatial WM, and visuospatial MI rely on most of the same brain areas (Zimmer, 2008). According to Zimmer, WM is an emergent property of any cognitive process that serves to retain a representation in order to continue processing.

Although our results do not directly provide information on the functional architecture of WM, providing evidence that the visual cache stores visual information in a depictive form poses constraints on theories of MI and WM. First, these results suggest that even after representations are encoded and consolidated in the visual cache, it can be accessed by direct visual input. Second, given that the visual cache stores depictive representations, it might also play a role in maintaining visual mental images. In Kosslyn’s theory (Kosslyn, 1994; Kosslyn et al. 2006), the spatial-properties-processing subsystem, which relies on the posterior parietal cortex, represents an object map (Mesulam, 1990)—laying out the location of objects in a scene or of the parts of an object. The object map not only specifies the locations of shapes, but also for each location has pointers to representations of these shapes (which are represented in the object-properties-processing subsystem within the inferior temporal lobe; Kosslyn, 1994). Given

that a new representation is created in the visual buffer every time the eyes move, the maintenance of visual mental images is unlikely to occur in the visual buffer itself. Rather, image maintenance may well occur in a visual cache, which relies on both the object-properties- and spatial-properties-processing subsystems (and thus relies on the posterior parietal and inferior temporal cortices).

The results reported here not only provide evidence that the visual cache relies, at least in part, on depictive representations, but they also shed light on the contradictory effects of DVN on visual WM reported in the literature. Arguably, one of the factors that could explain why DVN sometimes does not interfere with the short-term retention of visual information is the lack of overlap between the representations fed into the passive store (i.e., visual cache) and the representations produced by the DVN. If WM relies at least in part on depictive representations, interference will then occur only when the representations produced by the visual patterns overlap with the structure of the stored representations—namely, they must depict similar information. For example, Andrade et al. (2002) would have probably found interference with the short-term recognition or recall of Chinese characters if the visual patterns presented during the DVN contained fragments of Chinese characters rather than simply black and white dots.

In short, studying the format of representations used in MI and WM not only helps us understand those systems and previous findings regarding them, it also opens the door to new predictions and new studies that will further illuminate the nature of these mechanisms.

## References

- Andrade, J., Kemps, E., Wermiers, Y., May, J., & Szmalec, A. (2002). Insensitivity of visual short-term memory to irrelevant visual information. *Quarterly Journal of Experimental Psychology*, *55A*, 753–774. doi:10.1080/02724980143000541
- Avons, S., & Sestieri, C. (2005). Dynamic visual noise: No interference with visual short-term memory or the construction of visual images. *European Journal of Cognitive Psychology*, *17*, 405–424.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.
- Cowan, N. (2005). *Working memory capacity*. New York: Psychology Press.
- Darling, S., Della Sala, S., & Logie, R. H. (2007). Behavioural evidence for separating components of visuo-spatial working memory. *Cognitive Processing*, *8*, 175–181. doi:10.1007/s10339-007-0171-1
- Darling, S., Della Sala, S., & Logie, R. H. (2009). Dissociation between appearance and location within visuo-spatial working memory. *Quarterly Journal of Experimental Psychology*, *62*, 417–425. doi:10.1080/17470210802321984
- Dean, G. M., Dewhurst, S. A., & Whittaker, A. (2008). Dynamic visual noise interferes with storage in visual working memory. *Experimental Psychology*, *55*, 283–289.
- Delord, S. (1998). Which mask is the most efficient: A pattern or a noise? It depends on the task. *Visual Cognition*, *5*, 313–338. doi:10.1080/713756789
- Dent, K. (2010). Dynamic visual noise affects visual short-term memory for surface color but not spatial location. *Experimental Psychology*, *57*, 17–26.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47.
- Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Cognitive Brain Research*, *20*, 226–241. doi:10.1016/j.cogbrainres.2004.02.012
- Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S., & Mintun, M. A. (1993). Spatial working memory in humans as revealed by PET. *Nature*, *363*, 623–625. doi:10.1038/363623a0
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: MIT Press.
- Kosslyn, S. M. (1983). Mental representation. In J. R. Anderson and S. M. Kosslyn (Eds.), *Tutorials in Learning and Memory: Essays in Honor of Gordon Bower*. San Francisco, CA: W. H. Freeman.
- Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, *94*, 148–175. Reprinted in A. M. Collins & E. E. Smith (Eds.), (1988). *Readings in cognitive science: A perspective from psychology and artificial intelligence* (pp. 615–642). San Mateo, CA: Morgan Kaufman.
- Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
- Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Thompson, W. L., et al. (1999). The role of Area 17 in visual imagery: Convergent evidence from PET and rTMS. *Science*, *284*, 167–170. doi:10.1126/science.284.5411.167
- Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, *129*, 723–746. doi:10.1037/0033-2909.129.5.723
- Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York: Oxford University Press.
- Logie, R. H. (1995). *Visuo-spatial working memory*. Hillsdale, NJ: Erlbaum.
- Logie, R. H. (2003). Spatial and visual working memory: A mental workspace. In D. Irwin & B. Ross (Eds.), *Cognitive vision: The psychology of learning and motivation* (Vol. 42, pp. 37–78). New York: Elsevier Science.
- Logie, R. H., & van der Meulen, M. A. (2009). Fragmenting and integrating visuospatial working memory. In J. R. Brockmole (Ed.), *The visual world in memory* (pp. 1–32). Hove, U.K.: Psychology Press.
- McConnell, J., & Quinn, J. G. (2000). Interference in visual working memory. *Quarterly Journal of Experimental Psychology*, *53A*, 53–67. doi:10.1080/027249800390664
- McConnell, J., & Quinn, J. G. (2004). Complexity factors in visuo-spatial working memory. *Memory*, *12*, 338–350. doi:10.1080/09658210344000035
- Mesulam, M.-M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology*, *28*, 597–613. doi:10.1002/ana.410280502
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- Pearson, D. G. (2001). Imagery and the visuo-spatial sketchpad. In J. Andrade (Ed.), *Working memory in perspective* (pp. 33–59). Hove, U.K.: Psychology Press.
- Pearson, D. G., & Sahraie, A. (2003). Oculomotor control and the maintenance of spatially and temporally distributed events in visuo-spatial working memory. *Quarterly Journal of Experimental Psychology*, *56*, 1089–1111.

- Phillips, W. A., & Christie, D. F. M. (1977a). Components of visual memory. *Quarterly Journal of Experimental Psychology*, *29*, 117–133. doi:10.1080/00335557743000080
- Phillips, W. A., & Christie, D. F. M. (1977b). Interference with visualization. *Quarterly Journal of Experimental Psychology*, *29*, 637–650. doi:10.1080/14640747708400638
- Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*, 49–71.
- Postle, B. R., & D'Esposito, M. (1999). What-then-where in visual working memory: An event-related fMRI study. *Journal of Cognitive Neuroscience*, *11*, 585–597.
- Quinn, J. G. (2008). Movement and visual coding: The structure of visuo-spatial working memory. *Cognitive Processing*, *9*, 35–43.
- Quinn, J. G., & McConnell, J. (1996). Irrelevant pictures in visual working memory. *Quarterly Journal of Experimental Psychology*, *49A*, 200–215. doi:10.1080/027249896392865
- Quinn, J. G., & McConnell, J. (1999). Manipulation of interference in the passive visual store. *European Journal of Cognitive Psychology*, *11*, 373–389. doi:10.1080/713752322
- Quinn, J. G., & McConnell, J. (2006). The interval for interference in conscious visual imagery. *Memory*, *14*, 241–252. doi:10.1080/09658210500210019
- Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S., & Passingham, R. E. (2000). The prefrontal cortex: Response selection or maintenance within working memory? *Science*, *288*, 1656–1660.
- Schacter, D. L. (1996). *Searching for memory—The brain, the mind, and the past*. New York: Basic Books.
- Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*, 889–893. doi:10.1126/science.7754376
- Sereno, M. I., Pitzalis, S., & Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science*, *294*, 1350–1354.
- Smith, E. E., Jonides, J., Koeppel, R. A., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial versus object working memory: PET investigations. *Journal of Cognitive Neuroscience*, *7*, 337–356. doi:10.1162/jocn.1995.7.3.337
- Sperling, G. A. (1963). A model for visual memory tasks. *Human Factors*, *5*, 19–31.
- Sun, H., Zimmer, H. D., & Fu, X. (2011). The influence of expertise and of visual complexity on visual short-term memory consolidation. *Quarterly Journal of Experimental Psychology*, *64*, 704–729.
- Thompson, W. L., Kosslyn, S. M., Hoffman, M. S., & van der Kooij, K. (2008). Inspecting visual mental images: Can people “see” implicit properties as easily in imagery and perception? *Memory & Cognition*, *36*, 1024–1032.
- Tootell, R. B. H., Hadjikhani, N. K., Mendola, J. D., Marrett, S., & Dale, A. M. (1998). From retinotopy to recognition: fMRI in human visual cortex. *Trends in Cognitive Sciences*, *2*, 174–183.
- Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, *80*, 1–52. doi:10.1037/h0033872
- van der Meulen, M., Logie, R. H., & Della Sala, S. (2009). Selective interference with image retention and generation: Evidence for the workspace model. *Quarterly Journal of Experimental Psychology*, *62*, 1568–1580.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1436–1451. doi:10.1037/0096-1523.32.6.1436
- Zimmer, H. D. (2008). Visual and spatial working memory: From boxes to networks. *Neuroscience and Biobehavioral Reviews*, *32*, 1373–1395.
- Zimmer, H. D., & Speiser, H. R. (2002). The irrelevant picture effect in visuo-spatial working memory: Fact or fiction? *Psychologische Beiträge*, *44*, 223–247.
- Zimmer, H. D., Speiser, H. R., & Seidler, B. (2003). Spatio-temporal working-memory and short-term object-location tasks use different memory mechanisms. *Acta Psychologica*, *114*, 41–65. doi:10.1016/S0001-6918(03)00049-0