Journal of Experimental Psychology: Human Perception and Performance

Visual Working Memory is More Tolerant Than Visual Long-Term Memory

Mark W. Schurgin and Jonathan I. Flombaum

Online First Publication, May 7, 2018. http://dx.doi.org/10.1037/xhp0000528

American Psychological Association

Journal of Experimental Psychology: Human Perception and Performance

0096-1523/18/$12.00 http://dx.doi.org/10.1037/xhp0000528

Visual Working Memory is More Tolerant Than Visual Long-Term Memory

Mark W. Schurgin
University of California, San Diego

Jonathan I. Flombaum
Johns Hopkins University

Human visual memory is tolerant, meaning that it supports object recognition despite variability across encounters at the image level. Tolerant object recognition remains one capacity in which artificial intelligence trails humans. Typically, tolerance is described as a property of human visual long-term memory (VLTM). In contrast, visual working memory (VWM) is not usually ascribed a role in tolerant recognition, with tests of that system usually demanding discriminatory power—identifying changes, not sameness. There are good reasons to expect that VLTM is more tolerant; functionally, recognition over the long-term must accommodate the fact that objects will not be viewed under identical conditions; and practically, the passive and massive nature of VLTM may impose relatively permissive criteria for thinking that two inputs are the same. But empirically, tolerance has never been compared across working and long-term visual memory. We therefore developed a novel paradigm for equating encoding and test across different memory types. In each experiment trial, participants saw two objects, memory for one tested immediately (VWM) and later for the other (VLTM). VWM performance was better than VLTM and remained robust despite the introduction of image and object variability. In contrast, VLTM performance suffered linearly as more variability was introduced into test stimuli. Additional experiments excluded interference effects as causes for the observed differences. These results suggest the possibility of a previously unidentified role for VWM in the acquisition of tolerant representations for object recognition.

Public Significance Statement

Human visual memory is tolerant, meaning that it supports object recognition despite variability across encounters at the image level. Tolerance is not yet fully understood in neural or computational terms. Typically, tolerance is ascribed to visual long-term memory (VLTM)—recognition taking place over longer durations and relying on passive storage mechanisms. Visual working memory (VWM), in contrast, requires active maintenance, operates only on the order of seconds (perhaps minutes), and is typically thought of as discriminating, noticing differences not tolerating them. But here, comparing VWM and VLTM tolerance directly and for the first time revealed that VWM is more tolerant. These results suggest a previously unidentified role for VWM in the acquisition of tolerant representations for object recognition.

Keywords: long-term memory, working memory, object recognition, tolerance, invariance

Visual object recognition is an exceptionally challenging computational problem whenever pixel-level image analysis is an inadequate solution (Dicarlo, Zoccolan, & Rust, 2012; Logothetis & Sheinberg, 1996). For humans, many organisms, and for general-purpose artificial systems (AI), image variability arises inescapably because of changes in viewpoint, orientation, and lighting across encounters (Cox & DiCarlo, 2008; Cox, Meier, Oertelt, & DiCarlo, 2005; DiCarlo & Cox, 2007; Rust & Stocker, 2010; Wallis & Bülthoff, 2001). Yet humans are remarkably good at recognizing objects despite these and other perturbations, even recognizing objects when they are partially occluded, when they are simplified as line drawings, and when one has only encoun-

Mark W. Schurgin, Department of Psychology, University of California, San Diego; Jonathan I. Flombaum, Department of Psychological and Brain Sciences, Johns Hopkins University.

This research was funded by NSF BCS-1534568 and a seed grant from the Johns Hopkins University Science of Learning Institute to Jonathan I. Flombaum. We thank Jack Schurgin for his invaluable support. Pilot data from Experiments 1a and 1b were published in *Visual Cognition* in 2015 as part of a conference paper (Schurgin & Flombaum, 2015). Some of the results and ideas in this article were also presented at the annual meeting on Object Perception, Attention and Memory (OPAM) in 2016, and at the annual meeting of the Visual Sciences Society (VSS) in 2016.

Correspondence concerning this article should be addressed to Mark W. Schurgin, Department of Psychology, University of California, 9500 Gilman Drive #0109, McGill Hall, La Jolla, CA 92093. E-mail: mschurgin@ucsd.edu

tered other exemplars of a category, not the specific exemplar to be recognized (Biederman, 1987; Rust & Stocker, 2010). Object recognition is a rare case in which humans are still competitive with (and in many cases surpass) AI (Pinto, Cox, & DiCarlo, 2008). The ability to recognize objects despite considerable input variability is usually referred to as tolerance or invariance. The exact computational and neural mechanisms that afford tolerance remain unknown (Andreopoulos & Tsotsos, 2013; Dicarlo, Zoccolan, & Rust, 2012; Logothetis & Sheinberg, 1996; Rust & Stocker, 2010).

Here we compare human memory performance over longer and shorter durations in terms of tolerance. This was done in order to contrast tolerance in what are thought to be distinct memory systems, visual working memory (VWM) and visual long-term memory (VLTM). The system level distinction between visual working and visual long-term memory is among the most entrenched in research on human cognition. Among the motivators of the distinction is that VWM and VLTM are neurally dissociable (Baddeley, 2003; Cabeza, Dolcos, Graham, & Nyberg, 2002; Ryan & Cohen, 2004; Shallice & Warrington, 1970; Speer, Jacoby, & Braver, 2003). Another motivation for distinguishing these two systems is the need for active maintenance in working memory compared with passive maintenance in long-term memory. Persistent neural firing is not an available mechanism over years, months, days, or even minutes, though it appears to be necessary for visual working memory maintenance on the order of seconds (Awh & Jonides, 2001; Awh, Vogel, & Oh, 2006; Baddeley, 2003; Brady, Konkle, & Alvarez, 2011; Cowan, 2008; Harrison & Tong, 2009; Serences, Ester, Vogel, & Awh, 2009).

Tolerance is usually a property ascribed to long-term memory. Indeed, typical visual working memory experiments demand intolerance from observers. They ask for participants, more often than not, to notice small changes and to mark stimuli as changed. This is not an inherent property of any paradigm used, but instead, a property of what expectations researchers have for accurate performance; we often want participants in working memory experiments to judge slight changes as constituting something different. This is not without good motivation. As an active venue for the manipulation of information one might expect that visual working memory prizes precision (Awh, Barton, & Vogel, 2007; Bays, Catalao, & Husain, 2009; Brady, Konkle, Gill, Oliva, & Alvarez, 2013; Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001; Zhang & Luck, 2008). In contrast, the need for passive and massive long-term storage may inherently require more permissive criteria for identifying inputs as “the same” (Rust & Stocker, 2010; Wolfe, 1998; Yonelinas, Aly, Wang, & Koen, 2010). But visual working memory and visual long-term memory have never been compared with respect to tolerance.

Experiment 1: Comparing VWM and VLTM Tolerance

We compared tolerance in VWM and VLTM. Participants were initially exposed to two (or more) real-world objects in a trial. After a brief delay, one of the objects was paired with a new object and the task was to indicate which was in the recently seen set. After completing the visual working memory task, a long-term memory test probed the previously untested objects against foils using the same testing procedure. Thus, both objects in a trial were encoded at the same time, with both as candidates for a short-term or a long-term test. And both objects were tested in exactly the same way, with participants asked to pick out the previously seen object from a pair that included only one that was previously shown (along with a foil).

We used a two-alternative-forced-choice procedure (2AFC) because it affords more straightforward data analysis than, for example, an old/new procedure, and requires recognition mechanisms. From a standard signal detection account, an error should only arise when a previously seen object fails to evoke a memory trace that is stronger than the memory trace evoked by the foil. On the aggregate therefore, error rates in the task reflect the effectiveness of recognition (Bayley, Wixted, Hopkins, & Squire, 2008; Smith & Duncan, 2004).

In Experiment 1a we sought to establish task performance levels when no image perturbations were present. Then, in Experiment 1b, we directly assessed the relative tolerances of these two systems by injecting noise (via randomly scrambling pixels) into the images at test. We reasoned that a more tolerant system would afford better performance as noise at test increased.

Method

Participants. A group of 18 Johns Hopkins University undergraduates participated in Experiment 1a and a separate group of 20 in Experiment 1b. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

In all experiments except for Experiment 3, we sought to test between 18 and 25 participants, applying the lab’s standard stopping rule wherein we test all participants who have signed up on the day that we cross a preset threshold, set at 18 for these experiments. This rule is applied for logistical purposes, allowing undergraduate volunteers to complete experiments for course related credit. In both Experiments 1a and 1b the results from one participant was excluded due to noncompliance with the instructions.

Power analysis. We sought this number of participants because a recently published study of ours confirmed reliable effects using similar sample sizes, including incidental encoding of the same image sets and test conditions with noise and orientation changes (Schurgin & Flombaum, 2017, Experiment 1b). Specifically, in that study we compared overall memory performance among 20 subjects for recognition of test images shown in low or high noise of the same variety of noise used here. We observed a within-subject Cohen’s dz of 1.51. A power analysis with those results demonstrated that nine participants would be sufficient, with a power of 0.95 and a 0.05 significance level. While a sample size of nine would be sufficient we decided to adhere to a larger sample, more consistent with those typical in the lab (between 18 and 25 in visual memory experiments).

Apparatus. Experiment 1 took place in a dimly lit sound-attenuated room. Stimuli were presented on a Macintosh iMac computer with a refresh rate of 60 Hz. The viewing distance was 60 cm so that the display subtended $39.43° \times 24.76°$ of visual angle.

Stimuli and procedure. Stimuli were generated using MATLAB and the Psychophysics toolbox (Brainard, 1997; Pelli, 1997).

All stimuli were presented within the full display frame of 39.43° × 24.76°. In Experiment 1a, participants first completed a VWM test. In each trial of the experiment, participants briefly saw two real-world objects in one of four possible locations. Their task was to maintain the two pictures in VWM. Additionally, there were set size one trials, to provide a baseline performance for VWM. Participants would press the space key to initiate each trial. They would then see a fixation cross and four boxes (left, right, above, below) where potential stimuli could appear for 1,000 ms, followed with the addition of two images of real-world objects appearing in two of the possible locations for 500 ms. Afterward a mask was presented for 900 ms. Masks were created by randomly selecting approximately 20 images from the stimulus set which were then pixelated using a mosaic filter in Photoshop in order to remove identity information of the objects while maintain other visual information (i.e., color distributions). A total of eight masks were created and randomly used during each trial.

At test, participants faced a two-alternative forced (2AFC) judgment involving a randomly selected object from the encoding display and a completely new categorically distinct object, with the task of identifying the old object (see Figure 1). New images were never repeated across any trial and were always distinct. Participants completed 180 trials in total. Our primary motivation for this many trials was to allow us to evolve methods for subsequent experiments with more individual conditions, while maintaining an adequate number of trials per condition bin. After completing the VWM task participants faced a surprise VLTM test. Participants initiated each trial again by pressing the space key, and they were then immediately presented the test images. On each trial, the previously untested object from each of the encoding displays was

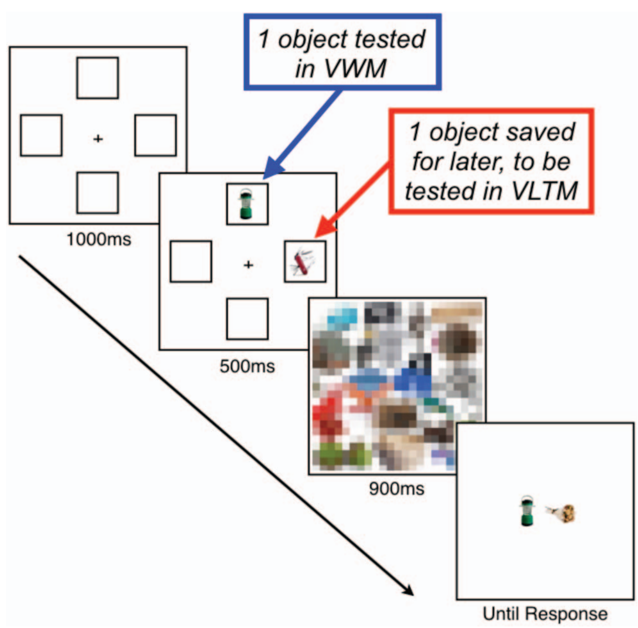

Figure 1. Illustration of the general experimental procedure. Two objects appeared briefly in one of four locations. One object was tested immediately in VWM using a 2AFC procedure. The other object in the display was tested later using the same alternative forced choice procedure. See the online article for the color version of this figure.

paired with a new object, and participants reported the one that was "old," that is, the one that had appeared at some point in the encoding phase. Their task was exactly the same as before—to indicate which of the two images they had seen in the past.

Experiment 1b was identical to Experiment 1a, with the following exceptions: During the VWM task, all test objects were embedded in 75% noise by randomly scrambling 75% of the pixels in an image. In the VLTM task, all test objects were embedded in either 25%, 50%, or 75% noise.

Data analysis. For the VLTM task, we did not include responses that were faster than 200 ms, as these responses were likely the result of random error. We did not exclude these responses for the VWM task, due to the delay provided by the mask before a response could be made. This accounted for a total of 1.81% of trials in Experiment 1a and 2.13% of VLTM trials in Experiment 1b.

Results

In Experiment 1a we established performance expectations, testing only noiseless images. We included a set size one condition in the VWM task to identify performance in the most basic condition. At Set Size 1, mean VWM accuracy was 99.56% (SD = 0.7%), which was significantly better than VWM accuracy at Set Size 2 (M = 97.5%, SD = 1.7%), $t(17) = 5.08$, $p < .001$, Cohen's dz = 1.21.

Next, we compared performance for images presented both in VWM (i.e., SS2) and VLTM. We observed that VLTM accuracy was significantly worse compared to VWM performance, with participants averaging 81.7% correct (SD = 6.3%), $t(17) = 9.63$, $p < .001$, Cohen's dz = 2.54. This level of performance is somewhat lower than what has been observed in some experiments that have looked at the massive capacity of long-term memory (where performance is sometimes as high as 90% as opposed to 80%). This is not surprising given the specific constraints of the present experiments, which included simultaneous exposure to two items, and for only half a second (as opposed to 2 s or 3 s). Moreover, in some massive memory experiments, 80% has been the norm (Guerin, Robbins, Gilmore, & Schacter, 2012; Vogt & Magnussen, 2007). Thus, these results suggest relatively normal long-term memory for unperturbed images in our stimulus set. Our question of interest concerned how the introduction of perturbations at test might affect this performance. Therefore, we turn to Experiment 1b.

In Experiment 1b, we discovered that VWM is robustly tolerant to image-noise introduced at test, while VLTM performance declines as more noise is added (see Figure 2). To introduce image-level noise at test, we randomly scrambled a percentage of the pixels (25%–75%) in the previously shown test image and the foil in each test pair. We expected that less scrambled images would be recognized more easily, allowing us to assess the relative tolerance of VWM and VLTM.

To evaluate whether noise affected VWM and VLTM, we conducted a repeated-measures ANOVA, with memory type as a within-subject factor and noise (0% and 75%) as a between-subjects factor. We observed a main effect of memory type, $F(1, 34) = 405.63$, $p < .001$, $\eta^2_G = 0.92$, reflecting that VWM performance was significantly better than VLTM performance. Critically, we also observed a significant interaction between memory

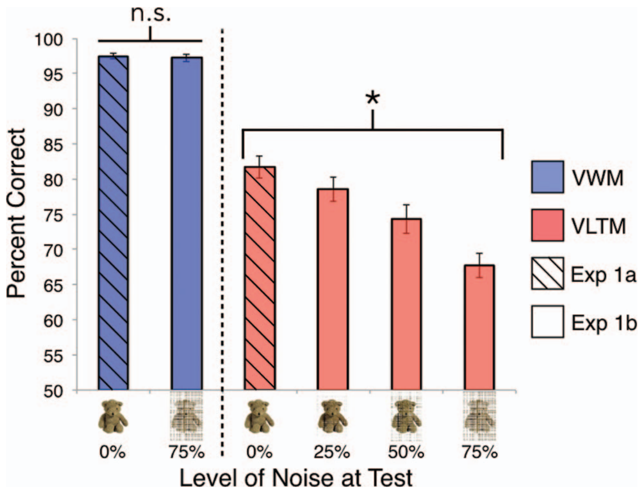

Figure 2. Results of Experiment 1a and 1b. The 0% noise conditions in the figure were obtained in Experiment 1a and all other noise condition results came from Experiment 1b (i.e., a different group of participants). Across both experiments, memory performance was unaffected by noise level at test in VWM, with comparable performance across 0%–75% noise. In contrast, VLTM was greatly affected by noise at test, with a linear decrease in performance observed across 0%–75% noise. * $p < .01$. Error bars represent SEM. See the online article for the color version of this figure.

type and noise, $F(1, 34) = 36.84$, $p < .001$, $\eta^2_G = 0.52$, suggesting there was a difference between how noise affected VWM and VLTM.

Further analyses revealed VWM performance was unaffected by noise at test, even when images were embedded in 75% noise. (We only tested at 75% noise for VWM in this experiment because of pilot results, not reported, that showed no effect on performance at 25 or 50% levels.) Participants averaged 97.2% correct (SD = 2.24%), which was not significantly different than the performance observed in Experiment 1a (in the two objects condition), $t(34) = 0.42$, $p = .68$, Cohen's $d = 0.14$.

In contrast, VLTM performance was significantly affected by noise, $F(2, 36) = 20.28$, $p < .001$, $\eta^2_G = 0.53$. As noise increased there was a clear drop in performance, with 78.5% correct (SD = 7.6%) at 25% noise, 74.3% correct (SD = 8.8%) at 50% noise, and 67.7% (SD = 7.6%) at 75% noise.

Experiment 2: Greater VWM Tolerance to Orientation Change

In Experiment 1, we contrasted how tolerant VWM and VLTM were when random noise was injected at test. We assumed a tolerant system should be able to overcome the variability introduced by our random noise manipulation. However, noise is only one way to introduce variability into images. Orientation changes are a common type of variability that object recognition mechanisms must face (DiCarlo & Cox, 2007; Pinto et al., 2008; Rust & Stocker, 2010). We therefore sought to compare VWM and VLTM tolerance with orientation changes. Experiment 2 was identical to Experiment 1, with the following exceptions: each presentation of an object during encoding was shown from a specific orientation. At test, "old" objects were shown either at the original orientation ("same" relative to encoding) or a second, never-before-seen orientation ("different"). The task was still to pick which object in a pair had been shown before, such that it was correct to identify a rotated object as the one seen before if it had been shown at a different orientation.

Method

Participants. A new group of 24 Johns Hopkins University undergraduates participated in Experiment 2. The results from two participants were excluded due to noncompliance with the instructions. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Apparatus, stimuli, and procedure. Experiment 2 was identical to Experiment 1a, with the following exceptions: During the VWM task, participants were shown color images of objects from a specific orientation (taken from Geusebroek, Burghouts, & Smeulders, 2005). At test, participants made a 2AFC judgment, but the old object was shown from the same or a different orientation. After 79 trials of this task, participants faced a surprise VLTM test. Again, at test participants made a 2AFC judgment, but the old object was shown either from the same or a different orientation.

Data analysis. For the VLTM task we did not include responses that were faster than 200 ms, which accounted for a total of 0.12% of trials in Experiment 2.

Results

We first conducted a repeated measures ANOVA with memory type and orientation change as factors. We observed a main effect of memory type, $F(1, 21) = 391.23$, $p < .001$, $\eta^2_G = 0.95$, reflecting that VWM performance was significantly better than VLTM performance. We also observed a main effect of orientation change, $F(1, 21) = 7.79$, $p = .01$, $\eta^2_G = 0.27$, suggesting that there was an effect on performance relative to orientation test. Finally, we observed a significant interaction between memory type and orientation, $F(1, 21) = 4.45$, $p < .05$, $\eta^2_G = 0.18$. To identify the source of the interaction, we conducted follow-up analyses individually looking at memory type and orientation.

VWM performance was not significantly affected by orientation changes. Same orientation test performance was 95.66% on average (SD = 4.25%), and different orientation test performance was 94.76% (SD = 4.15), $t(21) = 0.68$, $p = .50$, Cohen's dz = 0.16. VLTM performance at the same orientation averaged 67.03% correct (SD = 9.07%), whereas recognition at a different orientation was significantly lower, averaging 61.39% (SD = 9.69%), $t(21) = 3.03$, $p < .01$, Cohen's dz = 0.65 (see Figure 3).

Experiment 3: VLTM Performance Not Caused by Interference

Interference is a well-known cause of difficulty in long-term memory retrieval. For example, in verbal long-term memory recall performance for a single syllable does not degrade over time, even if other types of memory contents need to be stored. But if multiple

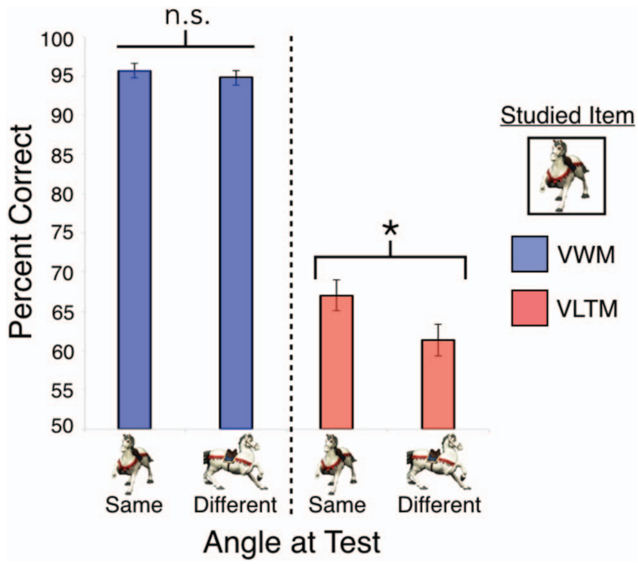

Figure 3. Results of Experiment 2. VWM performance was equivalent across recognizing a previously seen object at the same or a different orientation. In contrast, VLTM performance was significantly worse when recognizing a previously seen object at a different orientation than the one it originally appeared in. * $p < .01$. Error bars represent SEM. See the online article for the color version of this figure.

syllables need to be remembered performance decreases as the retention interval increases, suggesting a role for mutual interference between memoranda in long-term memory (Keppel & Underwood, 1962). A concern in Experiments 1 and 2 therefore is that long-term relative intolerance (compared with visual working memory) may be caused by interitem interference because a set of items larger than two must be stored before the long-term memory test takes place.

We addressed this concern in Experiment 3, a one trial version of Experiment 1a. Participants did only one trial of the visual working memory task, in which they were shown two objects and then tested immediately for memory of one of them. The other item was saved while the participant completed an unrelated study (one of a number of the other experiments in the lab that did not include images of real-world objects). Finally, the participant was asked to identify the second previously shown object from a pair of images that included a foil, that is, the participant was given a surprise VLTM test probing the saved image from the previous VWM trial. In this way, the VLTM test took place without the potential for interference among a large set of exposure images.

Method

Participants. A new group of 60 Johns Hopkins University undergraduates participated in Experiment 3. We sought to collect three times as many subjects as normal to assure we had sufficient power, given that we only collected a single trial for each memory condition (we had no previous basis, as this was to our knowledge the first one-trial experiment conducted in visual episodic long-term memory). All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Apparatus, stimuli, and procedure. Experiment 3 was identical to Experiment 1a, with the following exceptions: Participants received only one VWM trial. After completing this trial, they would then complete an unrelated study, which did not involve any stimuli containing real-world images of objects. After completing this study (approximately 40 min), they were then given one trial of the surprise VLTM test. Objects encoded and tested were drawn randomly from 378 possible images, and the "new" images in the 2AFC test were drawn randomly from a separate catalogue of 378 images.

Results

We tested 60 participants in this experiment, and astonishingly performance across both the VWM and VLTM tests was similar to the baselines established in Experiment 1a, with participants averaging 98.33% correct (SEM = 5.5%) and 76.67% correct (SEM = 1.7%), respectively, $t(59) = 3.69$, $p < .01$, Cohen's dz = 0.50 (see Figure 4). Thus, average performance differences in Experiment 1a were not the result of interference between memoranda. This is the only single trial experiment on VLTM that we are aware of. It provides evidence that VLTM intolerance is not caused exclusively by interference, and also reinforces the motivation to distinguish between VLTM and VWM.

Experiment 4: VWM Robustly Tolerant Despite Pressured Capacity

In Experiments 4a and 4b we sought to investigate tolerance in VWM when memory capacity is pressured. Capacity limitation is among the defining characteristics of VWM, with apparent constraints on the number of items a person can maintain in memory simultaneously (Awh et al., 2007; Vogel et al., 2001) and on the resolution of representations when more items are maintained, as opposed to fewer (Ma, Husain, & Bays, 2014). Experiments 1–3 make it clear that VWM has a high-degree of tolerance when only two items are maintained in memory. But what happens when memory capacity is stretched? Perhaps the effort to allocate resources or to manage competition between representations comes with a cost in terms of tolerance.

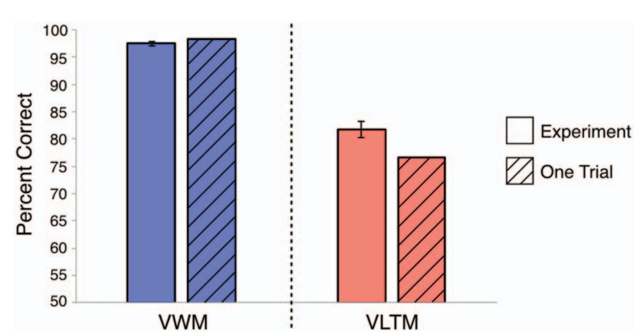

Figure 4. Results of Experiment 3, comparing one-trial performance to the baseline observed in Experiment 1a. Even with a single trial, VLTM performance was significantly worse than VWM. See the online article for the color version of this figure.

To investigate this possibility, we conducted an experiment like the ones previous, but involving only a VWM component (i.e., without a long-term test session). Participants saw trials that included images of four objects (i.e., a memory load of four). Previous research has revealed vastly worse memory performance when four items are maintained in VWM as opposed to two. Participants were instructed to remember all items in each trial, and then after the items disappeared, participants faced an alternative forced choice test in which they indicated which item in a pair was present in the previous set of four. In Experiment 4a, the test images were embedded in either 0% or 75% noise.

Experiment 4b extended this type of manipulation to an orientation change. In this experiment, we also varied memory load within subject; each trial included either two or four images. To test memory tolerance, the 2AFC test that followed each trial included one of the previous objects at either the same or a different orientation relative to encoding, along with a previously unseen foil (see Experiment 2).

Method

Participants. A new group of 20 Johns Hopkins University undergraduates participated in Experiment 4a and 21 in Experiment 4b. The results from one participant was excluded in Experiment 4b due to noncompliance with the instructions. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Apparatus, stimuli, and procedure. Experiment 4a was identical to Experiment 1b, with the following exceptions: only the VWM task was administered. All test stimuli were embedded in either 0% or 75% noise. Experiment 4b was identical to Experiment 2, with the following exceptions: Only the VWM task was administered. In the encoding array, participants were shown either two or four images of real-world objects.

Results

If capacity limits interact with tolerance, then in Experiment 4a we would expect to find worse performance in the higher noise condition. However, we did not observe a significant difference; average 0% noise performance was 83.3% (*SD* = 4.96%), which was statistically indistinguishable from a mean of 83.6% (*SD* = 6.21%) in the 75% noise condition, $t(19) = 0.17$, $p = .87$, Cohen's dz = 0.04. (Performance in this experiment was significantly worse than in Experiment 1a, $M = 83.5\%$ vs. $M = 97.5\%$, respectively, $t(35) = 12.15$, $p < .001$, Cohen's $d = 4.14$. This difference demonstrates that capacity was challenged by the memory load of four compared with two).

In Experiment 4b, we observed a significant performance deficit at a larger set size compared to trials with a load of two, $M = 79.03\%$ versus 96.25%, respectively, $F(1, 19) = 58.15$, $p < .001$, $\eta^2_G = 0.75$, confirming that capacity was stressed in the memory load of four condition. However, we did not find a main effect for orientation, $F(1, 19) = 0.50$, $p = .49$, $\eta^2_G = 0.03$, or an interaction between load and orientation, $F(1, 19) = 0.34$, $p = .56$, $\eta^2_G = 0.02$ (see Figure 5). At a load of two this lack of an orientation effect replicates the lack of an effect observed in Experiment 2. Further, the absence of an effect of orientation at a load of four suggests that limited capacity in visual working memory does not interact with tolerance. Tolerance appears to be an affordance of visual working memory such that limitations on capacity can reduce performance overall—perhaps by leading to objects that are not represented, or by reducing representational precision for individual visual features (e.g., color; see Brady et al., 2013). But capacity limits appear not to constrain the tolerance afforded by a visual working memory representation, any load considerations equal.

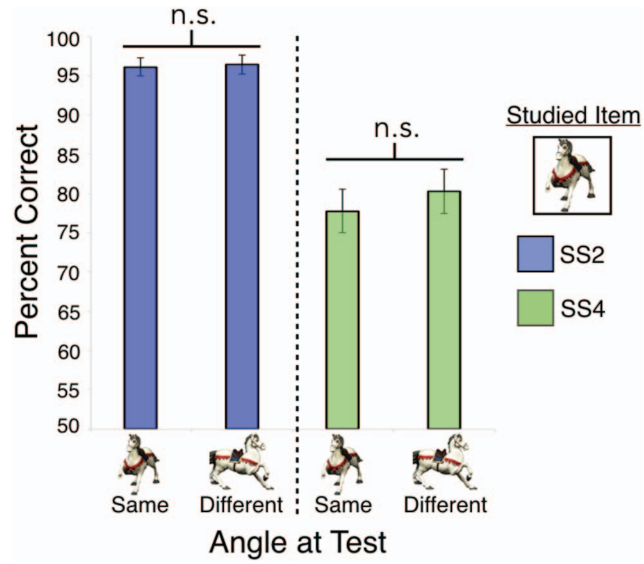

Figure 5. Results of Experiment 4b. There was no difference in VWM performance for recognizing a previously seen object at the same or a different orientation, regardless of set size. However, performance was significantly worse at SS4 relative to SS2. Error bars represent SEM. See the online article for the color version of this figure.

Experiment 5: Controlling for Encoding Time

As detailed below, we conducted three additional experiments in order to address potential alternative accounts or confounding factors in the primary experiments. Experiment 5, was designed to investigate potential effects of encoding time.

It may be that the relative intolerance demonstrated by VLTM reflects inadequate encoding time. To investigate the possibility, we replicated Experiment 2 increasing encoding time from 0.5 s to 1 s.

Method

Participants. A new group of 22 Johns Hopkins University undergraduates participated in Experiment 5. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Apparatus, stimuli, and procedure. Experiment 5 was identical to Experiment 2, with the following exceptions: encoding time for the VWM array was increased from 0.5 s to 1 s.

Data analysis. For the VLTM task we did not include responses that were faster than 200 ms, which accounted for a total of 1.09% of trials in Experiment 5.

Results

As in Experiment 2, we first conducted a repeated measures ANOVA. We observed a main effect of memory type, $F(1, 21) = 234.59$, $p < .001$, $\eta^2_G = 0.92$, reflecting that VWM performance was significantly better than VLTM performance. We also observed a main effect of orientation change, $F(1, 21) = 10.46$, $p < .01$, $\eta^2_G = 0.33$, suggesting that there was an effect on performance caused by orientation at test. Finally, we observed a significant interaction between memory type and orientation, $F(1, 21) = 8.15$, $p < .01$, $\eta^2_G = 0.28$. To identify the source of the interaction, we conducted follow-up analyses individually looking at memory type and orientation.

Again, in VWM we did not observe any significant difference for recognizing a previously seen object at the same ($M = 96.70\%$, $SD = 2.93\%$) or a different orientation ($M = 96.15$, $SD = 4.17\%$), $t(21) = 0.63$, $p = .54$, Cohen's dz = 0.13. But VLTM performance at the same orientation averaged 71.89% correct ($SD = 9.49\%$), whereas recognition at a different orientation was significantly lower, averaging 65.94% correct ($SD = 9.54\%$), $t(21) = 3.42$, $p < .01$, Cohen's dz = 0.73 (see Figure 6).

Experiment 6: Controlling for the Role of Color Information

We have emphasized the encoding of objects into visual working and long-term memory. It is possible though that discrimination responses in the tasks rely on lower-level display features. Color, in particular, could be available as a means to discriminate choices in the immediate VWM test, but not the VLTM test. Of particular concern is that color could be a diagnostic cue as to which item was not in a previous VWM array, as opposed to which item was, meaning that performance in the task would be more about rejecting an item known not to have been present, as opposed to recognizing one that was. We therefore replicated Experiment 1b but with completely grayscale images.

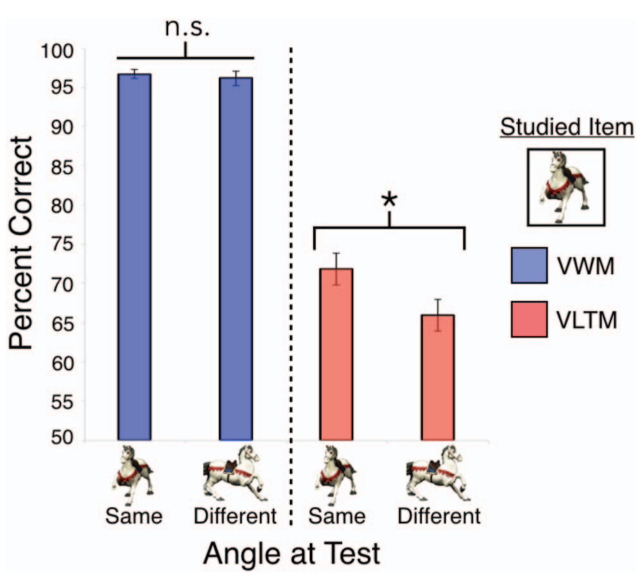

Figure 6. Results of Experiment 5 (Encoding Time Control). Overall, with the increased encoding time performance was higher for both VWM and VLTM relative to Experiment 2. VWM was robustly tolerant to recognizing a previously seen object at a new orientation. However, even with increased encoding time (relative to Experiment 2) VLTM performance suffered. * $p < .01$. Error bars represent SEM. See the online article for the color version of this figure.

Method

Participants. A new group of 21 Johns Hopkins University undergraduates participated in Experiment 6. The results from two participants were excluded due to noncompliance with the instructions. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Apparatus, stimuli, and procedure. Experiment 6 was identical to Experiment 1b, with the following exceptions: all stimuli were converted to grayscale images.

Data analysis. For the VLTM task we did not include responses that were faster than 200 ms, which accounted for a total of 1.11% of trials in Experiment 6.

Results

As in Experiment 1, we first conducted a repeated-measures ANOVA with memory type as a within-subject factor and noise (0% and 75%) as a between-subjects factor. We observed a main effect of memory type, $F(1, 34) = 433.73$, $p < .001$, $\eta^2_G = 0.93$, reflecting that VWM performance was significantly better than VLTM performance. We also observed a significant interaction between memory type and noise, $F(1, 34) = 42.85$, $p < .001$, $\eta^2_G = 0.56$, suggesting that there was a difference between whether noise affected VWM and VLTM.

Despite all stimuli being presented without color information, performance was remarkably similar to Experiment 1b (see Figure 7). Even when given 75% noise at test, participants averaged 95.41% correct ($SD = 3.56\%$) in VWM. This was not significantly different than what was observed in Experiment 1b (97.2%), $t(36) = 1.87$, $p = .07$, Cohen's $d = 0.61$.

VLTM performance also patterned as it did in Experiment 1, with a linear decrease in performance observed as noise increased from 25%–75%. We observed no significant differences comparing individual noise levels with Experiment 1 (all p's > 0.26).

Experiment 7: Controlling for Test-Induced Forgetting

The general instructions for the experiments so far presented the VLTM test as a surprise, with the participants at first only told that they would be tested following each trial. Perhaps testing one item in a display immediately led to the strategic forgetting of the other. To evaluate this possibility, we modified the basic paradigm so observers were told that they would be tested for long-term memory after the VWM portion of the experiment. Additionally, in half of trials (unpredictably sorted) there was no immediate VWM test after the encoding display. Thus, items were available for VLTM testing that were not paired with an already-tested image.

Method

Participants. A new group of 20 Johns Hopkins University undergraduates participated in Experiment 7. All participants re-

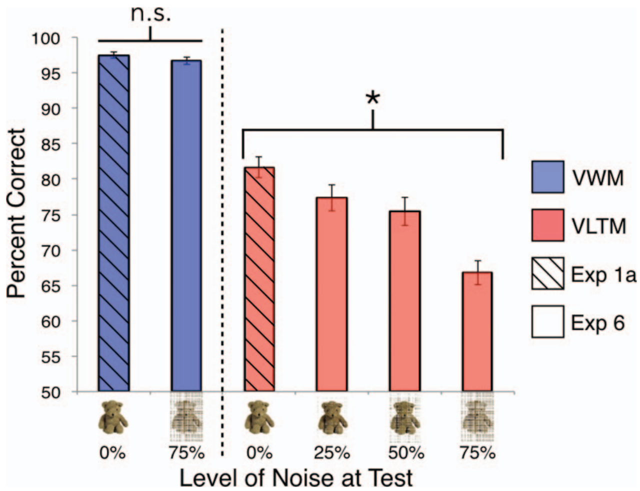

Figure 7. Results of Experiment 1a and 6 (Color Information Control). The 0% noise conditions in the figure were obtained in Experiment 1a and all other noise condition results came from Experiment 6 (i.e., a different group of participants). Consistent with Experiment 1b, Experiment 6 replicated with grayscale stimuli that VWM was robustly tolerant to noise at test whereas VLTM performance suffered. * $p < .01$. Error bars represent SEM. See the online article for the color version of this figure.

ported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Apparatus, stimuli, and procedure. Experiment 7 was identical to Experiment 1a, with the following exceptions: All trials during the first phase of the experiment contained two objects in the array. Participants were explicitly told they would be tested on all the items in the VWM array during a later test of VLTM. In 50% of trials they received a test of one of the objects directly after the array. In the other 50% of trials there was no immediate test of VWM. Trial types were assorted randomly.

Data analysis. For the VLTM task we did not include responses that were faster than 200 ms, which accounted for a total of 2.38% of trials in Experiment 7.

Results

We did not find a significant difference in VLTM performance for items that appeared in immediately tested compared to untested displays, respectively, $M = 67.17\%$ ($SD = 6.29\%$) compared with $M = 70.0\%$, ($SD = 8.94\%$), $t(19) = 1.70$, $p = .11$, Cohen's dz = 0.38 (see Figure 8). This suggests that the lower performance in VLTM was not the result of participants disregarding information about objects after being tested in VWM, incidental encoding, or the surprise nature of the VLTM test.

Discussion

We sought to compare VWM and VLTM performance in terms of tolerance. After a brief duration and while actively keeping memoranda in mind—that is, when relying on VWM—participants easily recognized objects that they had just seen, and they did so without accruing costs if the test images where noisy or showed an object at a new orientation (compared with the studied image of the object). Stressing VWM capacity by including four objects in each trial reduced performance overall, but it did not interact with the presence or absence of image-level perturbation at test (Experiment 4); at a given memory load, participants recognized objects equally well when they were shown exactly as presented during study, when random noise was introduced to the test image via pixel scrambling, and when the objects were tested at new orientations.

In contrast to VWM, object recognition after a delay of 30 min–45 min was not only poorer than during immediate test, but it suffered from the introduction of image-level perturbation. In marked contrast to VWM-based recognition, VLTM-dependent recognition declined as more pixels were scrambled in test compared with study images, and performance was also worse when objects were tested at orientations different from the ones originally studied. Worse performance over the long-term compared with immediate testing was unlikely to have been caused by interference between memoranda because a one trial experiment produced the same relative long- and short-term performance differences. Tolerance refers to the ability to recognize objects despite viewing conditions that are different from past encounters. The results obtained here show visual working memory to be highly tolerant, and more so than long-term memory. This has several important implications with respect to visual object recognition and visual episodic memory.

A System-Level Distinction

An important implication to note first is that the results reinforce the system-level distinction between visual working and visual long-term memory. That is, working memory performance was not only better than long-term memory performance in these experiments, it was also qualitatively different in being unaffected by noise or orientation changes, even as memory load increased. This

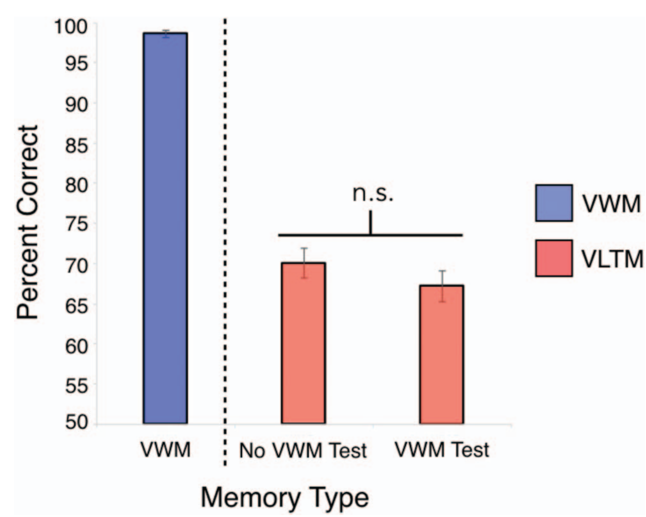

Figure 8. Results of Experiment 7 (Test Interference Control). We observed no significant difference in VLTM performance for items that appeared in immediately tested displays in VWM compared with those that were untested. This was despite explicit instructions during encoding that all items would be tested subsequently in VLTM. Error bars represent SEM. See the online article for the color version of this figure.

suggests that the two systems could traffic in different representations and algorithms. In his well-known book on how to investigate human vision, Marr (1982) explains that representations differ in what they make explicit, in less technical terms, "how information is represented can greatly affect how easy it is to do different things with it" (p.244) A classic example (also courtesy of Marr, 1982) is the difficulty of multiplication with Roman numeral representations of number. Here we have observed a difference in what a person *can do* with VWM and VLTM representations, in particular whether or not she will be able to recognize a perturbed image as the same as an image encountered just once previously. Thus, the systems seem to possess different affordances, either because they utilize different representations or because they use those representations in different ways (with different procedures and algorithms). A great deal more research is necessary, and hopefully imminent, to explain exactly how differences in tolerance arise.

Recognition Versus Novelty

One way that VWM and VLTM could *use* their representations differently is with respect to how they rely on recognition as opposed to novelty in order to determine whether something was seen already. In the experiments reported, the difference between recognition and novelty would be the difference between accurately choosing the seen object in a test pair because it is affirmatively recognized as being an object shown (recognition) or because the foil is notably an object that was not shown (novelty). The difference is meaningful outside this specific paradigm; in general, do recognizing and rejecting depend on the same mechanisms and representations, or are they substantively different, with the possibility that VWM and VLTM rely on them differentially?

This is a question that has interested researchers in VLTM for some time, and to some degree, researchers looking at VWM as well. In the case of VWM, the question was framed by Hyun, Woodman, Vogel, Hollingworth, and Luck (2009) with respect to the common change detection paradigm: Do subjects record a change response when they believe that something that was previously seen is now absent (a recognition-dependent mechanism), when they believe that something that was not previously seen is present (a novelty-dependent mechanism), or both? But the results could not distinguish between the two, with some evidence available in favor of each. More recently, computational models have applied optimal Bayesian decision-making to the task of change (and sameness) detection, a framework wherein absence and presence decisions are always made together, and a given task response is made if and only if its probability is greater than the computed probability of the alternative response (Keshvari, van den Berg, & Ma, 2013; van den Berg et al., 2012). This point-of-view is consistent with the consensus view in long-term memory research, where dual-process models have been replaced by single process memory strength models, for which an error should only arise when a previously seen object fails to evoke a memory trace that is stronger than the memory trace evoked by the foil (Bayley et al., 2008; Smith & Duncan, 2004).

Returning to the experiments reported here, then, is it possible that participants will choose a target over a foil only because one fails to recognize the foil, as opposed to affirmatively recognizing the target? The prevailing perspective from both the literature on VWM and VLTM is that the distinction is not actually substantive because both types of decisions rely on the same representational content, which is compared with both the target and the foil. With respect to the differential tolerances of the systems, the prevailing decision-making models demand differentially tolerant representations to produce the results that we observed.

But because the issue is so highly relevant to tolerance specifically, it is perhaps worth further entertaining the possibility that our participants relied on novelty detection, not affirmative recognition, when they chose targets over foils in our VWM experiments. Could it be that the tolerance observed in VWM was caused by novelty detection, excluding foils by detecting them as different, and not by affirmatively choosing targets because they were detected as the same (as what was shown) despite perturbation?

The possibility seems inconsistent with what we observed. A novelty detection mechanism should ultimately be an intolerant one; but we observed tolerance in VWM. As noise was introduced to test images, and actual signal decreased, the novelty signal generated by the target and the foil should become more similar. Forget for a moment that there was a sample image, and consider the foil and the target merely as paired images. How easily can they be discriminated *from each other?* Presumably, less easily when they each contain more noise. Thus, the novelty signal that the same two images produce should be less distinguishable from one each other, and performance should degrade as noise is added. That is, if novelty is an independent and sole basis for responses in the task. In order for noise to have not affected performance, some degree of signal matching is necessary, especially as memory load was increased.

Consider the case of a load of four in VWM. Specifying a hypothetical novelty detection mechanism might be useful to consider here in order to further see why added noise should produce intolerance. So suppose that instead of trying to recognize a specific match between one of the test images and the memoranda, a participant just compares each of the test images to some summary statistics of the memoranda display as a whole, asking which test image is more novel in this context. With more noise in both test images, the target should appear out of place more frequently than it does with less noise. But noise did not impact performance at a load of four. This suggests that some affirmative matching took place.

Finally, on this point, we do want to acknowledge that we do not view the issue as concluded. Instead, we are comfortable with the possibility that decisions derived from VWM representations rely extensively on novelty detection, even the possibility that the apparent tolerance of VWM arises from this tendency: That somehow, asking whether a stimulus is novel promotes tolerance, albeit a more operational or process driven tolerance than a representational one. The fact remains that tolerance in VWM has been scantly documented or compared with VLTM, and so even if only operationally, the results reported suggest something different for VWM tolerance than would have been predicted. In general, it will be useful to investigate both VWM and VLTM with a range of test procedures, given some memoranda, in order to better understand the representations and procedures that the systems rely on.

Features and Objects

Another important distinction that may be useful for characterizing differences between VWM and VLTM may be in how they handle memory of an object's low-level features compared with its arrangement and identity. It is unusual to classify tolerant performance as "good" when it comes to VWM, and also unusual to test it with images of real-world objects, as opposed to low-level features such as hue or angle. In contrast, and with only one notable exception (Brady et al., 2013), testing feature memory in VLTM is never done, with whole objects always the focus of attention. Quite a lot of work needs to be done, therefore, in order to relate extant research on the two systems, and to understand how feature memories and object memories interact. An interesting recent result is that VWM capacity for real-world objects is higher than for abstract features in a way that seems dependent on active maintenance mechanisms (Brady, Störmer, & Alvarez, 2016). In other words, VWM does better when it is asked to store the features of real-world objects than when just asked to store features, and not because it can exploit long-term knowledge of those real-world objects. Minimally this result and the ones that we have produced here suggest that VWM is optimized for processing objects, and asking visual working memory to remember individual abstract features amounts to feeding it pieces of things that it expects to be parts of more complex wholes.

Object Recognition

More broadly these results suggest that VWM could be part of the set of systems involved in visual object recognition. Such a role is not among the functions typically ascribed to VWM. So what might the roles of VWM and VLTM be with respect to object recognition? Before making a suggestion, we emphasize that the results should not be taken to suggest that VLTM is not tolerant. It is tolerant, in these experiments in practice—because recognition performance was always above chance, even in the most difficult conditions—and as we know from many other experiments (e.g., Cox & DiCarlo, 2008; Cox et al., 2005; DiCarlo & Cox, 2007; Rust & Stocker, 2010; Yonelinas et al., 2010). What these experiments show is *less* tolerance than VWM, at least after only a single and brief encounter with an image. Tolerance in VLTM may increase when more than one exposure is allowed (see Schurgin & Flombaum, 2017).

In retrospect, we think this makes sense because in addition to the challenge of tolerance, VLTM faces a second challenge, one that VWM can concern itself with less. Specifically, VLTM also needs to be explicit, meaning that it needs to discern when similar appearing objects are actually not the same (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). A spoiled food may look very similar to an unspoiled one, my bike may look very similar to your bike, and a poisonous species of snake may look very similar to a nonpoisonous species—and in all these cases differences matter despite physical similarity. The pressure to be explicit must place constraints on how tolerant VLTM can be.

In contrast, VWM can use spatiotemporal cues and token representations to know with relative certainty when an object is the same and when it is not, regardless of physical appearance (Flombaum & Scholl, 2006; Kahneman, Treisman, & Gibbs, 1992; Schurgin & Flombaum, 2017; Schurgin, Reagh, Yassa, & Flombaum, 2013; Yi et al., 2008). We and others have shown that visual working memory employs featureless, pointer-like representations (cf. Chun & Cavanagh, 1997; Kahneman et al., 1992; Pylyshyn & Storm, 1988), sometimes called "object-files," representations which track objects as the same through time, space, and motion, regularly updating the physical features that should be associated with a given pointer (Chun & Cavanagh, 1997; Kahneman et al., 1992; Odic, Roth, & Flombaum, 2012; Pylyshyn & Storm, 1988). This is a reliable strategy so long as an object can be tracked continuously through time and space. Therefore, the strategy is unavailable to VLTM, while critical for VWM.

As a consequence, VWM can demonstrate greater tolerance to variability arising from the same input, as it can rely on spatiotemporal cues and token representations to facilitate object tracking and identity. Given this circumstance, a potential role available to VWM is to serve as a venue for the integration of variability during encounters with objects over the short-term. In doing so VWM can furnish VLTM memory with information about what kinds of variability to tolerate from objects later, and what kind of variability not to tolerate. Tracking over short duration supplies VWM with independent—and reliable—awareness about when inputs at different moments in time are likely to have come from the same objects, regardless of any differences in physical properties. This should allow VWM to be permissive with respect to physical variability, where VLTM needs to be explicit.

Recent work is consistent with the hypothesis that visual working memory is a site for the acquisition of appropriately tolerant long-term memories. We have recently shown that objects seen multiple times, and moving along spatiotemporally continuous motion trajectories are better remembered at a later test and with greater tolerance to orientation and noise manipulations, that is, compared with objects encountered an equal number of times but along spatiotemporally discontinuous motion paths that violate expectations about object permanence (Schurgin & Flombaum, 2017). Similarly, Cox and colleagues have shown that if a person or a monkey makes a saccade to an object, they will integrate the image that they find at the landing place of the saccade together with the image they saw at that location before they executed the saccade. In other words, they expect the object of and the target of a saccade to be the same—a reasonable expectation—regardless of physical appearance. In these experiments, surreptitiously changing the image at the target location causes the acquisition of mongrel memories that integrate physical features that should not have been associated (Cox & DiCarlo, 2008; Cox et al., 2005; see also Isik, Leibo, & Poggio, 2012; Poth, Herwig, & Schneider, 2015; Poth & Schneider, 2016).

But overall, research rarely considers VWM as part of a broader object recognition system. Beyond the descriptive conclusion that at least under some conditions visual working memory is more tolerant than visual long-term memory we therefore propose a novel hypothesis: Visual working memory is a tolerant site for the integration of physical variability so that appropriately tolerant and explicit representations can be encoded into visual long-term memory.

References

Andreopoulos, A., & Tsotsos, J. K. (2013). 50 years of object recognition: Directions forward. *Computer Vision and Image Understanding, 117,* 827–891. http://dx.doi.org/10.1016/j.cviu.2013.04.005

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18,* 622–628. http://dx.doi.org/10.1111/j.1467-9280.2007.01949.x

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences, 5,* 119–126. http://dx.doi.org/10.1016/S1364-6613(00)01593-X

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience, 139,* 201–208. http://dx.doi.org/10.1016/j.neuroscience.2005.08.023

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4,* 829–839. http://dx.doi.org/10.1038/nrn1201

Bayley, P. J., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2008). Yes/no recognition, forced-choice recognition, and the human hippocampus. *Journal of Cognitive Neuroscience, 20,* 505–512. http://dx.doi.org/10.1162/jocn.2008.20038

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10), article 7. http://dx.doi.org/10.1167/9.10.7

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147. http://dx.doi.org/10.1037/0033-295X.94.2.115

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

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science, 24,* 981–990. http://dx.doi.org/10.1177/0956797612465439

Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 7459–7464. http://dx.doi.org/10.1073/pnas.1520027113

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Cabeza, R., Dolcos, F., Graham, R., & Nyberg, L. (2002). Similarities and differences in the neural correlates of episodic memory retrieval and working memory. *NeuroImage, 16,* 317–330. http://dx.doi.org/10.1006/nimg.2002.1063

Chun, M. M., & Cavanagh, P. (1997). Seeing two as one: Linking apparent motion and repetition blindness. *Psychological Science, 8,* 74–79. http://dx.doi.org/10.1111/j.1467-9280.1997.tb00686.x

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169,* 323–338. http://dx.doi.org/10.1016/S0079-6123(07)00020-9

Cox, D. D., & DiCarlo, J. J. (2008). Does learned shape selectivity in inferior temporal cortex automatically generalize across retinal position? *The Journal of Neuroscience, 28,* 10045–10055. http://dx.doi.org/10.1523/JNEUROSCI.2142-08.2008

Cox, D. D., Meier, P., Oertelt, N., & DiCarlo, J. J. (2005). “Breaking” position-invariant object recognition. *Nature Neuroscience, 8,* 1145–1147. http://dx.doi.org/10.1038/nn1519

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11,* 333–341. http://dx.doi.org/10.1016/j.tics.2007.06.010

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73,* 415–434. http://dx.doi.org/10.1016/j.neuron.2012.01.010

Flombaum, J. I., & Scholl, B. J. (2006). A temporal same-object advantage in the tunnel effect: Facilitated change detection for persisting objects. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 840–853. http://dx.doi.org/10.1037/0096-1523.32.4.840

Geusebroek, J. M., Burghouts, G. J., & Smeulders, A. W. (2005). The Amsterdam library of object images. *International Journal of Computer Vision, 61,* 103–112. http://dx.doi.org/10.1023/B:VISI.0000042993.50813.60

Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language, 66,* 68–78. http://dx.doi.org/10.1016/j.jml.2011.07.002

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458,* 632–635. http://dx.doi.org/10.1038/nature07832

Hyun, J.-S., Woodman, G. F., Vogel, E. K., Hollingworth, A., & Luck, S. J. (2009). The comparison of visual working memory representations with perceptual inputs. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1140–1160. http://dx.doi.org/10.1037/a0015019

Isik, L., Leibo, J. Z., & Poggio, T. (2012). Learning and disrupting invariance in visual recognition with a temporal association rule. *Frontiers in Computational Neuroscience, 6,* 37. http://dx.doi.org/10.3389/fncom.2012.00037

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219. http://dx.doi.org/10.1016/0010-0285(92)90007-O

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1,* 153–161. http://dx.doi.org/10.1016/S0022-5371(62)80023-1

Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No evidence for an item limit in change detection. *PLoS Computational Biology, 9,* e1002927. http://dx.doi.org/10.1371/journal.pcbi.1002927

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19,* 577–621. http://dx.doi.org/10.1146/annurev.ne.19.030196.003045

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

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience, 17,* 347–356. http://dx.doi.org/10.1038/nn.3655

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: Henry Holt.

Odic, D., Roth, O., & Flombaum, J. I. (2012). The relationship between apparent motion and object files. *Visual Cognition, 20,* 1052–1081. http://dx.doi.org/10.1080/13506285.2012.721405

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. http://dx.doi.org/10.1163/156856897X00366

Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLoS Computational Biology, 4,* e27. http://dx.doi.org/10.1371/journal.pcbi.0040027

Poth, C. H., Herwig, A., & Schneider, W. X. (2015). Breaking object correspondence across saccadic eye movements deteriorates object recognition. *Frontiers in Systems Neuroscience, 9,* 176. http://dx.doi.org/10.3389/fnsys.2015.00176

Poth, C. H., & Schneider, W. X. (2016). Breaking object correspondence across saccades impairs object recognition: The role of color and luminance. *Journal of Vision, 16*(11), article 1.2. http://dx.doi.org/10.1167/16.11.1

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197. http://dx.doi.org/10.1163/156856888X00122

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature, 435,* 1102–1107. http://dx.doi.org/10.1038/nature03687

Rust, N. C., & Stocker, A. A. (2010). Ambiguity and invariance: Two fundamental challenges for visual processing. *Current Opinion in Neurobiology, 20,* 382–388. http://dx.doi.org/10.1016/j.conb.2010.04.013

Ryan, J. D., & Cohen, N. J. (2004). Processing and short-term retention of relational information in amnesia. *Neuropsychologia, 42,* 497–511. http://dx.doi.org/10.1016/j.neuropsychologia.2003.08.011

Schurgin, M. W., & Flombaum, J. I. (2015). Visual long-term memory has weaker fidelity than working memory. *Visual Cognition, 23,* 859–862. http://dx.doi.org/10.1080/13506285.2015.1093243

Schurgin, M. W., & Flombaum, J. I. (2017). Exploiting core knowledge for visual object recognition. *Journal of Experimental Psychology: General, 146,* 362–375. http://dx.doi.org/10.1037/xge0000270

Schurgin, M. W., Reagh, Z. M., Yassa, M. A., & Flombaum, J. I. (2013). Spatiotemporal continuity alters long-term memory representation of objects. *Visual Cognition, 21,* 715–718. http://dx.doi.org/10.1080/13506285.2013.844969

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science, 20,* 207–214. http://dx.doi.org/10.1111/j.1467-9280.2009.02276.x

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *The Quarterly Journal of Experimental Psychology, 22,* 261–273. http://dx.doi.org/10.1080/00335557043000203

Smith, D. G., & Duncan, M. J. (2004). Testing theories of recognition memory by predicting performance across paradigms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 615–625. http://dx.doi.org/10.1037/0278-7393.30.3.615

Speer, N. K., Jacoby, L. L., & Braver, T. S. (2003). Strategy-dependent changes in memory: Effects on behavior and brain activity. *Cognitive, Affective & Behavioral Neuroscience, 3,* 155–167. http://dx.doi.org/10.3758/CABN.3.3.155

van den Berg, R., Vogel, M., Josić, K., & Ma, W. J. (2012). Optimal inference of sameness. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 3178–3183. http://dx.doi.org/10.1073/pnas.1108790109

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 92–114. http://dx.doi.org/10.1037/0096-1523.27.1.92

Vogt, S., & Magnussen, S. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology, 54,* 298–303. http://dx.doi.org/10.1027/1618-3169.54.4.298

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 4800–4804. http://dx.doi.org/10.1073/pnas.071028598

Wolfe, J. M. (1998). Visual memory: What do you know about what you saw? *Current Biology, 8,* R303–R304. http://dx.doi.org/10.1016/S0960-9822(98)70192-7

Yi, D. J., Turk-Browne, N. B., Flombaum, J. I., Kim, M. S., Scholl, B. J., & Chun, M. M. (2008). Spatiotemporal object continuity in human ventral visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 8840–8845. http://dx.doi.org/10.1073/pnas.0802525105

Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus, 20,* 1178–1194. http://dx.doi.org/10.1002/hipo.20864

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453,* 233–235. http://dx.doi.org/10.1038/nature06860

Received May 8, 2017
Revision received December 7, 2017
Accepted December 14, 2017 ■