

Sequential dynamics in visual short-term memory

Wouter Kool · Andrew R. A. Conway · Nicholas B. Turk-Browne

Published online: 17 September 2014
© The Psychonomic Society, Inc. 2014

Abstract Visual short-term memory (VSTM) is thought to help bridge across changes in visual input, and yet many studies of VSTM employ static displays. Here we investigate how VSTM copes with sequential input. In particular, we characterize the temporal dynamics of several different components of VSTM performance, including: storage probability, precision, variability in precision, guessing, and swapping. We used a variant of the continuous-report VSTM task developed for static displays, quantifying the contribution of each component with statistical likelihood estimation, as a function of serial position and set size. In Experiments 1 and 2, storage probability did not vary by serial position for small set sizes, but showed a small primacy effect and a robust recency effect for larger set sizes; precision did not vary by serial position or set size. In Experiment 3, the recency effect was shown to reflect an increased likelihood of swapping out items from earlier serial positions and swapping in later items, rather than an increased rate of guessing for earlier items. Indeed, a model that incorporated responding to non-targets provided a better fit to these data than alternative models that did not allow for swapping or that tried to account for variable precision. These findings suggest that VSTM is updated in a first-in-first-out manner, and they bring VSTM research into closer alignment with classical working memory research that focuses on sequential behavior and interference effects.

Keywords Visual short-term memory · Working memory · Sequential dynamics · Retroactive interference · Computational modeling

Electronic supplementary material The online version of this article (doi:10.3758/s13414-014-0755-7) contains supplementary material, which is available to authorized users.

W. Kool (✉) · A. R. A. Conway · N. B. Turk-Browne
Department of Psychology and Princeton Neuroscience Institute,
Princeton University, Princeton, NJ 08544, USA
e-mail: wkool@princeton.edu

Introduction

Visual input is constantly in flux, with frequent changes occurring due to eye movements, locomotion and object motion. Visual short-term memory (VSTM)—a subset of the working memory (WM) system that maintains visual information for a few seconds with relatively low executive demands—has often been proposed to help bridge across such changes. For example, VSTM plays a key role in integrating information across saccadic eye movements (Irwin, 1991), possibly by aligning subsequent snapshots of visual information (Henderson & Hollingworth, 1999). However, despite its oft-cited role in dynamic visual processing, few studies have examined how VSTM deals with input that is intrinsically dynamic (cf. Kumar & Jiang, 2005; Woodman, Vogel, & Luck, 2012; Xu & Chun, 2006).

Recent years have seen great advances in scientific knowledge about VSTM. The hallmark property of VSTM is that its storage capacity is exceedingly limited, as evidenced by a marked drop in performance when people are asked to remember more than 3–4 items (Cowan, 2001; Luck & Vogel, 1997). Some researchers have argued that this capacity limit reflects a fixed number of items that the system can maintain (Awh, Barton, & Vogel, 2007; Cowan, 2001; Fukuda, Awh, & Vogel, 2010; Luck & Vogel, 1997; Zhang & Luck, 2008), whereas others argue that it reflects a limited resource that can be distributed flexibly among any number of items (Alvarez & Cavanagh, 2004; Bays, Catalao, & Husain, 2009; Bays & Husain, 2008, 2009; Fougner, Suchow, & Alvarez, 2012; Frick, 1988; van den Berg, Shin, Chou, George, & Ma, 2012).

This research provided important findings and behavioral paradigms, but, for the most part, focused on static memory arrays in which all items are presented simultaneously. Because of the dynamic nature of visual information, it is important to examine how VSTM deals with such input. How does the system deploy its limited capacity in a dynamic

environment? In what way is a sequence of incoming information processed and maintained over time?

Similar questions have long been entertained in research on the more canonical form of WM—a system dependent on prefrontal cortex that maintains and manipulates information in the service of goal-directed behavior (Baddeley, 2003). For example, when asking participants to maintain a list of sequentially presented words, recall is better for early and late items (primacy and recency effects, respectively) than items in the middle of the list (Baddeley & Hitch, 1993; Murdock, 1962). Importantly, these two effects have been interpreted to reflect two different storage systems: the primacy effect stems from storage in and retrieval from long-term memory, whereas the recency effect relies on active maintenance of items in WM (Glanzer & Cunitz, 1966). Another key factor that influences memory for sequential information is proactive interference. In WM tasks, proactive interference is found when memory representations of earlier items interfere with the storage or retrieval of items later in a sequence (Underwood, 1969). The degree to which people are subject to proactive interference contributes to individual differences in WM capacity (Kane & Engle, 2000; Lustig, May, & Hasher, 2001) and fluid intelligence (Bunting, 2006; Burgess, Gray, Conway, & Braver, 2011; Gray, Chabris, & Braver, 2003).

With these findings about WM in mind, we turn to the question of how VSTM deals with sequential information. Specifically, we were interested in the extent to which VSTM shows similar temporal dynamics to WM. Does retention of input in VSTM exhibit primacy and recency biases? Is the system subject to interference effects?

The answers to these questions are not obvious. On the one hand, VSTM and WM are often treated as distinct constructs, based on psychometric and neuroimaging evidence. Psychometric studies show that VSTM and WM load on separate factors and differentially predict general cognitive abilities, such as fluid intelligence (Burgess et al., 2011; Shipstead, Redick, Hicks, & Engle, 2012). Neuroimaging studies suggest VSTM and WM may recruit distinct neural substrates: VSTM and its capacity limits have been associated primarily with occipital and parietal cortices (Harrison & Tong, 2009; Serences, Ester, Vogel, & Awh, 2009; Todd & Marois, 2004; Xu & Chun, 2006), whereas WM capacity and proactive interference are typically associated with prefrontal cortex (Braver et al., 1997; Courtney, Ungerleider, Keil, & Haxby, 1997; D'Esposito, Postle, Jonides, & Smith, 1999; Kane & Engle, 2002; Postle, Brush, & Nick, 2004). Furthermore, the two memory systems are markedly different in the tasks used to study their effects. For example, the retention interval of VSTM is on the order of a few seconds (Zhang & Luck, 2009), whereas maintenance in WM tasks can last up to a minute or more (Conway et al., 2005; Daneman & Carpenter, 1980). This difference in experimental approach may be hinting at a functional dissociation between

the systems: the flexible and volatile nature of VSTM may be useful to support binding between rapid successions of visual input, whereas WM may be more robust in order to sustain long-term goal-directed behavior. Moreover, it is unclear whether recency or interference effects commonly seen in WM tasks with longer retention intervals, would manifest over the shorter intervals typically used in VSTM tasks.

On the other hand, despite these differences, both systems share a general ability to maintain limited information over relatively short intervals. In addition, recent work has highlighted a correlation between VSTM capacity and performance on a fluid intelligence test, commonly thought to tap into working memory capacity (Fukuda, Vogel, Mayr, & Awh, 2010). Adding to this potential correspondence, neurons in areas of prefrontal cortex that are typically associated with WM also reflect capacity limitations in VSTM (Buschman, Siegel, Roy, & Miller, 2011).

Broadly speaking, there have already been several studies of serial visual memory (Hay, Smyth, Hitch, & Horton, 2007; Johnson & Miles, 2009; Phillips & Christie, 1977; Smyth, Hay, Hitch, & Horton, 2005; Smyth & Scholey, 1996). These studies have revealed hints of primacy and recency effects when entire spatial sequences are reproduced from visual memory (Hay et al., 2007; Smyth et al., 2005; Smyth & Scholey, 1996). However, it remains unclear whether these effects are intrinsic to the way in which items are stored in VSTM or whether they only arise during the retrieval of full sequences. Furthermore, all of these studies employed categorical (correct/incorrect) accuracy scores when measuring VSTM (cf. Gorgoraptis, Catalao, Bays, & Husain, 2011). Over the last decade, modeling techniques have been developed to describe several components of VSTM in a more sophisticated manner, allowing the probability that an individual item has been stored, along with the precision of this representation, to be estimated quantitatively. The current study uses this general approach, incorporating three of the most prominent statistical models from the literature, to shed new light on the nature and dynamics of sequential VSTM.

In summary, less is known about how VSTM copes with dynamic input than static input, and the WM literature and recent developments in VSTM research provide a useful framework for addressing these issues.

Experiment 1

Our first step in investigating the temporal dynamics of VSTM was to adapt a continuous-report VSTM task in which participants view an array of colored squares and recall the color of a probed item on a color wheel (Wilken & Ma, 2004). Using statistical likelihood estimation techniques, response deviations in this task allow for measurement of both the probability of storage and the precision of the stored

representation (Zhang & Luck, 2008). Specifically, the distribution of response deviations can be modeled as a mixture of two different trial types. On trials when the target color is not stored in memory, the observer responds randomly, resulting in a uniform distribution over color space; one minus the height of this distribution thus reflects the probability of items being stored. On trials when the target is stored, the response will be centered on the original color, giving rise to a normal distribution; the width of this distribution reflects the precision with which the items are stored. We adapted this paradigm to present items sequentially, so that we could assess the temporal dynamics of storage probability and precision (Fig. 1).

This model has most prominently been used by proponents of the view that VSTM consists of a discrete number of slots to maintain a fixed number of items, since it assumes that items are either stored or completely forgotten. As mentioned before, other theorists have argued that VSTM’s limited

capacity is set by a continuous, flexible resource (see Experiment 3), and new, related, computational models have been developed to distinguish between these two possibilities. In this paper, we will fit all of these models to the sequential VSTM task.

Methods

Participants

Twenty-two adults (aged 22–34 years, 15 females) participated for monetary compensation. Some participants completed two of the three conditions in the experiment, and others completed just one condition (resulting in ten participants per condition). Set size was manipulated between rather than within participants, due to the need to estimate VSTM parameters that every serial position (unlike the standard spatial VSTM task, where separate parameters are not estimated for each spatial position). The participants provided informed consent, following a protocol approved by the Princeton University Institutional Review Board. All participants reported normal color vision and normal or corrected-to-normal visual acuity.

Materials

Stimuli were generated with the Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997) and presented on a 17-inch CRT monitor (refresh rate 100 Hz) in a dimly lit room. Participants sat approximately 60 cm from the computer monitor. The color configuration of the monitor was calibrated with an X-Rite i1Display 2 colorimeter (X-Rite, Grand Rapids, MI).

The experiment comprised the sequential presentation of colored squares, each subtending $2 \times 2^\circ$ of visual angle. The squares were presented in a randomly selected subset of eight equally spaced locations on an invisible circle (4.5° radius). Next, a color wheel and outlined squares at the location of each item in the sample array were presented. One outline, thicker than the others, indicated the item to be recalled. The wheel consisted of 180 colors that were evenly distributed along a color circle in the L*a*b color space of the 1976 Commission Internationale de l’Eclairage (centered at L = 70, a = 20, b = 38). Colors in the arrays were also selected from this set, with a minimum distance of 24° in color space between selected values. To eliminate response biases and minimize contributions from spatial memory, both the color wheel and the stimulus circle were randomly oriented on each trial.

Procedure

Before modifying the standard static version of the continuous-report task (Wilken & Ma, 2004; Zhang & Luck, 2008), we first successfully replicated the findings from that

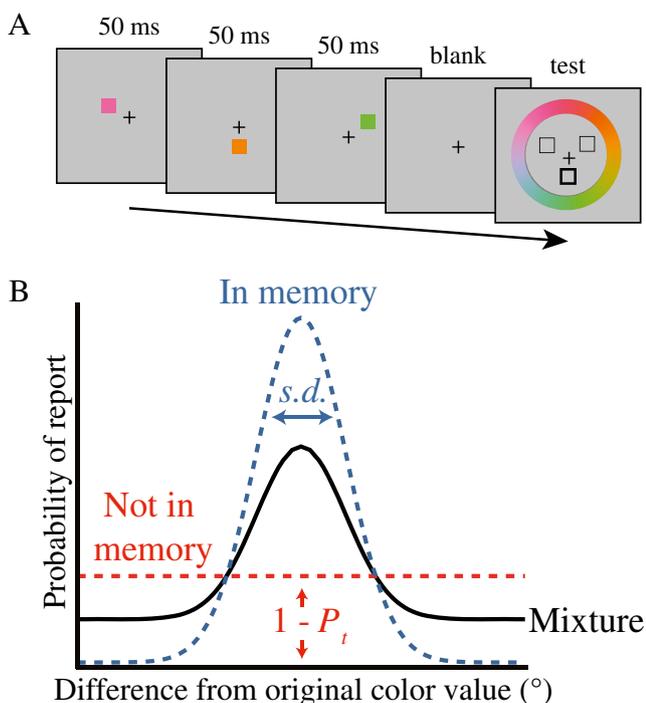


Fig. 1 **a** Visual short-term memory (VSTM) color recall task. Participants maintained fixation and were instructed to remember all colors in the array of squares (set size = three, four, or five) displayed on screen. After a short blank interval, participants were instructed to recall, by clicking on a color wheel, the color of the square at the location indicated by the thicker outline. **b** Two-component mixture model of response deviation (adapted from Zhang & Luck, 2008), showing the probability of reporting each deviation from the true color value. Two different sets of trials contributed to the probability function: when the probed target item was maintained, the response was normally distributed around the original value (blue line); when the target item was not maintained, the response was randomly drawn from a uniform distribution (red line). The aggregation of all trials produces a mixture of these two distributions (black line), with parameters: P_t probability that target items were stored in memory, $s.d.$ precision with which stored items were represented

task in a separate group of 11 participants. In our modified version, each square in the sample array was presented for 50 ms in a random order. This stimulus presentation time was chosen so as to make the presentation time per item as close as possible to the original design by Zhang and Luck (2008), while not rendering the stimulus time too short to be visible. Using a short stimulus duration time likely increased the average *s.d.* score and guessing rate (Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011), but as will become clear through this paper, we still obtained numerous reliable memory results across set sizes and serial positions, holding these sensory/encoding constraints constant.

For each trial, the duration of the blank interval between the offset of the last item and the onset of the color wheel was chosen so that exactly 1000 ms passed between the onset of the to-be-probed item and the onset of the color wheel. This way, any differences in storage probability or precision estimates across serial position could not simply reflect differences in retention time (as described by Zhang & Luck, 2009). We probed each serial position in the sequence with equal frequency. After the onset of the test display, participants were instructed to click on the color wheel to indicate the color that had appeared in the probed location.

The set size of the array was manipulated across participants to be three, four, or five (ten participants each). Depending on the set size, the experiment took 1–1.5 h to complete and was composed of ten blocks of 45, 60, or 50 trials (for set sizes three, four, or five, respectively). This resulted in 150 total trials per serial position for set sizes three and four, and 100 total trials per position for set size five (reduced to avoid lengthening the experiment beyond 1.5 h). Participants were offered a break after every block.

Analysis

We used a quantitative model to estimate the proportion of trials on which the target color was stored and the precision of stored representations (Zhang & Luck, 2008), but now separately for each serial target position. To minimize practice effects, we excluded the first block of trials from analysis. Maximum likelihood estimation was used to fit the distribution of response deviations with a mixture of a von Mises distribution (the circular analogue of a standard Gaussian) and a uniform distribution. This model was fitted to the data with parameters *s.d.*, the width of the von Mises distribution (inversely proportional to mnemonic precision), and P_t , the proportion of responses accounted for by this distribution (the proportion of responses in which the target was recalled). The model also includes a uniform component to describe the remaining share of responses ($1-P_t$), corresponding to those trials in which subjects failed to recall the target color and made a random guess over the color space. The mathematical

specification of the probability density function of this model is as follows:

$$p(\hat{x}) = P_t \phi(\hat{x}-x, s.d.) + (1-P_t) \frac{1}{360}$$

where x is the original target color (in degrees), \hat{x} is the reported color value, P_t is the proportion of responses on which the target color was successfully recalled, and ϕ denotes the von Mises distribution with standard deviation *s.d.* and mean 0.

We used maximum-likelihood estimation, implemented with Matlab's *fmincon* function, to fit this model for each serial position in the stimulus arrays, and in each set size condition, separately. The value of parameter P_t was constrained with an upper bound of 1 and a lower bound of 0. The value of parameter *s.d.* was constrained with an upper bound of 180 degrees and a lower bound of 1 degree. To avoid local optima in the estimation solution, we ran 500 iterations for each model with randomly selected starting positions for each parameter. The final estimations of *s.d.* and P_t were extracted from the iteration with maximal log-likelihood (Myung, 2003).

Results

Model-free descriptive statistics

The raw distributions of error are shown in Fig. 2 and the raw circular SDs of error can be found in Table 1.

Storage probability

Figure 3a shows how P_t estimates varied as a function of the serial position in which the target appeared in the sequence, separately for each set size condition. Collapsing over target serial position revealed a reliable main effect of set size on P_t , $F(2, 27) = 24.30$, $P < 0.001$, $\eta^2 = 0.64$, and significant linear contrast as a function of set size, $P < 0.001$.

There was clear evidence of a recency effect for larger set sizes—a benefit for the later serial positions. Three repeated-measures ANOVAs with the within-subject factor “target serial position” and P_t as the dependent measure indicated a significant effect of target serial position on storage probability for set sizes four, $F(3, 27) = 10.50$, $P < 0.001$, $\eta^2 = 0.54$, and five, $F(4, 36) = 5.80$, $P < 0.01$, $\eta^2 = 0.39$, but not for set size three, $F(2, 18) = 3.04$, $P = 0.07$, $\eta^2 = 0.25$. More specifically, for set sizes four and five, the linear trend of serial position was significant, $F(1, 9) = 16.43$, $P < 0.01$, $\eta^2 = 0.65$, and $F(1, 9) = 13.69$, $P < 0.01$, $\eta^2 = 0.60$, respectively. This indicates that later items were recalled with higher probability than earlier items. The linear trend was not significant for set size three, $F < 1$, suggesting that all items were stored with equal probability.

To test for differences in these linear trends between set size conditions, we conducted a linear regression for each

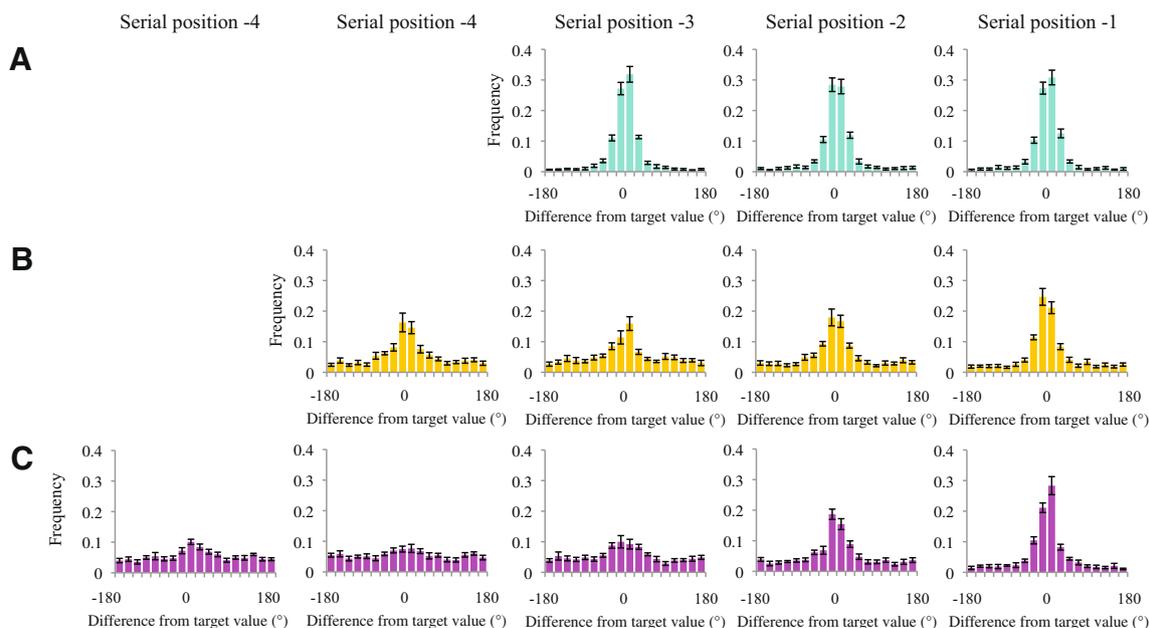


Fig. 2 Raw distributions of response error for each serial position in the set size **a** three, **b** four, and **c** five conditions of Experiment 1. Error bars indicate across-subject standard error of the mean (SEM)

participant predicting P_t as a function of serial position. The slopes of these regressions were used as the dependent variable in a one-way ANOVA with “set size condition” as a between-subjects factor. There was a significant main effect of set size, $F(1, 27) = 29.81, P < 0.001, \eta^2 = 0.53$. Post-hoc Bonferroni-corrected t -tests revealed that this reflected a smaller slope in set size three vs four, $t(18) = 2.83, P < 0.05$, Cohen’s $d = 1.79$, and three vs five, $t(18) = 4.22, P < 0.01, d = 1.84$. There was no difference between set sizes four and five, $t(18) = 1.39, P = 0.53, d = 0.52$.

There was also a hint of a primacy effect—a benefit for the first serial position. The most obvious primacy effect would be if the first position had a higher P_t than the second position, but this was found inconsistently: set size three, $t(9) = 1.94, P = 0.08, d = 0.61$; set size four, $t(9) = 2.30, P < 0.05, d = 0.73$; set size five, $t(9) = 1.52, P = 0.16, d = 0.48$. However, under the assumption that the recency effect also applies to the first position, it may drive P_t for the first position in the opposite direction, thereby obscuring the primacy effect. To address this, we ran new linear regression analyses to predict P_t as a function

of serial position for all but the first item in the sequence. The resulting slope and intercepts were then used to predict P_t for the first item, as if the curves were indeed linear (open data points in Fig. 3a), with the restriction that these extrapolated scores could not be negative. These predicted values provide a baseline for testing the primacy effect, having accounted for the mitigating influence of the recency effect. Storage probability for the first item was reliably higher than this baseline for both set size four, $t(9) = 3.17, P < 0.05, d = 1.00$, and five, $t(9) = 2.43, P < 0.05, d = 0.77$, but not for set size three, $t(9) = 2.07, P = 0.07, d = 0.65$. For comparison, estimates for P_t of the second position, as predicted using the same approach of a linear regression based on subsequent items (not possible for set size three), did not differ from the observed P_t for set size four or five, $P_s > 0.41$.

Memory precision

Figure 3b depicts estimates of precision, $s.d.$, over the target serial positions and for each set size. Collapsing over target serial position did not reveal a main effect of set size on $s.d.$,

Table 1 Average raw circular SDs (in degrees) in Experiment 1

Set size	Serial position					Linear contrast	
	-5	-4	-3	-2	-1	$F(1, 9)$	P
3	–	–	35.7 (0.8) ^a	39.1 (0.9)	36.3 (0.7)	0.3	0.539
4	–	62.7 (1.4)	66.6 (1.3)	60.6 (0.7)	41.8 (2.1)	17.5	< 0.01
5	73.6 (1.0)	75.7 (1.1)	71.0 (1.9)	62.7 (2.9)	48.4 (3.3)	49.6	< 0.001

^a The error terms, in parentheses, reflect the within-subjects SEM

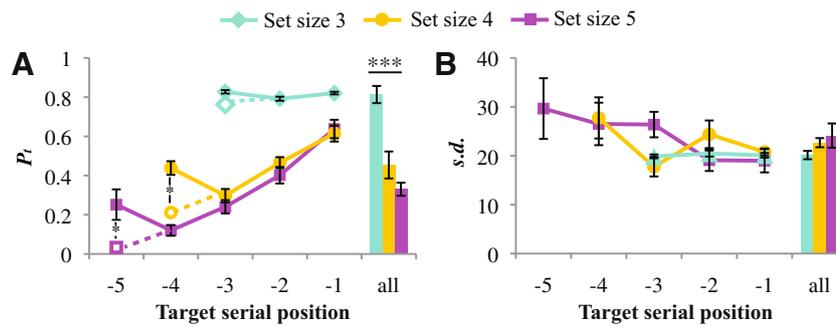


Fig. 3 **a, b** Results from Experiment 1. **a** P_i estimates over target serial position (-1 = latest in the array) for the different set size conditions. While storage probability did not reliably differ over the positions for set size three, a positive linear relationship was found for set sizes four and five. The *open points* and *dashed lines* represent the extrapolation of the linear trend observed for all but the first item in the array. Comparison of this projected value with the observed value suggested that there was a

$F(2, 27) = 1.34, P = 0.28, \eta^2 = 0.09$. In repeated-measures ANOVAs for each set size, there was not a main effect of target serial position on *s.d.*, $P_s > 0.16$. As a result, these data were not analyzed further.

Discussion

Our results revealed dissociations in the sequential dynamics of VSTM across set sizes but also between memory parameters. Consistent with the previous literature (Zhang & Luck, 2008), we found no straightforward effect of set size or serial position on precision estimates. The measure of storage probability, however, was strongly affected by serial position. For larger set sizes, items that were presented later in the array were recalled with higher probability than those presented earlier (a recency effect), whereas all colors in a sequence of three items were equally likely to be stored. Interestingly, the set size at which the storage curve started showing a recency effect, about four items, corresponds well to previously established capacity limits of VSTM (Luck & Vogel, 1997). As considered in more detail in the General Discussion, this suggests that VSTM is updated in a first-in-first-out (FIFO) fashion: when capacity is not exceeded, all items are stored with equal probability, but with larger memory arrays, the items stored earlier are given less priority.

An important exception to this rule is the first item, which seems to receive some form of a primacy benefit at the larger set sizes. At first glance, this benefit may seem inconsistent with the FIFO principle. However, evidence for FIFO is obtained by comparing performance *across* larger vs smaller set sizes. Specifically, the probability of storing the first item in a trial is a function of how many other items follow in sequence (compare serial position -3 for set size 3, to -4 for set size 4, to -5 for set size 5). The primacy effect, on the other hand, is observed *within* the storage probability curve for a given set size. We interpret the overall response profile as reflecting the operation of two or more processes that independently contribute recency and primacy

effects. A similar perspective is widely accepted in the broader memory literature. For example, in immediate free recall, recency effects occur because of persisting item activation and because the temporal context at retrieval more closely matches that of later items, and primacy effects occur because there is more time for rehearsal and consolidation of earlier items and because of a lack of proactive interference during their encoding.

Experiment 2

Although Experiment 1 provided one of the first systematic investigations of the sequential dynamics of VSTM using the continuous-report task, it involved a design decision that may have played an important role in the results. Specifically, the length of the blank interval between the final item and test display depended on the serial position of the target. This design was used to equate the total maintenance time for the probed item across serial positions (1000 ms). However, it is possible that participants exploited these differences, such as to know that if the blank interval was long enough, that the earlier items were not going to be probed. In other words, participants might have cleverly cleared their memory during the blank interval, leading to storage probability curves with “strategic” recency effects. The alternative—using a fixed blank interval—results in the maintenance time varying across serial positions. Here we attempt to replicate the results of Experiment 1 for the middle set size using this design, in order to demonstrate that this decision was not consequential.

Methods

Participants

Ten adults (aged 21–34 years, 7 females) participated for monetary compensation.

Materials, procedure, and analysis

The design was nearly identical to that of Experiment 1. All sample arrays contained four squares. The only modification was that the blank interval between the offset of the final item and the onset of the color wheel was always 1000 ms, regardless of the serial position of the probed item. The same maximum-likelihood estimation technique from Experiment 1 was used to find parameter fits for $s.d.$ and P_t for each serial position and each participant separately.

Results and discussion

Two repeated-measures ANOVAs with the within-subject factor “target serial position” indicated that there was a significant effect of serial position on storage probability, $F(3, 27) = 7.29$, $P < 0.001$, $\eta^2 = 0.45$, but again not on precision, $F(3, 27) = 1.71$, $P = 0.19$, $\eta^2 = 0.16$. Storage probability showed a recency effect, as demonstrated by a significant linear trend over target positions, $F(1, 9) = 8.50$, $P < 0.05$, $\eta^2 = 0.49$. There was also a primacy effect in terms of storage probability for the first item with respect to what was predicted by a purely linear trend computed from the other serial positions, $t(9) = 2.36$, $P < 0.05$, $d = 0.74$. Thus, we replicated the key results of Experiment 1 despite using a fixed blank interval for all serial positions. This casts doubt on an alternative explanation of the recency effect in that experiment, whereby participants detected the length of the blank interval and used it strategically to manage memory resources.

Experiment 3

The two-component mixture model (Zhang & Luck, 2008), with which the data from Experiments 1–2 were analyzed, makes the explicit assumption that the distribution of response deviations reflects two types of trials: those in which the target is stored and recalled with a certain amount of noise (contributing the circular Gaussian component) and those in which the target is not stored and the participant guesses randomly over the color space (resulting in the uniform component). However, response deviations in a continuous-report VSTM task sometimes reflect errors in stimulus location (Bays et al., 2009). Specifically, on some trials, participants “swap” the color-location associations and recall one of the non-target colors. To capture these trials, a three-component model was developed that additionally included Gaussians centered on non-target colors (Fig. 4).

Here, we use this approach to examine not only which items in a sequence are remembered and with what precision, but also to characterize interference between items in VSTM. In the case of proactive interference, non-target responses

should happen more frequently when the target appears later in the sequence. Alternatively, in the case of retroactive interference, one would predict that when targets appear early in the sequence, non-target responses are more likely. Indeed, these two forms of interference might be related to the observed primacy and recency effects in Experiments 1–2, respectively.

The sequential nature of our task allows us to go one step further in the analysis of interference effects. Not only can we model at which target serial positions swapping is likely to occur, but we can also investigate *which* non-target items in the sequence are most likely to be “swapped in”. One natural prediction, based on our previous results of robust recency effects, would be that later items are most likely to be swapped in, as a form of retroactive interference.

Here, we examine such swapping in a replication of the boundary set size conditions (three and five) from Experiment 1. It has been argued in the literature that a minimum distance in feature space between items of the same array (as was used in Experiments 1 and 2) is problematic for modeling non-target responding, as the feature values are not assigned independently (Fougnie, Asplund, & Marois, 2010). Thus, in Experiment 3, we assign all color values randomly.¹

Methods

Participants

Twenty-eight adults (aged 18–23 years, 7 females) participated for monetary compensation or course credit.

Materials and procedure

The design was nearly identical to that of Experiment 1. One group of participants ($n = 14$) always saw sample arrays with three items, whereas the other group ($n = 14$) saw sample arrays with five items. The colors in these arrays were selected from the color set described in Experiment 1. Crucially, there was no minimal distance in color space between the colors in the arrays—they were chosen randomly from a uniform distribution.

Analysis

We used a three-component model to estimate the proportion of trials on which the target color was reported, the proportion

¹ We thank Paul Bays for noting that this concern may not apply in principle to maximum likelihood estimation. Indeed, we obtained nearly identical results when the three-component model was applied to Experiment 1 (see [Supplemental Material](#)).

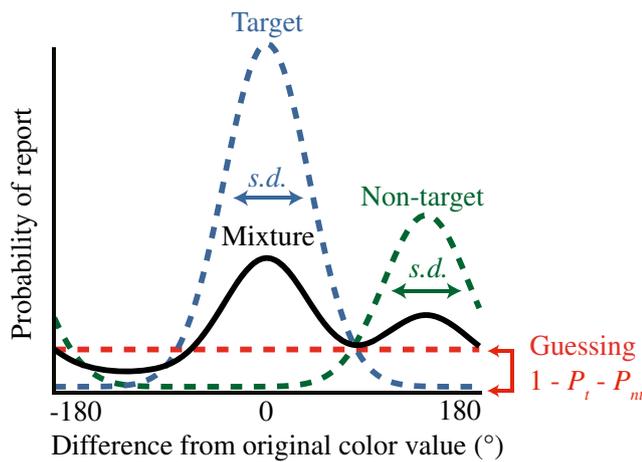


Fig. 4 Three-component mixture model of response deviation (adapted from Bays et al., 2009). The model is largely similar to the one depicted in Fig. 1b, but performance on the task is now also explained by responses to non-target colors with additional Gaussians of precision $s.d.$. The model has parameters for the proportions of target responses, P_t , and non-target responses, P_{nt}

of trials on which a non-target color was reported, and the precision of these reports (Bays et al., 2009). However, to fully exploit the sequential nature of our task, we allowed the weights of these non-target components to vary as a function of the serial position of the non-target color:

$$p(\hat{x}) = P_t \phi(\hat{x} - x, s.d.) + P_{nt} \sum_{i=1}^m w_i \phi(\hat{x} - x_i^*, s.d.) + (1 - P_t - P_{nt}) \frac{1}{360}$$

where P_{nt} is the proportion of responses to non-target colors, $\{x_1^*, x_2^*, \dots, x_m^*\}$ are the color values of the m non-target items (in serial order), and $\{w_1, w_2, \dots, w_m\}$ are the relative proportions of each of the non-target Gaussian components (in serial order) with $\sum w_i = 1$. The value of the parameters P_t , P_{nt} , and their sum were constrained with an upper bound of 1 and a lower bound of 0. Similarly, the values of parameters w_i and their sum were constrained with an upper bound of 1 and a lower bound of 0. The value of parameter $s.d.$ was constrained with an upper bound of 180 degrees and a lower bound of 1 degree.

As in the previous experiments, we used maximum-likelihood estimation to fit this model for each serial position in the stimulus arrays, and each participant separately. To avoid local optima in the estimation solution, we ran 500 iterations for each model with randomly selected starting values for each parameter. The final estimations of $s.d.$, P_t , P_{nt} , and w_i s were extracted from the iterations with maximal log-likelihood.

Results

Model-free descriptive statistics

The raw distributions of error are shown in Fig. 5 and the raw circular SDs of error can be found in Table 2.

Target storage probability

The temporal dynamics of P_t as estimated by the three-component model were comparable to the two-component model (Fig. 6a). Collapsing over target serial position, P_t was reliably higher for set size three vs five, $t(26) = 5.94$, $P < 0.001$, $d = 2.25$. Storage probability varied over serial positions for set size five, $F(4, 52) = 16.88$, $P < 0.001$, $\eta^2 = 0.57$, but not for set size three, $F < 1$. Storage probability showed a recency effect, as evidenced by a significant linear contrast over serial positions for set size five, $F(1, 13) = 22.97$, $P < 0.001$, $\eta^2 = 0.64$, but not set size three, $F(1, 13) = 1.12$, $P = 0.31$, $\eta^2 = 0.08$. The slopes of these linear fits were reliably higher for set size five vs three, $t(26) = 3.01$, $P < 0.01$, $d = 1.14$. We also replicated the primacy effect for set size five. The storage probability of the first item was greater than for the second item, $t(13) = 2.12$, $P = 0.05$, $d = 0.57$, and again reliably larger than predicted by a linear response profile over the other serial positions (with a minimum predicted value of 0), $t(13) = 4.72$, $P < 0.001$, $d = 1.26$. For set size three, there was no primacy effect, even compared to the projected linear trend, $t < 1$.

For comparison with Experiment 1, we ran two repeated-measures ANOVAs with the within-subject factor “target serial position”, the experiment as a between-subject factor, and P_t as the dependent variable for both set sizes. The resulting non-significant interaction terms, $P_s > 0.60$, suggest that the color selection method did not reliably affect target storage probability.

Precision

Collapsing over target serial position did not reveal a main effect of set size on $s.d.$, $t(26) = 1.57$, $P = 0.13$, $d = 0.59$. Consistent with our earlier findings, precision again did not reliably vary over serial positions for set size three, $F(2, 26) = 2.39$, $P = 0.11$, $\eta^2 = 0.16$, or for set size five, $F(4, 52) = 1.70$, $P = 0.17$, $\eta^2 = 0.12$ (Fig. 6b).

Guessing

In the two-component model used for Experiments 1 and 2, the rate of guessing is one minus the storage probability ($P_g = 1 - P_t$). In the three-component model, however, there is one term for storage probability, but also a term for the rate of non-target responding, and thus the guessing rate reflects one

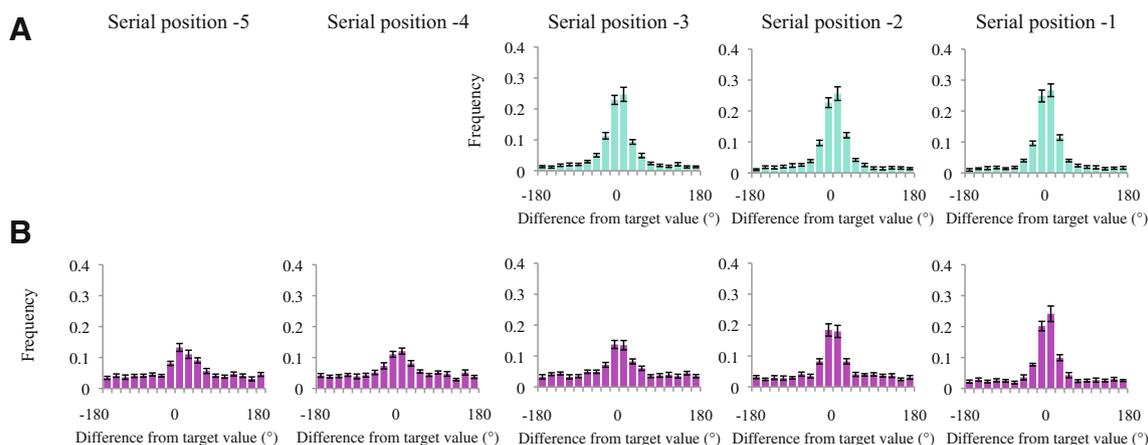


Fig. 5 Raw distributions of response error for each serial position in the set size **a** three and **b** five conditions of Experiment 3. Error bars indicate across-subject SEM

minus both terms ($P_g = 1 - P_t - P_{nt}$). Participants showed a considerable amount of guessing for both set size three, $M = 0.21$, $SD = 0.17$, and set size five, $M = 0.24$, $SD = 0.12$ (Fig. 6c). However, this measure did not vary over serial target positions for either set size, $P_s > 0.53$, nor was it different between set sizes, $t < 1$.

Non-target responding

Collapsing over target serial position, the proportion of non-target responses, P_{nt} (reflecting both the probability of storage of non-targets and the tendency to respond with the color of the non-target items), was reliably higher for set size five vs three, $t(26) = 9.69$, $P < 0.001$, $d = 3.66$. There was a main effect of target serial position for set size five, $F(4, 52) = 6.12$, $P < 0.001$, $\eta^2 = 0.32$, but not for set size three, $F < 1$. Linear contrasts revealed that non-target responses became less likely when later items were probed for set size five, $F(1, 13) = 4.55$, $P < 0.01$, $\eta^2 = 0.53$, but not set size three, $F < 1$. The slopes of these linear fits were reliably higher for set size five vs three, $t(26) = 3.08$, $P < 0.01$, $d = 1.39$.

It is possible that the proportion of non-target responding was higher for earlier targets because these items were not as likely to have been stored in VSTM (see Figs. 3a and 6a). If true, the slopes of the P_t and P_{nt} curves should be negatively related: the less likely you are to store earlier items, the more likely you should be to respond with a non-target color. We

tested this hypothesis by correlating the recency regression slopes for P_t (excluding the first serial position) and P_{nt} across participants for each set size. Because of our small sample size for correlational analyses, we calculated robust correlations using the skipped-correlation method of Pernet, Wilcox, and Rousselet (2012). This method weighs down or removes outliers (which can be particularly influential in small-sample correlations) and accounts for them in significance testing. Consistent with our prediction (Fig. 7), the slopes of P_t and P_{nt} were negatively correlated for set size five (Pearson $r = -0.55$, $CI = [-0.81, -0.21]$; Spearman $r = -0.55$, $CI = [-0.85, -0.04]$), but not for set size three (Pearson $r = 0.38$, $CI = [-0.15, 0.79]$; Spearman $r = 0.38$, $CI = [-0.26, 0.82]$).

To formally test the difference between these relationships, we conducted a regression analysis with the dependent variable being the slope of the P_{nt} curve, and the P_t slope, set size, and their interaction as the predictors. Importantly, the interaction term was significant, $t(24) = 2.27$, $P < 0.05$, suggesting that the strength of the relationship between the P_{nt} and the P_t slopes was different between the two set size conditions.

So far, the analyses have revealed that responses to non-targets contributed to VSTM performance, that the rate of this swapping followed an interesting trajectory over serial positions, and that these sequential dynamics were related to the rate at which people stored target items. However, our experimental design allowed us to go one step further and investigate *which* items in the sequence were most likely to be

Table 2 Average raw circular SDs (in degrees) in Experiment 3

Set size	Serial position					Linear contrast	
	-5	-4	-3	-2	-1	$F(1, 9)$	P
3	–	–	46.5 (1.3)	45.6 (1.1)	44.3 (1.1)	1.42	0.25
5	69.4 (2.0)	70.5 (1.3)	68.1 (1.3)	61.7 (0.8)	55.8 (2.2)	25.4	< 0.001

^a The error terms, in parentheses, reflect the within-subjects SEM

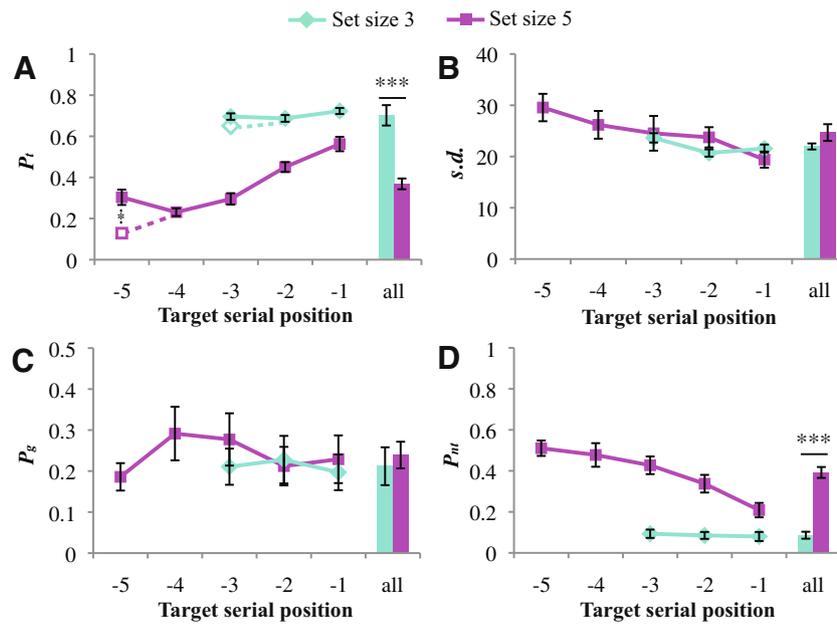


Fig. 6 a–d Results from Experiment 3. **a** The probability of storing targets, P_t , showed a recency effect only for set size five. **b** Precision estimates, $s.d.$, did not reliably vary over the target serial positions for either set size. **c** The guessing rate, P_g , also did not reliably vary over target serial position or set size. **d** The probability of reporting a non-

target color, P_{nt} , showed a strong decreasing trend for set size five, but not three. Mean estimates for each set size are depicted in the *bar graph* on the *right*. *Error bars* indicate within-subject SEM, except for the *bar graph*, which depicts across-subject SEM for each set size condition. * $P < 0.05$, *** $P < 0.001$

swapped in. Recall that, for each model fitted to a particular target serial position and set size, we estimated the relative share, or weight (w_i), of each non-target serial position to the non-target response Gaussians.

To analyze how these weights varied as a function of the non-target serial position, we computed for each participant

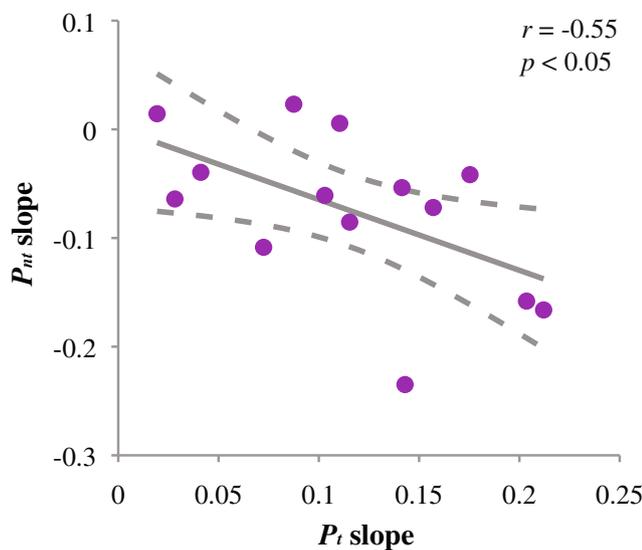


Fig. 7 Relationship between P_t and P_{nt} slopes over target serial positions for set size five. We found a negative correlation, suggesting that people with a stronger recency effect showed more of a bias to report non-targets when probed on earlier items. *Dashed lines* indicate the 95 % confidence interval

the average of the weights at a given non-target serial position over all of the models (for different target serial positions) at that set size (Fig. 8). For set size five, there was a main effect of non-target serial position on the average weight, $F(4, 52) = 3.00, P < 0.05, \eta^2 = 0.53$, but this was not the case for set size three, $F < 1$. The effect for set size five was characterized by a significant linear contrast, $F(1, 13) = 5.43, P < 0.05, \eta^2 = 0.30$, but not for set size three, $F < 1$; however, there was no difference in slope between set sizes, $t < 1$. Thus, items appearing later in larger arrays were more likely to be swapped in than earlier items; whether this occurs for smaller arrays, at all or to the same extent as for larger arrays, is unclear because of the null effects.

Finally, we tested whether the non-target serial position effect for set size five was driven by non-target weights from just a subset of models (i.e., for one or more particular target serial positions). A repeated-measures ANOVA with the within-subject factors “target serial position model” and “non-target serial position”, and “non-target serial position weights” as the dependent measure revealed no interaction, $F < 1$. This suggests that the likelihood of swapping in later items may not have depended upon which item was probed.

Discussion

The results of Experiment 3 confirmed our earlier findings, now using a three-component model. As before, storage probability was uniform over serial positions at the smaller set size, but showed primacy and recency effects at the larger set size.

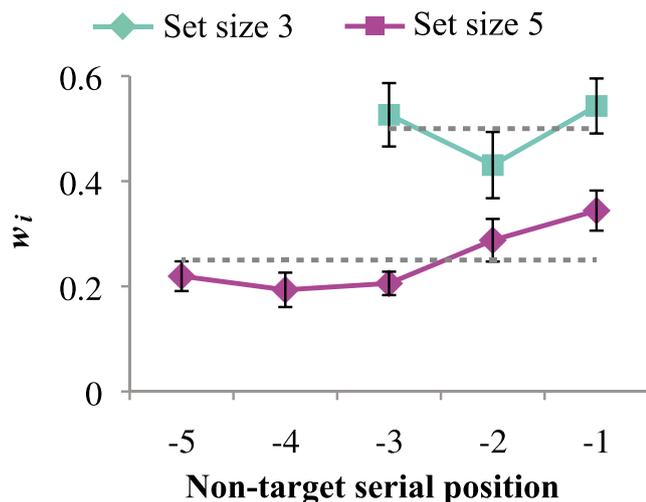


Fig. 8 Average weights of non-target responding for each non-target serial position, for set size three and five. In essence, these plots show the non-target serial positions from which items were reported instead of the target (irrespective of target serial position). *Dashed lines* indicate the expected average weight if distributed uniformly over the (set size – 1) non-target serial positions. We found a significant linear trend over serial positions for set size five, but not for set size three. *Error bars* indicate within-subject SEM

We again found no effect of set size or target serial position on mnemonic precision.

Most importantly, this new model allowed us to examine the sequential dynamics of non-target responding. Erroneous recall of non-target items was more likely when the probed item appeared earlier in the sequence. Together with the recency effect in target storage probability, one interpretation of this result is that swapping occurred when the target item was not in VSTM. Consistent with this view, over serial positions, the strength of the recency effect was inversely correlated with the amount of non-target responding. Moreover, when swapping occurred, items in later serial positions were more likely to be erroneously recalled than those in earlier positions.

Variable precision model

The FIFO updating profile we observed in the previous experiments is likely a consequence of the limited capacity of VSTM. However, it is largely unclear what form this capacity limit takes and how it interacts with sequential input on a mechanistic level. One possibility, derived from the work of Zhang and Luck (2008), is that the system can maintain three discrete items and discards earlier items when new items come in. An alternative view, however, is that VSTM does not consist of a discrete number of fixed-resolution slots, but rather consists of a continuous resource that can be flexibly distributed across an unlimited number of items. In fact, the three-component model described in Experiment 3 is an

adaptation of the model described by Bays and colleagues (2009), which was formulated to support this theory. In our task, this model would predict that the system devotes more of the resource to recent items, reallocating this resource from earlier items for larger set sizes.

More recently, a new class of resource models of performance in the continuous-report task has emerged (Fougnie et al., 2012; van den Berg et al., 2012). Similar to the work by Bays and colleagues (2009), these models also assume that the capacity of VSTM can be flexibly distributed across items in the display, but, additionally, that this capacity is variable within an individual. In other words, the amount of VSTM “resource” varies from trial to trial, and this means that any computational model should not only include a parameter for the mean precision, but also describe the variance in precision across items in the display.

The results described so far seem at odds with these continuous-resource models, since both serial position and set size have primarily showed an effect on the P_t and P_{nt} parameters, and not on $s.d.$ —our measure of mnemonic precision. However, it is possible that we did not find effects of precision because the previous models did not account for variability in precision. In fact, it has been argued that the variable precision model fits distributions of recall errors better than the standard models (Suchow, Brady, Fougnie, & Alvarez, 2013; van den Berg et al., 2012). We therefore fit the data from Experiment 3 with these models.

Fougnie and colleagues (2012) showed that when error is normally distributed around the true color value, and precision follows a gamma distribution, the resulting function is described by a Student’s t -distribution, so that the mixture model is as follows,

$$p(\hat{x}) = P_t \psi(\hat{x} - x, \overline{s.d.}, \nu) + (1 - P_t) \frac{1}{360}$$

where ψ is the wrapped generalized Student’s t -distribution, $\overline{s.d.}$ is the average precision, and ν is the SD of the precision. We used the maximum-likelihood estimation procedure in Memtoolbox (Suchow et al., 2013) to fit response deviations with this model for each serial position and each participant in Experiment 3.

Results

Target storage probability

The temporal dynamics of P_t estimates in the variable precision model followed the same pattern as the two- and three-component models (Fig. 9a). Average P_t was reliably higher for set size three vs five, $t(26) = 5.58, P < 0.001, d = 2.11$. Storage probability varied over serial positions for set size five, $F(4, 52) = 4.58, P < 0.01, \eta^2 = 0.26$, but not for set size

three, $F < 1$. Storage probability showed a recency effect, as evidenced by a significant linear contrast over serial positions for set size five, $F(1, 13) = 8.98$, $P < 0.05$, $\eta^2 = 0.41$, but not set size three, $F < 1$, and a comparison of slopes from linear regressions revealed that this effect was reliably stronger for set size five vs three, $t(26) = 2.36$, $P < 0.05$, $d = 0.89$. However, we did not replicate the reliability of the primacy effect for set size five with the variable precision model. The storage probability of the first item was not reliably greater than for the second item, $t(13) = 1.21$, $P = 0.25$, $d = 0.33$, or than for the prediction extrapolated from a linear response profile over the other serial positions, $t(13) = 1.89$, $P = 0.08$, $d = 0.50$. For set size three, we found no primacy effect, $ts < 1$.

Precision

The overall difference in $\overline{s.d.}$ between set size three and five did not reach significance, $t(26) = 1.90$, $P = 0.07$, $d = 0.71$. Average precision reliably varied over serial positions for set size three, $F(2, 18) = 4.37$, $P < 0.05$, $\eta^2 = 0.25$, but not for set size five, $F(4, 52) = 1.39$, $P = 0.25$, $\eta^2 = 0.10$. Post-hoc contrasts revealed a significant quadratic trend, $F(1, 13) = 7.05$, $P < 0.05$, $\eta^2 = 0.35$, but not a linear trend for set size three, $F(1, 13) = 1.78$, $P = 0.21$, $\eta^2 = 0.12$ (Fig. 9b).

The overall difference in the variability in precision, v , between set size conditions did not reach significance, $t(26) = 1.79$, $P = 0.09$, $d = 0.68$. Variability in precision also did not show a significant effect over serial positions for either set size three, $F(2, 18) = 1.02$, $P = 0.38$, $\eta^2 = 0.07$, or set size five, $F(4, 52) = 1.23$, $p = 0.31$, $\eta^2 = 0.09$ (Fig. 9c).

Model comparison

All of the models that we have used provided some insight into the temporal dynamics of VSTM. However, beyond describing the data in different ways, we were also interested in determining which of these models provided the best description. Thus, using the data from Experiment 3, we compared the goodness-of-fit of the two-component model (as used in Experiments 1–2) to both the three-component model with serial non-target responding weights and the variable precision model. We computed the AICc, a measure of goodness-of-fit with a penalty for each additional parameter, separately for each model, participant, and serial position. In all cases, average AICc was smallest for the three-component model with serial position non-target responding weights (Fig. 10).

When we looked at the individual participants separately (collapsed over target serial position), we found that the three-component model and the variable precision model were both the most likely for 7 of 14 participants in the set size three condition. For set size five, however, the three-component model was most likely for 12 of 14 participants, whereas the

variable precision model was most likely for 2 participants. Overall, these results suggest that the three-component model fit the data better than the classic two-component model and the variable precision model.

Discussion

We analyzed the data from Experiment 3 with a variable precision model (Fougnie et al., 2012) and again saw that storage probability was uniform over serial positions at the smaller set size, but showed a hint of a primacy effect and a robust recency effect at the larger set size. Importantly, we observed this pattern of results even after modeling precision in a more sophisticated fashion, providing stronger evidence that the capacity of sequential VSTM primarily affects storage probability.

Even though we found that average precision varied in a quadratic fashion over serial position for the smaller set size, this effect is hard to interpret as we would have expected to find parallel results for the larger set size. Furthermore, the lack of a serial position effect on the variability in precision makes it difficult to draw strong conclusions about how precision was affected in our task.

General discussion

Even though VSTM is often proposed as a glue that bridges across disruptions in visual input, much of our understanding of VSTM has come from static displays. The goal of these experiments was to characterize how different components of VSTM operate over more dynamic, sequential input. In Experiment 1, for larger set sizes, we demonstrated a higher probability of storage for items in later serial positions. In Experiment 2, this recency effect was shown to not depend on the exact length of the retention interval. Finally, in Experiment 3, we found a greater likelihood of swapping out items in earlier serial positions and swapping in items from later serial positions. Our results largely replicated when we fit a variable precision model, with target serial position impacting target storage probability and not precision or variability in precision. As a boundary condition, there was little effect of serial position when the set size was small and likely within capacity.

Recency effects

This collection of findings suggests that VSTM is updated in a FIFO fashion. When the number of items is below capacity, the system stores all items with equal probability. However, when the size of the memory array exceeds capacity, the contents of VSTM are drastically affected and a robust

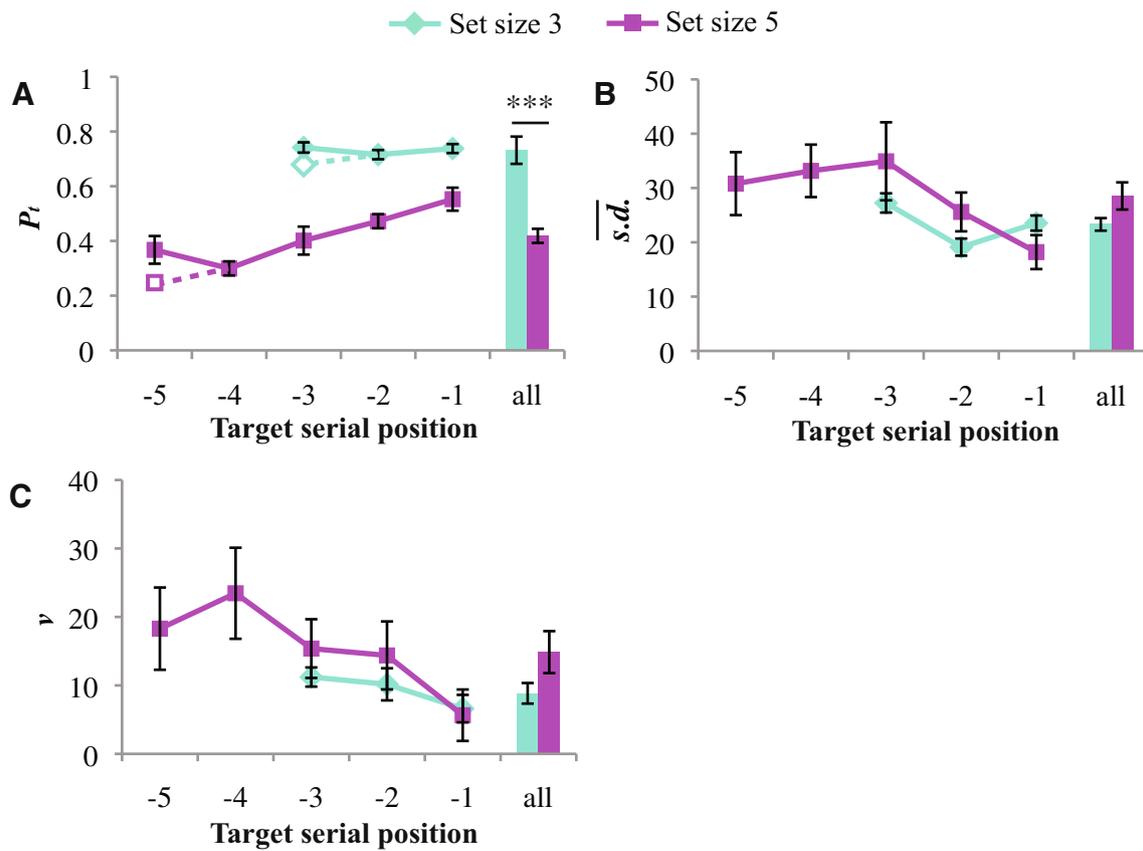


Fig. 9 a–c Results from fitting the variable precision model to the data of Experiment 3. **a** The probability of storing targets, P_t , showed a recency effect for set size five. **b** Mean precision estimates, $\overline{s.d.}$, showed a significant quadratic trend over the target serial positions for set size three, but not set size five. **c** The variability of precision, v , did not reliably

vary over target serial position. Mean estimates for each set size are depicted in the *bar graph* on the right. Error bars indicate within-subject SEM, except for the *bar graph*, which depicts across-subject SEM for each set size condition

recency effect emerges. Specifically, earlier items are discarded in favor of more recent items, resulting in less accurate subsequent memory. These inaccurate reports about earlier items do not seem to reflect random guessing, but rather erroneous recall of later items. The rate of non-target responding in this task might also explain the seemingly

puzzling finding that even though the recency effect manifests at set sizes larger than three (at the established 3–4 item limit in VSTM), a multiplication of the average storage probabilities with set size suggest an item capacity of approximately two. That is, this parameter may not directly capture the limit of VSTM due to retroactive interference at test.

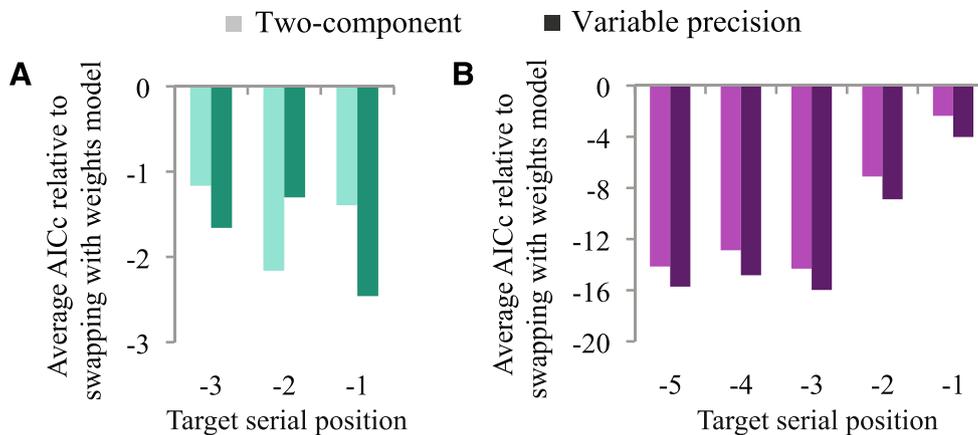


Fig. 10 a, b Average AICc fit measures of the standard two-component model and the variable precision model, relative to the three-component model with non-target serial position weights, separately for each serial position in the set size **a** three and **b** five conditions of Experiment 3

Why is it adaptive to have such a strong recency effect in VSTM? One obvious possibility is that the most recent information is most relevant in bridging seamlessly across changes in visual input (Henderson & Hollingworth, 1999; Irwin, 1991). However, a bias towards more recent information may not always be appropriate, especially if forgetting can occur within 200 ms. Indeed, it remains to be seen whether the FIFO nature of storage reflects a hard-wired consequence of how VSTM operates, or whether it can adapt more flexibly to changing demands. Indeed, perception and attention are highly sensitive to regularities in the environment (Chalk, Seitz, & Series, 2010; Zhao, Al-Aidroos, & Turk-Browne, 2013). In the case of VSTM, performance can be affected by spatial and relational structure (Brady & Alvarez, 2011; Brady, Konkle, & Alvarez, 2009; Umemoto, Scolari, Vogel, & Awh, 2010). If temporal structure was imposed on a sequential VSTM task, such as by probing certain serial positions more frequently than others, this might affect the baseline recency bias we observed in the absence of such structure.

Nature of capacity limits

We interpret the FIFO response profile as reflecting the capacity limit of sequential VSTM. However, as mentioned earlier, the form of this capacity limit remains hotly debated (Bays & Husain, 2008; van den Berg et al., 2012; Zhang & Luck, 2008). Our goal in this study was not to adjudicate between these accounts. However, to acknowledge these standing uncertainties, we fit our data with the three most influential models in the field. In doing so, we found a consistent effect of serial position and set size on storage probability, and less consistent evidence for such effects on precision. These findings are at odds with some of the most recent claims in the field about spatial VSTM (Fougnie et al., 2012; van den Berg et al., 2012), whereby capacity is thought to consist of a continuous resource that can be flexibly allocated to all items in a scene. In our study, this account predicts that precision would be gradually allocated away from the earliest items in order to store the more recent items, but we did not find positive evidence for this.

However, we hesitate to interpret the compatibility of our results and these claims too strongly at this stage, given that existing models were developed for a different task. As always, these models are only as good as the assumptions on which they are based, and various parameters may need to be tuned to incorporate sequential input. It remains fully possible in our minds that the FIFO effect is caused by a limited, continuous, flexible resource, which is simply not captured properly by these models. If so, our results provide empirical constraints for the development of new, more generalized models, by requiring an updating mechanism that resembles the FIFO pattern we observe here, deployed uniformly at

small set sizes but preferentially allocated to newer items at larger set sizes.

Relation to prior work

A recent study with a similar sequential recall task by Gorgoraptis and colleagues (2011) found increased precision for later items for all set sizes larger than one using the three-component model (but without serial non-target weights). They also reported recency effects in storage probability and decreasing rates of non-target responding for the latest items, but Gorgoraptis and colleagues (2011) observed these effects even for small set sizes (1–3). That is, our results were more selective than theirs, having found dissociations between storage probability and precision, and between small and large set sizes. There are several differences between the studies that could in principle account for these discrepancies.

First, the items used by Gorgoraptis and colleagues (2011) were oriented lines, in contrast to the colored squares we used. This stimulus difference may have led to differences in capacity and/or perceptual interference during the sequence (e.g., apparent rotation of the lines over time).

Second, their stimuli were presented for 500 ms each, in contrast to the 50-ms duration we employed. As mentioned earlier, we chose this duration as a compromise to stay as close as possible to Zhang and Luck (2008). It is important to note that the duration in Gorgoraptis and colleagues (2011) introduced substantially different maintenance requirements for each of the different serial positions (up to 6 s longer than in our design), since they did not alter the length of the retention interval depending on the set size. These different maintenance requirements may have impacted the response profile, reflecting both sequential storage dynamics and decay of the VSTM representations over time.

Third, Gorgoraptis and colleagues (2011) fit their models to data that was pooled across nine participants (25 trials per participant, serial position, and set size), whereas we fit the model separately for each participant (100–150 trials per serial position and set size). This controlled more appropriately for individual differences and used ~7–10 times more data per set size and serial position.

Fourth, arguably the most important difference is the way in which trials with different set sizes were presented. In our experiments, set size was manipulated between groups of participants, whereas they presented all serial positions of all set sizes in an interleaved manner. This introduces uncertainty in the expected set size on each trial with a bias towards larger set sizes (specifically, the expected set size was 3 2/3). Participants might have strategically and proactively weighed down items early in the sequence, giving rise to temporal performance curves that reflect both this strategic adjustment *and* the actual capacity limits of VSTM. This

potentially explains why Gorgoraptis and colleagues (2011) found recency effects on P_i even at very small set sizes (2–3).

Isolating and evaluating the importance of each of these factors to our results requires an extensive series of experiments that is beyond the scope of the current paper. Nevertheless, this is an important opportunity for future research, since it is conceivable that some aspects of our results are only revealed under the present stimulus conditions. For example, it would be useful to know how the results of this task are affected by varying stimulus durations or directly comparing within- and between-subject manipulations of set size.

Primacy effects

A priori, we could have found a pattern of results consistent with a first-in-*last*-out mechanism, where the earlier items would be prioritized over the later items. Although we saw little evidence for an overall trend of this sort, we did observe a bonus for the first serial position. This is consistent with a small primacy effect, which seems to accompany the more robust recency effect that dominates at all other serial positions.

Interestingly, the WM literature also often obtains a mixture of primacy and recency effects (Baddeley & Hitch, 1993; Murdock, 1962). These two types of effects may depend on separate systems in WM (Glanzer & Cunitz, 1966): primacy may be driven by a long-term component, which is biased toward earlier items because more time can be devoted to their representations; recency may be driven by a short-term component, which is biased toward later items because its representations are more fleeting. Although an appealing account for WM, it is unclear how this might apply to VSTM tasks. Remember that items were on screen for only 50 ms each, and that all colors were presented within 250 ms (often shorter). It is difficult to imagine a differential contribution of long-term storage to items separated by such short intervals.

An alternative explanation for the primacy effect in our task is that the item in the first serial position was unique in not being preceded by any other item. This could lead to a benefit for perceptual reasons, such as that the first item was not forward-masked. There are also potential attentional reasons, such as that later items may have suffered from a central bottleneck during the encoding and consolidation of the first item into VSTM, something akin to the attentional blink (Chun & Potter, 1995; Raymond, Shapiro, & Arnell, 1992).

Conclusions and future directions

These investigations were motivated by a desire to start bringing VSTM research into closer alignment with classical WM research. In particular, we sought to apply cutting-edge tools from the study of VSTM to behavioral paradigms that more

closely match the focus of WM research on sequential behavior. In doing so, we revealed robust and striking temporal dynamics over a very rapid timescale. These dynamics may also be at play, but harder to characterize, during more conventional static VSTM tasks, and they may provide a basis for further characterizing the neural substrates of VSTM. For example, electroencephalography could be used to study how event-related potentials that index the number of items in VSTM (e.g., contra-lateral delay activity, Vogel & Machizawa, 2004) evolve over time for longer sequences of single items (see also Ikkai, McCollough, & Vogel, 2010; Vogel, McCollough, & Machizawa, 2005). Furthermore, individual differences in the temporal dynamics, such as the slope of the recency effect, may provide new leverage for understanding the relationship between VSTM and intelligence (Fukuda, Vogel, et al., 2010; Shipstead et al., 2012), especially if these dynamics overlap with measures of WM typically linked to higher-level cognition.

Acknowledgments We thank Paul Bays, Daryl Fougny, Sam Gershman, and Jeremy Wolfe for helpful conversations. This work was supported by NIH grant R01 EY021755 (to N.T.B.).

References

- Alvarez, G., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111.
- Awh, E., Barton, B., & Vogel, E. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*, 622–628.
- Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839.
- Baddeley, A. D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? *Memory & Cognition*, *21*, 146–155.
- Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*, 7.
- Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*, *11*, 6.
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854.
- Bays, P. M., & Husain, M. (2009). Response to comment on “Dynamic shifts of limited working memory resources in human vision”. *Science*, *323*, 877d.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*, 384–392.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*, 487–502.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*, *5*, 49–62.

- Bunting, M. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 183–196.
- Burgess, G. C., Gray, J. R., Conway, A. R., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General*, *140*, 674–692.
- Buschman, T. J., Siegel, M., Roy, J. E., & Miller, E. K. (2011). Neural substrates of cognitive capacity limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 11252–11255.
- Chalk, M., Seitz, A., & Series, P. (2010). Rapidly learned expectations alter perception of motion. *Journal of Vision*, *10*. Article 2.
- Chun, M., & Potter, M. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 109–127.
- Conway, A., Kane, M., Bunting, M., Hambrick, D., Wilhelm, O., & Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786.
- Courtney, S., Ungerleider, L., Keil, K., & Haxby, J. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature*, *386*, 608–611.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114.
- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.
- D'Esposito, M., Postle, B., Jonides, J., & Smith, E. E. (1999). The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 7514–7519.
- Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, *10*, 27.
- Fougnie, D., Suchow, J., & Alvarez, G. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229.
- Frick, R. (1988). Issues of representation and limited capacity in the visuospatial sketchpad. *British Journal of Psychology*, *79*, 289–308.
- Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*, 177–182.
- Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*, 673–679.
- Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *5*, 351–360.
- Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *Journal of Neuroscience*, *31*, 8502–8511.
- Gray, J., Chabris, C., & Braver, T. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*, 316–322.
- Harrison, S., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*, 632–635.
- Hay, D., Smyth, M., Hitch, G., & Horton, N. (2007). Serial position effects in short-term visual memory: A SIMPLE explanation? *Memory & Cognition*, *35*, 176–190.
- Henderson, J., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, *10*, 438–443.
- Ikkai, A., McCollough, A. W., & Vogel, E. K. (2010). Contralateral delay activity provides a neural measure of the number of representations in visual working memory. *Journal of Neurophysiology*, *103*, 1963–1967.
- Irwin, D. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*, 420–456.
- Johnson, A., & Miles, C. (2009). Serial position effects in 2-alternative forced choice recognition: Functional equivalence across visual and auditory modalities. *Memory*, *17*, 84–91.
- Kane, M., & Engle, R. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 336–358.
- Kane, M., & Engle, R. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *2002*, 637–671.
- Kumar, A., & Jiang, Y. (2005). Visual short-term memory for sequential arrays. *Memory & Cognition*, *33*, 488–498.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–280.
- Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*, 199–207.
- Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology: Journal of Experimental Psychology*, *64*, 482–488.
- Myung, I. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90–100.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Pernet, C., Wilcox, R., & Rousselet, G. (2012). Robust correlation analyses: False positive and power validation using a new open source Matlab toolbox. *Frontiers in Psychology*, *3*, 606.
- Phillips, W., & Christie, D. (1977). Components of visual memory. *The Quarterly Journal of Experimental Psychology*, *29*, 117–133.
- Postle, B., Brush, L., & Nick, A. (2004). Prefrontal cortex and the mediation of proactive interference in working memory. *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 600–608.
- Raymond, J., Shapiro, K., & Arnell, K. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink. *Journal of Experimental Psychology: General*, *18*, 849–860.
- Serences, J., Ester, E., Vogel, E., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*, 207–214.
- Shipstead, Z., Redick, T. S., Hicks, K. L., & Engle, R. W. (2012). The scope and control of attention as separate aspects of working memory. *Memory*, *20*, 608–628.
- Smyth, M., Hay, D., Hitch, G. J., & Horton, N. J. (2005). Serial position memory in the visual—spatial domain: Reconstructing sequences of unfamiliar faces. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58A*, 909–930.
- Smyth, M., & Scholey, K. A. (1996). Serial order in spatial immediate memory. *The Quarterly Journal of Experimental Psychology*, *49*, 159–177.
- Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*, 9.
- Todd, J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*, 751–754.
- Umemoto, A., Scolari, M., Vogel, E. K., & Awh, E. (2010). Statistical learning induces discrete shifts in the allocation of working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1419–1429.
- Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, *76*, 559–573.
- van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*, 8780–8785.
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.

- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500–503.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*, 1120–1135.
- Woodman, G., Vogel, E., & Luck, S. (2012). Flexibility in visual working memory: Accurate change detection in the face of irrelevant variations in position. *Visual Cognition*, *20*, 1–28.
- Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91–95.
- Zhang, W., & Luck, S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235.
- Zhang, W., & Luck, S. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*, 423–428.
- Zhao, J., Al-Aidroos, N., & Turk-Browne, N. (2013). Attention is spontaneously biased towards regularities. *Psychological Science*, *24*, 667–677.