Emrich, S.M. [1], Lockhart, H.A. [1] & Al-Aidroos, N. [2] (*in press*). Attention Mediates the Flexible Allocation of Visual Working Memory Resources. *Journal of Experimental Psychology: Human Perception and Performance. DOI: 10.1037/xhp0000398*

1. Department of Psychology, Brock University
St Catharines, ON, Canada

2. Department of Psychology, University of Guelph

Guelph, ON, Canada

Correspondence should be addressed to:

semrich@brocku.ca
Stephen Emrich
Department of Psychology
Brock University
1812 Sir Isaac Brock Way
St Catharines, Ontario
L2S 3A1
Canada

Abstract

Though it is clear that it is impossible to store an unlimited amount of information in visual working memory (VWM), the limiting mechanisms remain elusive. While several models of VWM limitations exist, these typically characterize changes in performance as a function of the number of to-be-remembered items. Here, we examine whether changes in spatial attention could better account for VWM performance, independent of load. Across two experiments, performance was better predicted by the prioritization of memory items (i.e., attention) than by the number of items to be remembered (i.e., memory load). This relationship followed a power law, and held regardless of whether performance was assessed based on overall precision or any of three measures in a mixture model. Moreover, at large set sizes, even minimally attended items could receive a small proportion of resources, without any evidence for a discrete-capacity on the number of items that could be maintained in VWM. Finally, the observed data were best fit by a variable-precision model in which response error was related to the proportion of resources allocated to each item, consistent with a model of VWM in which performance is determined by the continuous allocation of attentional resources during encoding.

Public Significance Statement

Visual working memory supports the maintenance of visual information "on-line" for short periods of time, and is known to be able to store a limited amount of information. The present study examined whether the distribution of spatial attention across items in a visual scene could account for changes in visual working memory performance independent of the number of items that had to be maintained. Several analyses from two experiments revealed that performance was best explained by the proportion of attentional resources allocated to each item, rather than by the number of items that had to be remembered. These findings inform our understanding of how the brain maintains visual information on-line to support everyday behaviors, suggesting that the amount of attention allocated to an item is a limiting factor in how accurately that item can be stored in, and recalled from, visual working memory.

The ability to hold visual information in mind for short periods of time is critical to performing numerous everyday behaviors (e.g., remembering the number and location of cars in your mirror while driving), and has been linked to important aspects of cognition, including fluid intelligence (Unsworth, Fukuda, Awh, & Vogel, 2014). Despite the utility of this cognitive ability, one of the definitive attributes of visual working memory (VWM) is its severely limited capacity; only a small number of high-fidelity (i.e., high-resolution) representations can be maintained in VWM (Cowan, 2001; Luck & Vogel, 2013; Ma, Husain, & Bays, 2014); as the demands for VWM storage increase, there is an associated decrease in the number and/or fidelity of representations maintained.

The exact mechanism of information loss in VWM remains elusive. The two predominant theories in this debate are the discrete-capacity (Luck & Vogel, 2013; Zhang & Luck, 2008) and continuous-resource (Bays & Husain, 2008; Ma et al., 2014) models of VWM capacity, although variants on these theories exist (Fougnie, Suchow, & Alvarez, 2012; van den Berg, Awh, & Ma, 2014). According to the discrete-capacity model, VWM can store a fixed number of visual objects, and once this number is exceeded all items not encoded in these few storage “slots” are forgotten (Luck & Vogel, 1997; Zhang & Luck, 2008). Further, changes in memory load can affect fidelity, but only when the total number of remembered items is below capacity, possibly reflecting the sharing of resources among slots when memory loads are low (Machizawa, Goh, & Driver, 2012; Zhang & Luck, 2008). Changes in load above capacity should generally not affect memory fidelity.

By contrast, continuous-resource models posit that there is no upper limit on the number of items that can be maintained in VWM; rather, capacity is constrained by a limited pool of resources that must be allocated across all maintained items; increasing the number of items entails that each will receive a smaller proportion of the available resources, resulting in a proportional loss in representational fidelity (Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Wilken & Ma, 2004). As evidence of this continuous allocation of a shared resource, Bays and Husain (2008) demonstrated that VWM error (inverse of fidelity) varies with memory load according to a simple power law, consistent with the predicted change in noise of a neural population code (Bays, 2014).

Most studies examining models of VWM place significant emphasis on the relationship between *memory load*, as defined by the number (and occasionally complexity) of the physical stimuli to be remembered, and memory performance. While this approach has been very informative, it overlooks the potential flexibility of this memory resource. For example, many studies have demonstrated that VWM performance is sensitive to eye-movements (Bays & Husain, 2008), presentation order (Gorgoraptis, Catalao, Bays, & Husain, 2011; Zokaei, Gorgoraptis, Bahrami, Bays, & Husain, 2011), attentional cues (Zhang & Luck, 2008; Zokaei et al., 2011), attentional lapses (Fougnie et al., 2012), reward (Klyszejko, Rahmati, & Curtis, 2014), and even voluntarily shifts in performance (Machizawa et al., 2012). Thus, VWM performance cannot be fully explained by item load alone, but must also account for situations where resources are distributed unevenly across remembered items. However, most studies have tended to assume that memory resources are distributed evenly, potentially failing to capture the variability among items (Fougnie et al., 2012; van den Berg, Shin, Chou, George, & Ma, 2012).

One potential source of variation in VWM is attention. It has long been known that when trying to remember an array of letters, cueing attention to a subset of items in the array improves performance for those items (Sperling, 1960). Thus, restricting the scope of spatial attention to fewer items reduces the demand on resources and improves performance. Notably, the loss of encoding accuracy that occurs when attention must be distributed across multiple items occurs even in the absence of spatial attentional cues (Duncan, 1980). Consequently, when multiple items are to be remembered, it is potentially the distribution of attention that limits performance, regardless of whether explicit attentional cues are present or not; the number of available visual items to be remembered is simply a confounding variable over which participants must allocate attentional resources. Consistent with this account, several recent variable-precision models incorporate the allocation of attention during encoding into models of VWM capacity (Fougnie et al., 2012; van den Berg et al., 2012), suggesting that the allocation of attention, rather than storage limitations, may play a critical role in limiting VWM performance.

While the attentional mediation of VWM resources is a prediction of most continuous-resource models, there are also two variants of discrete-capacity models that predict comparable roles for attention, with some key differences (Machizawa et al., 2012; Zhang & Luck, 2008, 2011). First, according to the slots+averaging model, the precision associated with each memory slot is fixed, however memory precision can be improved by storing individual items across multiple slots. One prediction that separates this model from continuous-resource models is that the least amount of resources that can be allocated to a single item is determined by the amount of resources present in a single storage slot. Second, according to the slots+resources model, there are no restrictions on how resources are allocated across memory items; the main difference that separates this model from continuous-resource models is that resources can only be allocated across a fixed number of objects. Accordingly, when the number of items to be remembered exceeds this item limit, the slots+resources model predicts that extra-capacity items are forgotten, whereas continuous-resource models predict all items are remembered, but the precision of each item is proportional to the amount of resources allocated to it. Thus, continuous-resource models and these two hybrid discrete-capacity models differ primarily in how they predict load and attention can affect the precision and likelihood of items being stored in memory. Specifically, while fixed-capacity models posit that the effect of attention on performance should be constrained by the memory load, as well as by whether or not items can receive less than one "slot" worth of resources, continuous-resource models (in particular variable-precision models) impose no such constraints on the effect of attention. However, while numerous studies have examined the effect of load on VWM performance, fewer studies have examined the effects of attentional allocation.

In the present study, we examined how systematically varying the distribution of attention across a fixed set of stimuli influences the way those items are encoded in VWM. Participants saw memory arrays containing six (Experiment

1a), or four or one (Experiment 1b) colored squares and, after a brief delay, were required to report the color of one item using a continuous response (Wilken & Ma, 2004). To evaluate the role of attention, we presented predictive spatial cues during the memory array, and systematically varied both the number and predictive validity of the cues. We specifically tested the prediction that if VWM resources are flexibly allocated via attention, then the proportion of resources allocated to each item should be best described by a power law that follows the predictive value of each cue, independent of the load (i.e., the number of items with a greater than zero percent chance of being probed). We also examined the specific predictions of the continuous-resource and discrete-capacity models, to determine whether this effect was limited to a fixed number of items. Finally, we compared the fit of different memory models to our observed data, to determine how well the results could be explained by fixed-capacity as opposed to continuous-resource models.

Method

Participants

A total of 43 participants (20 and 23 in Experiments 1a and 1b, respectively) ages 18 – 29 participated in this study. Participants were recruited from Brock University and were either paid by honorarium 10$/hour or given 2 hours of course credit. All participants were screened for normal color vision using the Ishihara Test of Color Vision, with one participant excluded from Experiment 1b for not meeting normal criteria. Two additional participants did not complete the task and their data was not analyzed, resulting in 20 participants per experiment. All procedures were approved by the Research Ethics Board of Brock University.

Apparatus

Stimuli were presented using PsychoPy (Peirce, 2007) on a 20” LCD display at a distance of ~ 57 cm.

Procedure and Stimuli

Both Experiments (1a and 1b) used a similar task with the exception of the number and predictive validity of the memory items and cues (see Table 1).

In Experiment 1a, six equally spaced colored squares (1° x 1°) were presented for 500 ms centered around a central fixation dot (0.3° in diameter), along with 1 – 6 spatial line cues. The line cues indicated which of the six sample items were likely to be probed at the end of the trial. The predictive validity of these cues varied between 33 – 100%, and was indicated to the participant at the beginning of each block. Seven conditions were presented (see Table 1): one cue – 100% valid; one cue – 50% valid; one cue – 33% valid; two cues – 100% valid; two cues – 66% valid; three cues – 100% valid; six cues – 100% valid. Consequently, in some of the conditions the uncued items remained relevant to the participant, but were less likely to be probed. For example, in the two cues – 66% condition of Experiment 1a, each of the two cued items had a 33% likelihood of being probed, and each of the four uncued items had an 8.25% likelihood of being probed.

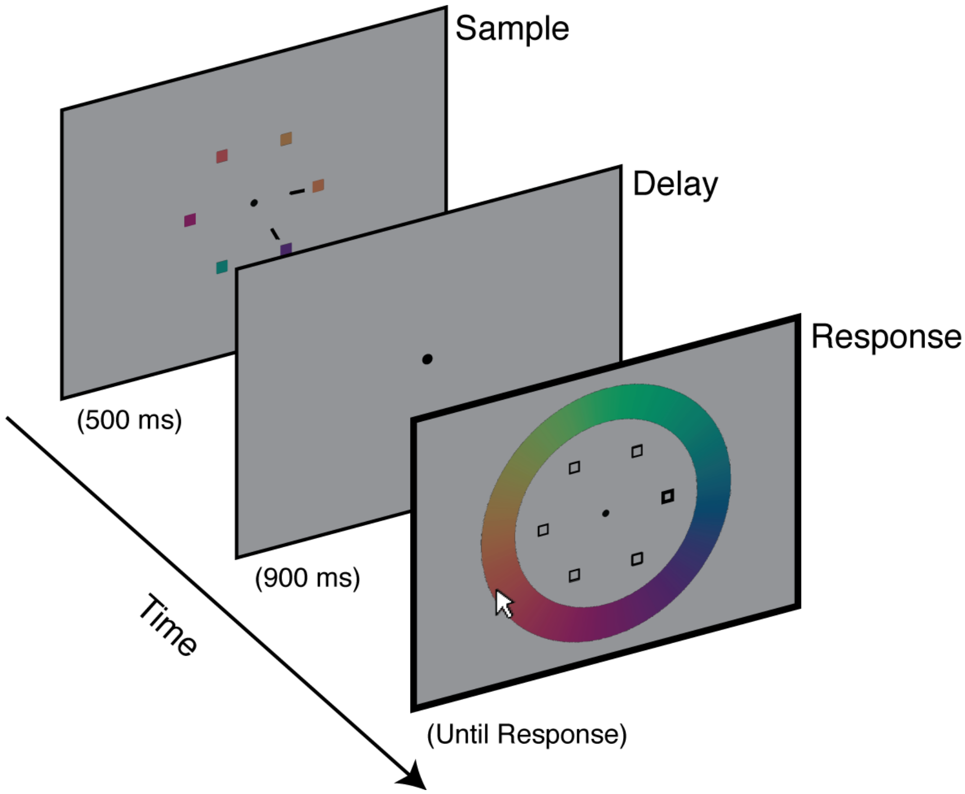

Figure 1 -. A schematic of experimental trials from Experiment 1a. Participants were instructed to report the color of the probed item (as indicated by the bold outline). The number of items to be remember was determined by the validity of the cues, which was indicated to participants prior to each block. Experiment 1b used total set sizes of 4 or 1, otherwise the procedures were identical.

The sample display was followed by a 900 ms delay in which only the fixation cross remained. Following the delay, a black box outline of all six sample items was presented, with one square presented with a bold outline to indicate the probed position (see Figure 1). In addition, a color wheel (12° radius) containing all potential color values was presented. The probed item was selected at random from the relevant items in accordance with the validity of the cues presented during the sample. As participants moved the mouse around the display, the color of the probed item was updated to reflect the position on the color wheel closest to the mouse location. Participants were instructed to press the mouse button when the color indexed by the probed location matched as closely as they could remember to the sample item presented in that location.

Participants performed a total of 1,200 trials in blocks of 50, with optional breaks in between. The number of trials per condition was balanced so that 100 valid trials were completed per condition (i.e., 100 trials in the 100% valid conditions, 150 trials in the 66% valid condition, 200 trials in the 50% valid conditions, and 300 trials in the 33% valid condition).

The aim of Experiment 1b was to replicate the Experiment 1a results with a smaller set size—four items—and with uncued probabilities that were equivalent or similar to cued probabilities. Experiment 1b also included a one-item set size condition to establish a baseline for maximal task performance. Seven total conditions were used: one cue – 100% valid; one cue – 33% valid; two cues – 100% valid; two cues – 66% valid; three cues – 100% valid; four cues 100% – valid; and one item, one cue – 100% valid. Again, 100 valid trials per condition were used, with a total of 1,100 trials.

Sample colors for both experiments were selected randomly from one of 360 unique colors obtained from a circular wheel on the CIE L*a*b* color space with coordinates of a = - 6 and b = 14 with a radius of 49, calibrated to the monitor.

Prior to performing the experimental task, participants also performed a standard change-detection task to estimate VWM capacity. These data were not analyzed for the purposes of the current study, and as such the results are not presented here.

Analysis

Response Error

We evaluated memory fidelity using numerous measures all based on raw response error (i.e., the angular distance in degrees between the actual color of the stimulus and the probed color). First, the amount of variance in participants' responses was assessed using the circular standard deviation of the response error ($SD_{response}$), and we calculated Precision as the inverse of response error (i.e., Precision = $1/SD_{response}$). Relative Precision was also calculated from both experiments using Relative Precision = Precision/Max Precision, where Max Precision was defined as precision obtained in the one item, one cue – 100% condition from Experiment 1b. Lower error values indicate higher fidelity working memory representations, and relative precision estimates the proportion of available resources allocated to the probed item.

Mixture Model Analysis

In order to test specific predictions of the fixed-capacity and continuous-resource models, raw error was also decomposed into a three-component mixture model describing different aspects of performance (Bays et al., 2009) using Maximum Likelihood Estimation (MLE). The mixture model is described as:

$$p(\hat{\theta}) = (1 - \gamma - \beta)\phi_\sigma(\hat{\theta} - \theta) + \gamma\frac{1}{2\pi} + \beta\frac{1}{m}\sum_{i}^{m}\phi_\sigma\left(\hat{\theta} - \theta_i^*\right)$$

where θ is the target color, $\hat{\theta}$ is the reported color, and the probability of a given response, $p(\hat{\theta})$, is determined by three distributions: the proportion of target responses with a normal circular (Von Mises) distribution of a given standard deviation (ϕ_σ); the proportion of non-target errors, β, which are responses centered around *m* non-probed items; and the proportion of guesses, γ, which is a uniform distribution. The circular SD of the mixture model, the probability of non-target responses, and the probability of guessing are referred to throughout as SD_{MM}, NT_P and G_P, respectively. MLE was performed using MATLAB and the MemToolBox (Suchow, Brady, Fougnie, & Alvarez, 2013).

Curve Fitting

To test the prediction that measures of behavior should follow a power law determined by the proportion of attentional resources allocated to each item, the measures of response error ($SD_{response}$), as well as each of the measures obtained in the mixture model were fit to a power law function:

$$y \propto ax^k,$$

where *x* is either the load or probe likelihood, and *y* is the measure of behavior ($SD_{response}$, SD_{MM}, NT_P, or G_P), k is the power constant, and *a* is a constant.

Curve fits were computed in MATLAB using nonlinear least squares regression with bisquare robust fitting across the group averaged data from each cued condition. Bisquares robust fitting minimizes the effect of outliers, although similar results were obtained without this method. For each fit type, 95%

confidence intervals were obtained, and data from invalidly cued items were used as validation data (see Supplementary Table 1).

Comparison between curve fits was performed via root mean-squared error (RMSE), as well as through the Akaike Information Criterion (*AIC*), which was estimated using SSE as

$$AIC = n \log\left(\frac{SSE}{n}\right) + 2(p+1), \tag{1}$$

where n is the number of data points and *p* is the number of free parameters estimated using the least squares method (Burnham & Anderson, 2002). *AIC* includes a penalty for the number of free parameters in the model, although assumes infinite data. The corrected *AIC*, AIC_c, is calculated as

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}. \tag{2}$$

Model Comparison

In order to directly compare whether fixed-capacity or continuous-resource models could best describe our data, we used the method of van den Berg et al (van den Berg et al., 2014, 2012) to estimate the parameters for four distinct models (see van den Berg et al., 2014 for a full description of the methods and model formulae). As with the mixture model described above, these models all assume responses follow a Von Mises distribution that has a concentration parameter of κ, and that the report of the target is corrupted by Von Mises-distributed noise. *J* is the measure of precision, which is inversely related to κ.

For fixed-capacity models, the maximum number of items that can be stored, K, is limited. We tested two specific variants of the fixed-capacity models described by Zhang and Luck (2008): the slots+averaging model, in which the lowest amount of precision that an item can be stored with is equivalent to that of one “slot”; and the slots+resource model, in which precision is equivalent for all items stored and follows a power-law relationship N, $J = J_1 N^{\alpha}$, where N is the number of items remembered and J_1 is the precision of a single item (i.e., $N = 1$).

For continuous-resource models, we examined two models: first, an equal-precision model (Bays & Husain, 2008), which is similar to the slots+resource model described above, although without the capacity parameter K; in addition, we also examined a variable-precision model (van den Berg et al., 2012), in which the precision of remembered items can vary randomly across items and trials according to a gamma distribution with a mean of $\bar{J} = \bar{J}_1 N^{\alpha}$ and a scale parameter τ.

Importantly, we fit each of these four models to the raw error data in two ways: once where error was predicted by memory load, and once where it was predicted by probe likelihood. Specifically, first, we fit the models where N = the total number of to-be remembered items, and then, second, we fit the models where N = 1/probe likelihood (rounded to the nearest whole number). That is, for each model we assumed that the resources allocated to each item should be proportional to the probability of each item being tested. Thus, if performance were primarily constrained by the total number of to-be-remembered items, we would expect the models defined by load to outperform those defined by the proportion of allocated resources, independent of other aspects of the model (i.e., fixed-capacity, slot-averaging, equal/variable precision).

For each model and each subject, parameter estimates were obtained using maximum-likelihood estimation (code available at http://www.cns.nyu.edu/malab/resources.html) using an evolutionary algorithm. Following the procedure of van den Berg et al (2014), models were compared using the Akaike information criterion (AIC).

Additional statistics were computed using JASP (JASP Team, 2016).

Results

In both versions of the experiment, participants performed a continuous-response task in which they had to report the color of a sample item maintained in VWM. Critically, we manipulated the number and predictive validity of spatial cues presented during the sample display. Consequently, in addition to the typical analysis examining memory performance as a function of the item load (i.e., the number of to-be-remembered items), we also examined performance as a function of the of probe likelihood of each item (i.e., cue validity/number of cues). We predicted that if VWM resources were proportionally distributed across memory items according to the likelihood that each item would be probed, performance would vary primarily by the validity of the cues. For example, if 2 cues are presented with 100% validity, then each item only has a 50% chance of being probed, and attention should, on average, be divided equally across those two items; if VWM performance is primarily affected by the voluntary allocation of attention, performance on this condition should be similar to that of the single cue – 50% valid condition, since in each case the cued items will only be probed on 50% of trials (see Table 1 for a full list of cue validity and Supplementary Table 2 for a complete breakdown of the data).

To test this prediction, we first examined VWM performance using a measure of response error ($SD_{response}$). Drawing from the predictions of resource allocation models, we modelled VWM performance using a power-law function (Bays & Huasin, 2008), and tested whether cue validity or memory load was the better predictor. To examine whether cueing effects were subject to item limits, and to examine specific predictions made by the slots+averaging and slots+resource models, we subsequently performed additional analyses on the mixture model proposed by Zhang and Luck (2008) and modified by Bays et al. (2009). We also compared the fits of fixed-capacity and flexible-resource models to the raw response error, using both load and probe likelihood as predictors.

$SD_{response}$

Response error ($SD_{response}$ or 1/*precision*) is plotted relative to load (i.e., the number of items to be remembered; Figure 2A) and probe likelihood (i.e., cue validity/number of cues; Figure 2B). As is evident from the figures, the amount of error tended to increase as a function of load, and decrease as a function of probe likelihood. That is, as the number of to-be-remembered items increased, the amount of resources allocated to each item decreased, and error increased. Probe likelihood, however, affected $SD_{response}$ over and above memory load. For example, significant differences were observed between the 1 cue – 33% (M = 35.2) and the 6 cues - 100% (M = 51.1) conditions, $t(19) = 5.73$, $p < .001$, Cohen's d = 1.281, even though both conditions required all six items be remembered. By contrast, there was no observed difference between the 3 cues – 100% (M = 40.0) and 1 cue – 33% (M = 35.2) conditions (cued trials), $t(19) = 1.617$, $p = .122$, Cohen's d = .385, even though there was a large difference in the total item load (3 vs 6). Thus, the size of the effect of doubling probe likelihood while holding load constant was very large, while the effect of doubling load while holding probe likelihood constant was modest.

To assess whether response error is explained by the amount of resources allocated to each item (probe likelihood), we fit response error to a power-law model, consistent with the predictions of Bays and Husain (2008). Fitting a power law function where $SD_{response}$ varies as a function of probe likelihood revealed that performance was strongly predicted by the predictive validity of the cues (Figure 2B), explaining over 90% of the variance in performance, adjusted $R^2 = .902$, RMSE = 3.497. By comparison, fitting a power law in which $SD_{response}$ varies as a function of memory load (Figure 2A) was a much poorer fit to the data, adjusted $R^2 = .515$, RMSE = 7.781.

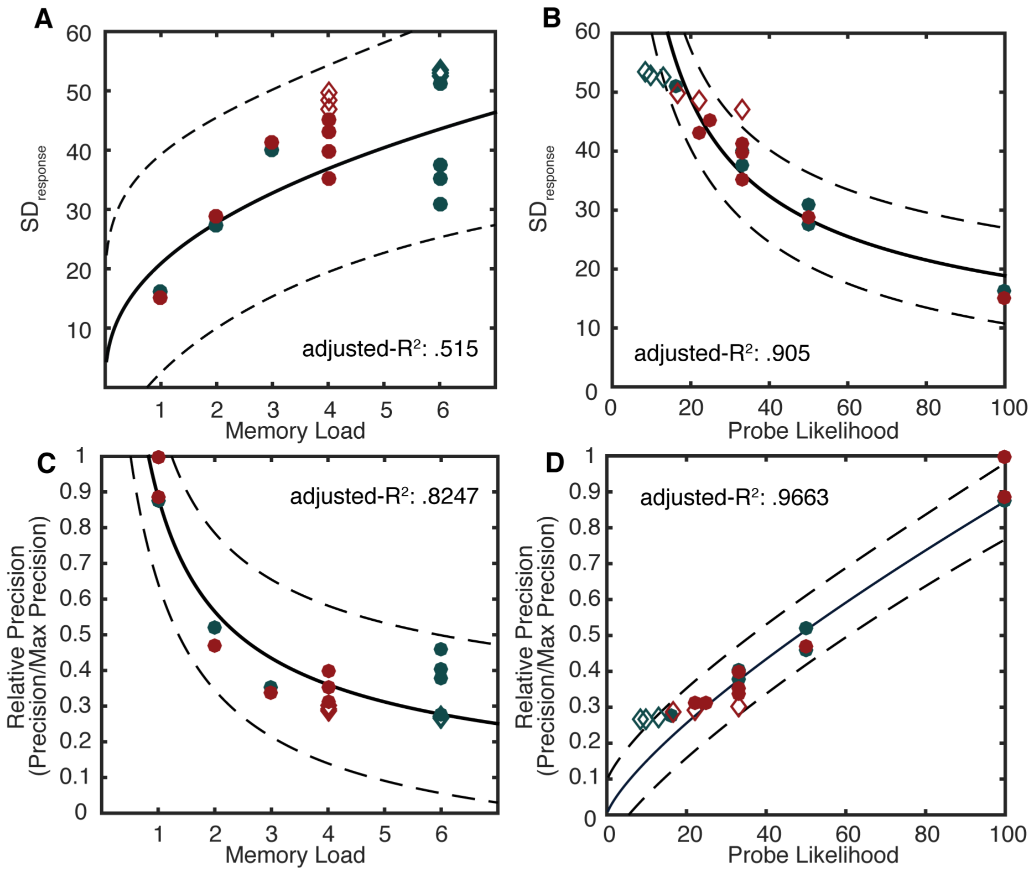

Figure 2 - Overall response error ($SD_{response}$) and relative precision, plotted by total number of to-be-remembered items (A,C) or the likelihood of an item being probed (B,D). Data are fit to a power law function as in Bays & Husain (2008). Green and red items represent data from Experiments 1a and 1b, respectively. Closed circles represent cued items, open diamonds represent uncued items (validation data). Dashed lines indicated 95% confidence intervals of the fitted line. Full means and SDs are available in Supplementary Table 2.

Moreover, the continuous-resource model of VWM predicts that performance should decrease according to the proportion of available resources that can be allotted to a given item (Bays & Husain, 2008). Accordingly, we also calculated precision as a proportion of available resources (*relative precision*), using performance from the one cue – 100% condition of Experiment 1b as a measure of maximum precision. Fitting a power law to relative precision as a function of probe likelihood provided a strong fit to the data (Figure 2D), explaining almost all variance in the cued trials, adjusted R^2 = .9663, RMSE = .0429. In contrast, modeling relative precision as a function of the number of items that were to-be-remembered (Figure 2C) resulted in a poorer fit, adjusted R^2 = .8247, RMSE = .0979. Comparing these two fits using AIC_c strongly favored the likelihood model (AIC_c = -89.6) over the model explaining performance by load (AIC_c = -64.861). Thus, overall and relative precision are both strongly predicted by a model in which VWM resources are proportionally distributed across all relevant items in the memory array based on the likelihood that each item can be probed.

Mixture Model Analysis

The analysis of overall response error ($SD_{response}$) suggests that participants can flexibly allocate VWM resources according to the predictive validity of spatial attention cues. However, it is possible that even if participants can distribute resources unevenly, this ability may be subject to item limits (Zhang & Luck, 2008). According to such fixed-capacity models, changes in the SD of reported targets from increases in memory load should asymptote at storage capacity (*K*), once the guess rate (i.e., random responses) is accounted for. Moreover, as outlined in the introduction, discrete-capacity models make specific predications about the effect of attentional cues on VWM resource allocation (as measured by precision), although previous studies examining these predictions have not accounted for the possibility that resource allocation is determined by the predictive validity of those cues.

Consequently, we further examined performance by breaking down responses using the three-component mixture model (Bays et al., 2009). Drawing from discrete-capacity models, the mixture model analysis provides a measure of response error (SD_{MM}) on those trials in which participants correctly reported the target, as well as an estimate of the proportion of trials on which subjects guess (G_p), and the proportion of trials in which subjects incorrectly report a non-target item (NT_p).

Mixture Model Error (SD_{MM})

The amount of response error on those trials in which participants correctly reported the target increased with both load and probe likelihood (Figure 3). One of the main predictions of the discrete-capacity model is that no changes in precision ($1/SD_{MM}$) should be observed once the number of items to be remembered exceeds capacity. Contrary to this prediction, we observed moderate evidence for decreases in precision (increased error) from conditions with 3 to 6 cues (100% cue validity, *Ms* = 21.5, 28.2, respectively) in Experiment 1a, $t(19) = 1.916$, $p = .035$ (one tailed), $d = .585$, as well as from conditions with 4 cues (Experiment 1b; $M = 21.4$) to 6 cues (100% cue validity, Experiment 1a; $M = 28.2$), $t(20.51) = 1.836$, $p = .040$ (one tailed), $d = 0.698$. These increases in response error suggest no fixed capacity on the number of items that can be stored in memory, and instead are more consistent with a continuous-resource model (see Model Comparison section below for a more complete test between models).

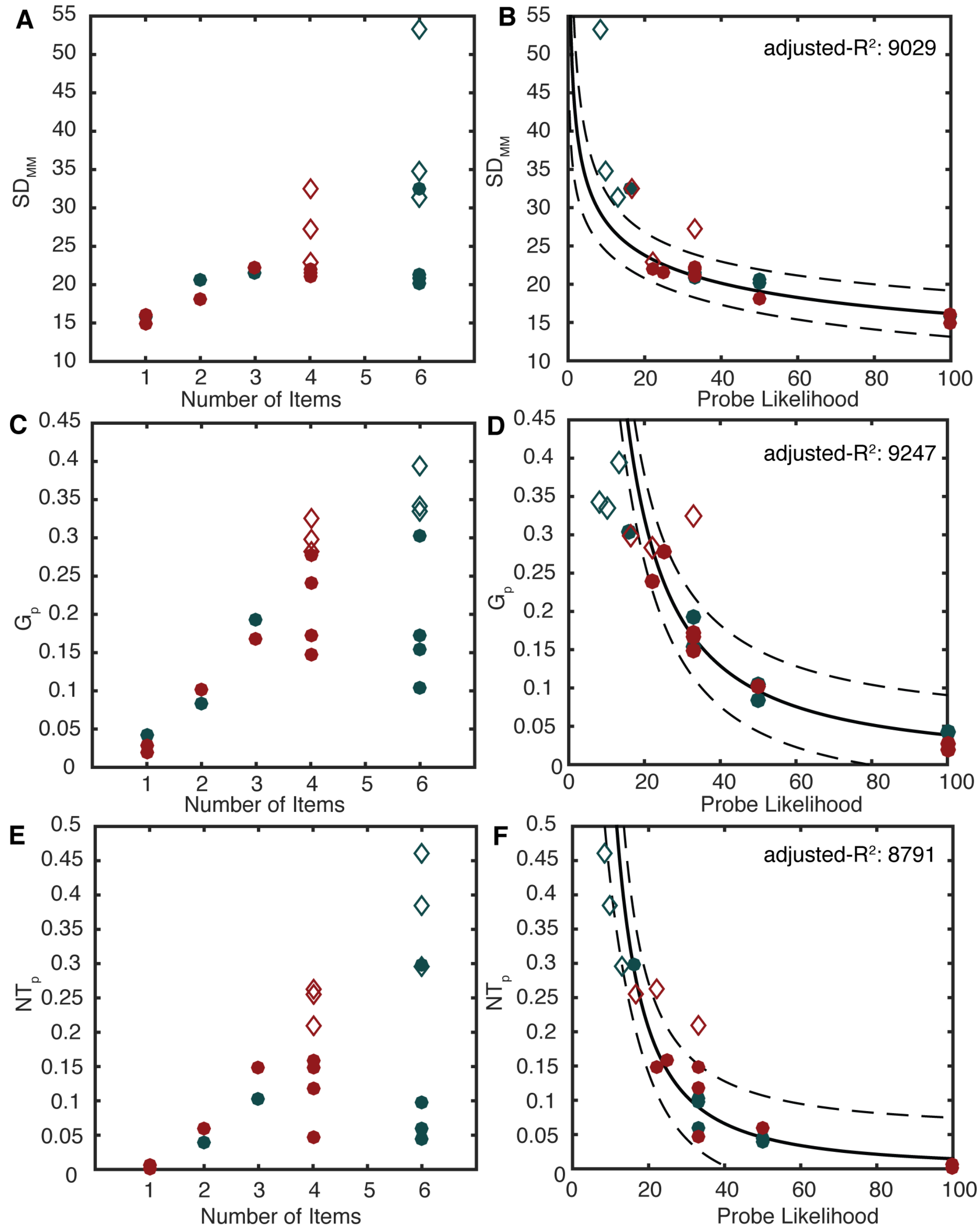

Figure 3 Error (SD_{MM}), guess rate (G_p) and non-target errors (NT_p) obtained from the three-component mixture model (Bays et al., 2009). (Left) Data are plotted by total number of items to-be-remembered. (Right) Plotting the data by the likelihood of an item being probed and fitting it to a power law function is a strong predictor across all three measures. Green and red items represent data from Experiments 1a and 1b, respectively. Closed circles represent cued items, open diamonds represent uncued items (validation data). Dashed lines indicated 95% confidence intervals of the fitted models. Full means and SDs are available in Supplementary Table 2.

A further prediction of the slots+averaging model (Zhang and Luck, 2008) is that items in memory should not be able to receive fewer resources than a single slot. To test this prediction, we used SD_{MM} on neutral-cued trials with three items (3 cues – 100%) as an estimate of the precision afforded by one slot, and tested for decreases in precision on uncued trials, in particular those trials where the likelihood of the item being probed was very small. The results of these comparisons are presented in Table 2. We observed evidence in favor of a decrease in precision for uncued trials relative to neutral trials, but only when the proportion of resources allocated to the uncued items (as determined by probe likelihood) was small. For example, in the 2 cues – 66% condition of Experiment 1a, each uncued item was probed only 8.25% of the time, compared to the 33% probe validity of the neutral condition (3 cues – 100%). This large difference in cue validity led to a significant decrease in precision, suggesting that the uncued items received just a small amount of available memory resources. Consequently, these findings suggest that the proportion of VWM resources allocated to uncued items is not limited to the resolution of a single slot, but rather is determined by the distribution of attention.

Following from the analysis of $SD_{response}$, we also examined whether SD_{MM} varied as a function of probe likelihood according to a power law. This analysis revealed that the power law function provided a close fit to the data when presented as a function of the predictive value of the cues, adjusted R^2 = .9029, RMSE = 1.264. Thus, the results suggest that within a mixture model, VWM precision is strongly predicted by the proportion of resources allocated to each item (as determined by the predictive validity of spatial attention cues), independent of the memory load.

Mixture Model Guess Rate (G_p)

As is evident in Figure 3D, although the proportion of guesses increased with the overall load, G_P appeared to be more closely related to probe likelihood. Indeed, following from the analyses of $SD_{response}$ and SD_{MM}, modeling the proportion of guesses as a power law function relative to the predictive value of the spatial cues demonstrated a close fit, Adjusted R^2 = .9247, RMSE = 0.02356.

Mixture Model Non-Target Errors (NT_p)

In addition to looking at both error and guess rate, the three-component model created by Bays et al. (2009) supports a parameter for reporting one of the non-target items, referred to sometimes as “swaps”. The probability of making non-target errors (NT_p) has been shown to increase with load (Bays et al., 2009), and increase as the spatial distance between items is reduced (Emrich & Ferber, 2012). Examining NT_p revealed that swap errors increase with load and cue validity (Figure 3E and F). Following the analyses above, a model in which NT_p varies by cue validity according to a power law explained nearly 88% of the variance in the data, adjusted R^2 = .8791, RMSE = .02741. Thus, both G_P and NT_P appear to vary as a function of the allocation of attention, independent of the memory load. That is, both G_P and NT_P are affected by resource allocation independently of the total number of to-be-remembered items. For example, in Experiment 1a, despite doubling the number of to-be-remembered items from 3 to 6, there was no difference in the proportion of guesses or swap errors between the 3 cues – 100% and 1 cue – 33% conditions, $t(19) = 1.323$, $p = .201$, d = .296 and t(19) = .981, $p = .339$, $d = .219$.

Model Comparison

Altogether, the preceding analyses consistently revealed that probe likelihood is a better predictor of working memory performance than load—regardless of whether performance is measured as overall response error, relative precision, or any of the tested mixture model parameters—and that the relationship with likelihood appears to follow a power law. Moreover, in testing specific conditions for which fixed-capacity and continuous-resource models make different predictions, we were unable to find evidence of a fixed memory capacity. Consequently, to assess whether the effect of probe likelihood holds up when other theorized effects on memory performance are taken into account, — as well as to directly compare how well different models perform when conditions are described by probe likelihood rather than by mnemonic load — we next evaluated how well four leading models of working memory performance fit our data. Specifically, we examined two fixed-capacity models already discussed, the slots+averaging and slots+resoures models (Zhang & Luck, 2008), as well as two capacity-free, continuous-resource models: an equal-precision model, in which each item in the array is assumed to have equal precision (Bays & Husain, 2008), and a variable-precision model, in which precision varies across items and trials (van den Berg et al., 2012; see also, Fougnie et al., 2012). Importantly, both the equal- and variable-precision models assume that precision varies according to a power law relationship to set size. The parameters of the four models were fit directly to the raw data using the procedure described by van den Berg et al. (2014).

In order to examine the effects of the deliberate (as opposed to the random) allocation of attention across each of these models, we analyzed the data both by memory load and by probe likelihood. That is, for each model and condition, we performed two analyses: one in which the model parameter *N* was defined as the total number of to-be-remembered items, which has previously been the standard; in the second analysis the model parameter *N* was defined as 1/probe likelihood (rounded to the nearest item). That is, if an item was probed with 33% probability, we assumed that this is equivalent to a condition in which three items were probed with equal probability. Thus, independent of whether VWM is limited by a fixed-capacity or a continuous-resource, we would predict that if attentional allocation best determines the precision of VWM representation, then the second group of models defined by probe likelihood should outperform those defined by memory load alone. Moreover, if the relationship between attentional allocation and VWM performance is related by a power law, we would expect the continuous-resource models to be a better fit to the data than the fixed-capacity models.

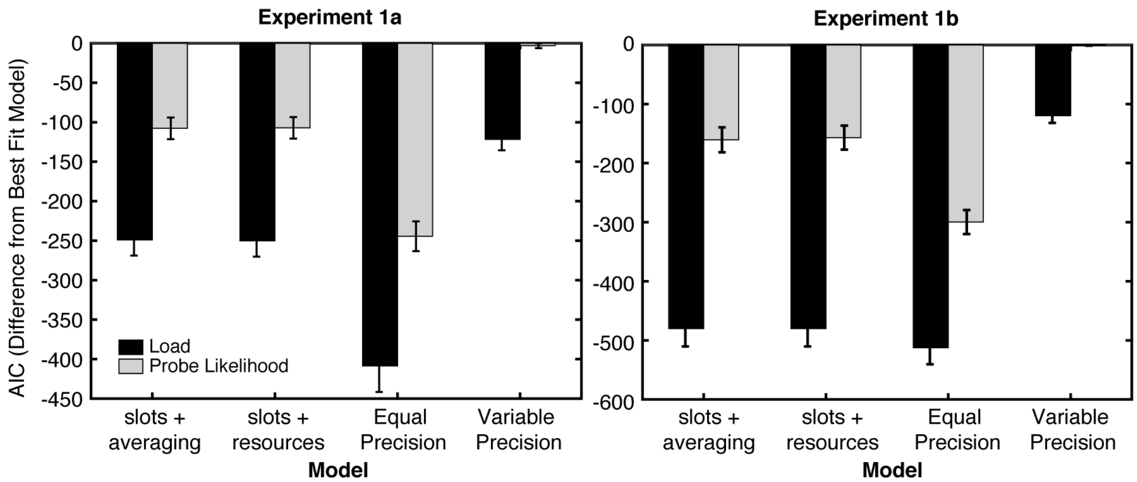

Figure 4 - Change in AIC values relative to the best fitting model for Experiments 1a (left) and 1b (right). Values closer to zero indicate better model fits. Black bars represent the model fits when data were modeled as a function of load (the total number of to-be-remembered items). Grey bars represent the model fits when data were modeled as a function of probe likelihood. Error bars represent 1 standard error of the mean across subjects.

The change in AIC values for each model and each experiment (relative to the best-fitting model) are presented in Figure 4. (The complete list of model parameters for each model, fit type, and experiment can be found in Supplementary Tables 3 – 4). As can be seen from the figure, across both experiments, the best fitting model was the variable-precision model fit to probe likelihood (power constant alpha = -1.43 and -1.2 in Experiments 1a and 1b, respectively). This was true for 19/20 individual subjects in each experiment, with the exceptions being a variable-precision model fit by load in Experiment 1a and the slot-averaging model fit by probe likelihood in Experiment 1b. This finding is consistent with previous studies that have demonstrated strong evidence in favor of variable precision models over other models (Fougnie et al., 2012; van den Berg et al., 2012). However, our findings suggest that a significant proportion of fluctuations in attention across items can be accounted for the probe likelihood, as the variable-precision model fit to probe likelihood outperformed the same model fit to memory load for all but 1 of 40 subjects across two experiments. Thus, although there remains a significant amount of variability in memory precision across items and trials, the data are consistent with a model in which precision scales according to a power law defined by the proportion of resources allocated to each item.

Discussion

In the current study, we examined VWM performance on a continuous report task by manipulating both the number of items to be remembered (load) as well as the relative likelihood that a sample item would be probed (cue validity). Our results revealed four novel findings: First, the amount of error in a VWM recall task ($SD_{response}$) is better predicted by probe likelihood than by the overall memory load. Second, uncued items can receive only a small proportion of memory resources relative to neutral-cued items, and the proportion of resources received is determined by the probe validity of each item. Third, the measures obtained through the three-component mixture model (Bays et al., 2009) are all strongly predicted by the probe likelihood of each item, and this effect follows a simple power law. Finally, raw error was best fit by a model in which precision varied as a function of the proportion of resources allocated to each item according to a power law, in addition to varying randomly across items and trials. Together, the results favor a model of VWM in which performance is limited by the allocation of a continuous attentional resource.

There are a number of implications that arise from these findings. First, we reveal that although load (i.e., the number of items to be remembered) can be a strong predictor of VWM performance, it is often confounded with the proportion of resources allocated to each item; when considered at the same time, resource allocation is the more useful predictor. By using cue validity as a proxy for attentional allocation, we observed that changes in load have little effect on VWM performance when attention is held constant, whereas changes in attention have a large effect on performance when load is held constant. This finding mirrors those observed in visual search, which have demonstrated that the relevant set size, rather than the sensory load per se, is a limiting factor (Palmer, 1994; Palmer, Ames, & Lindsey, 1993). Although these previous studies did not vary the probability of the attentional cues in visual search, they are consistent with the finding that the effect of attentional allocation during encoding is independent of the actual stimulus load, and is better predicted by the relative priority of items in the display. Our findings consequently have significant implications for models of VWM, which are largely focused on how many items must be maintained, rather than on how resources are flexibly distributed across items. Specifically, these results suggest that future models and empirical investigations of VWM capacity would be better informed by assessing the effects of item probability rather than memory load.

Second, although previous studies have shown that attentional priority can affect overall precision (Klyszejko et al., 2014), we are the first to demonstrate that this effect is predicted by a power law function. Moreover, the finding that a power law function could explain between 88 – 98% of the variance regardless of whether we were describing overall response error, relative precision, or any of the measures from the three-component mixture model speaks to the explanatory power of the continuous resource model over the discrete-capacity model. That is, continuous-resource models make few assumptions about how VWM performance should change as a function of load (or the distribution of attention) beyond the fact that it should change proportionally to the amount of resources allocated to each item (Ma et al., 2014; Zokaei et al., 2011). Here, we show that this holds independent of which aspect of performance is measured (i.e., precision, guessing, or swap errors). Thus, although we should be cautious in interpreting the power law fits from the mixture model (van den Berg & Ma, 2014), the finding that each measure can be explained by the same function (and that this function explains upwards of 90% of the variance across conditions) provides compelling support for the continuous-resource model. In other words, the finding that power law functions provide a strong fit regardless of which measure is used provides a parsimonious explanation for the relationship between memory load, attention, and behavior. Moreover, the finding that both non-target errors and guessing rates increased with decreasing probe likelihood suggests that these two types of errors may both be attributable to insufficient allocation of attention to individual items, thereby increasing the likelihood of memory decay (Pertzov, Bays, Joseph, & Husain, 2013), or replacement through competitive processes during encoding (Emrich & Ferber, 2012), although the exact effect attention has on guess rates and non-target errors likely requires further study.

Third, we demonstrate that VWM precision (including when measured using a mixture model) can vary across items within memory, and the proportion of resources allocated to a given item can be very small. These results contrast those of Zhang and Luck (2008) who observed no differences in the precision between neutral cued items and invalidly cued items. The authors suggested this finding was evidence for a slots+averaging model, in which resources are allocated via discrete slots, making it impossible for items to receive "just a few drops" of resources. There are a few possible reasons why our results differ from those of Zhang and Luck (2008). First, we used a three-component mixture model (Bays et al., 2009), which includes a parameter for non-target errors, as opposed to the two-component model used by Zhang and Luck (2008). These responses have been shown to potentially make up a significant proportion of responses in a continuous recall task (van den Berg at al., 2014), and can therefore affect the precision measure of a mixture model (Bays et al., 2009). Second, as can be observed in Table 2, our results suggest that the effect size of the difference in the circular SD of the mixture model between conditions depends in part on the magnitude of the difference in resource allocation. Thus, in the study by Zhang and Luck (2008), the difference in resource allocation between neutral conditions (25% per item) and invalid conditions (10% per item) may have been too small to have observed a significant difference with the size of their sample. Ultimately, however, differences between the two studies may simply be due to the unreliability of the mixture model measures at the tested number of trials (van den Berg & Ma, 2014).

We also found that variable-precision models are the best overall fit to the data, consistent with a number of previous findings (Fougnie et al., 2012; van den Berg et al., 2012) that demonstrate that VWM performance can be explained by spontaneous item-to-item fluctuations in memory fidelity, in addition to load demands. Whereas these models are either equivocal about the source of variability (attributing it to attention, arousal, or random noise (van den Berg et al., 2012)) or ascribe it to stochastic degradation of representations over time (Fougnie et al., 2012), our results reveal a decidedly non-stochastic source of variance: spatial attention. Although the variable-precision model fit to probe likelihood was the best overall fit to the data, it is important to note that across all four tested models, the models fit to probe likelihood out-performed the same models fit to overall load. Thus, even though it is possible that there exist other models which may better predict performance (van den Berg et al., 2014), our findings suggest that future models would benefit from modeling behavior as a function of attentional priority, instead of or in addition to modeling behavior as a function of memory load alone.

How does attention regulate the variability of precision? While it is possible that attention offsets degradation during maintenance (Murray, Nobre, Clark, Cravo, & Stokes, 2013), the presence of the predictive spatial cues during the sample is consistent with a mechanism in which VWM performance is limited by encoding processes (Emrich & Ferber, 2012; Linke, Vicente-Grabovetsky, Mitchell, & Cusack, 2011; Mazyar, van den Berg, & Ma, 2012; van den Berg et al., 2012). Our results are also consistent with a population-coding account of the effect of attention on errors in VWM (Bays, 2014). Namely, Bays (2014) demonstrated that an increase in response errors with increasing memory load could be accounted for by a decrease in the signal-to-noise in a population of neurons tuned to the features of the memory items. This study also demonstrated that the effect of attention on memory could be modeled as a gain boost to the neurons coding the features of the attended (cued) item or location. Critically, the gain factor required to optimize performance increased as a function of load. In other words, because additional items were occupying additional resources (increasing the signal-to-noise ratio), a greater gain signal was required to prioritize a single cued (attended) item over the remaining uncued items. Importantly, the results also demonstrated that human observers could perform at close to optimal levels in order to prioritize the cued item and minimize overall errors. Thus, this study provides a neural basis for observers in the current study to have efficiently and optimally allocated attentional resources according to the predictive value and number of cues.

One additional finding of note is the discrepancy between our results and those of Zhang and Luck (2011), who tested whether decreasing the required precision for a correct response in an VWM task would lead participants to store each item with a lower precision, in turn increasing the number of remembered items. Consistent with a discrete capacity, they observed no changes in precision, even when performance was rewarded with financial incentives. There are a number of factors that could account for these differences. First and foremost, our findings are the direct result of manipulations that explicitly encourage participants to flexibly allocate attention over varying numbers of stimuli. If, as our results suggest, the allocation of VWM resources hinges on the allocation of attention during encoding, Zhang and Luck's manipulation of response criterion may have had no effect on the proportion of VWM resources allocated to each item, resulting in null effects. Second, it is possible that while participants in the Zhang and Luck (2011) study were trading off capacity and precision on individual items, the analysis method used could not detect any such trade off; that is, by treating all items equally, it is possible that the study of Zhang and Luck (2011) failed to capture the variance between individual items, as is reported here. Third, previous studies have demonstrated that larger and more varying incentives can lead subjects to increase precision for individual items (Klyszejko et al., 2014). Importantly, this effect was found when manipulations emphasized prioritizing individual items, rather than manipulations that prioritized high vs. low precision. Thus, it appears possible to flexibly allocate resources according to task demands, at least when performance for individual items is accounted for.

Ultimately, our results provide further support for the proposal that attention and VWM are inextricably linked (Awh & Jonides, 2001; Gazzaley & Nobre, 2012; Postle, 2006; Souza, Rerko, Lin, & Oberauer, 2014), although other factors likely also contribute to the capacity limits of VWM (Brady & Alvarez, 2015). A recent controversy in the field relates to whether information in VWM is stored in areas of frontal and parietal cortices or in sensory visual cortex (Christophel, Cichy, Hebart, & Haynes, 2015; Emrich, Riggall, Larocque, & Postle, 2013; Ester, Sprague, & Serences, 2015; Riggall & Postle, 2012; Roggeman, Klingberg, Feenstra, Compte, & Almeida, 2014). Given that the results here speak to the importance of top-down attentional control, one interesting proposal is that although storage itself may occur in sensory cortex, the performance-limiting attentional allocation occurs via signals from fronto-parietal networks. Regardless, the results speak to the importance of further elucidating the nature of the relationship between neural mechanisms of attentional selection and VWM storage (Awh & Jonides, 2001; Postle, 2006).

Acknowledgements

The authors would like to thank Thomas Nelson for help with experiment programming. This work was supported by Natural Sciences and Engineering Research Council of Canada grant [435945] awarded to S.M.E.

References

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126.

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(10), 3632–3645. http://doi.org/10.1523/JNEUROSCI.3204-13.2014

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*(10), 7.1–11. http://doi.org/10.1167/9.10.7

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science (New York, N.Y.)*, *321*(5890), 851–854. http://doi.org/10.1126/science.1158023

Brady, T. F., & Alvarez, G. A. (2015). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *41*(3), 921–929. http://doi.org/10.1037/xlm0000075

Christophel, T. B., Cichy, R. M., Hebart, M. N., & Haynes, J.-D. (2015). Parietal and early visual cortices encode working memory content across mental transformations. *NeuroImage*, *106*, 198–206. http://doi.org/10.1016/j.neuroimage.2014.11.018

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87–114; discussion 114–185.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*(3), 272–300.

Emrich, S. M., & Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision*, *12*(4). http://doi.org/10.1167/12.4.12

Emrich, S. M., Riggall, A. C., Larocque, J. J., & Postle, B. R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(15), 6516–6523. http://doi.org/10.1523/JNEUROSCI.5732-12.2013

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*. http://doi.org/10.1016/j.neuron.2015.07.013

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229. http://doi.org/10.1038/ncomms2237

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129–135. http://doi.org/10.1016/j.tics.2011.11.014

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *The Journal of Neuroscience:*

The Official Journal of the Society for Neuroscience, *31*(23), 8502–8511. http://doi.org/10.1523/JNEUROSCI.0208-11.2011

JASP (Version 0.7.5.5)[Computer software].

Klyszejko, Z., Rahmati, M., & Curtis, C. E. (2014). Attentional priority determines working memory precision. *Vision Research*, *105*, 70–76. http://doi.org/10.1016/j.visres.2014.09.002

Linke, A. C., Vicente-Grabovetsky, A., Mitchell, D. J., & Cusack, R. (2011). Encoding strategy accounts for individual differences in change detection measures of VSTM. *Neuropsychologia*, *49*(6), 1476–1486. http://doi.org/10.1016/j.neuropsychologia.2010.11.034

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. http://doi.org/10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. http://doi.org/10.1016/j.tics.2013.06.006

Machizawa, M. G., Goh, C. C. W., & Driver, J. (2012). Human visual short-term memory precision can be varied at will when the number of retained items is low. *Psychological Science*, *23*(6), 554–559. http://doi.org/10.1177/0956797611431988

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. http://doi.org/10.1038/nn.3655

Mazyar, H., van den Berg, R., & Ma, W. J. (2012). Does precision decrease with set size? *Journal of Vision*, *12*(6), 10. http://doi.org/10.1167/12.6.10

Murray, A. M., Nobre, A. C., Clark, I. A., Cravo, A. M., & Stokes, M. G. (2013). Attention restores discrete items to visual short-term memory. *Psychological Science*, *24*(4), 550–556. http://doi.org/10.1177/0956797612457782

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, *34*(13), 1703–1721.

Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 108.

Pertzov, Y., Bays, P. M., Joseph, S., & Husain, M. (2013). Rapid forgetting prevented by retrospective attention cues. *Journal of Experimental Psychology. Human Perception and Performance*, *39*(5), 1224–1231. http://doi.org/10.1037/a0030947

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23–38. http://doi.org/10.1016/j.neuroscience.2005.06.005

Riggall, A. C., & Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(38), 12990–12998. http://doi.org/10.1523/JNEUROSCI.1892-12.2012

Roggeman, C., Klingberg, T., Feenstra, H. E. M., Compte, A., & Almeida, R. (2014). Trade-off between capacity and precision in visuospatial working memory. *Journal of Cognitive Neuroscience*, *26*(2), 211–222. http://doi.org/10.1162/jocn_a_00485

Souza, A. S., Rerko, L., Lin, H.-Y., & Oberauer, K. (2014). Focused attention improves working memory: implications for flexible-resource and discrete-capacity models. *Attention, Perception & Psychophysics*, *76*(7), 2080–2102. http://doi.org/10.3758/s13414-014-0687-2

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29.

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10). http://doi.org/10.1167/13.10.9

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. http://doi.org/10.1016/j.cogpsych.2014.01.003

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. http://doi.org/10.1037/a0035234

van den Berg, R., & Ma, W. J. (2014). “Plateau”-related summary statistics are uninformative for comparing working memory models. *Attention, Perception*

& Psychophysics, *76*(7), 2117–2135. http://doi.org/10.3758/s13414-013-0618-7

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 of the United States of America*, *109*(22), 8780–8785. http://doi.org/10.1073/pnas.1117465109

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135. http://doi.org/10:1167/4.12.11

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. http://doi.org/10.1038/nature06860

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, *22*(11), 1434–1441. http://doi.org/10.1177/0956797611417006

Zokaei, N., Gorgoraptis, N., Bahrami, B., Bays, P. M., & Husain, M. (2011). Precision of working memory for visual motion sequences and transparent motion surfaces. *Journal of Vision*, *11*(14). http://doi.org/10.1167/11.14.2

Tables

Table 1. Stimulus parameters for conditions in Experiments 1a and 1b

	Condition	Number of Items	Number of Cues	Memory Load	Cue Validity (%)	Probability of Cued Item Probed (%)	Probability of Uncued Item Probed (%)
Experiment 1a							
	1 cue – 100%	6	1	1	100	100	0
	2 cues – 100%	6	2	2	100	50	0
	3 cues – 100%	6	3	3	100	33	0
	1 cue – 50%	6	1	6	50	50	10
	2 cues – 66%	6	2	6	66	33	8.25
	1 cue – 33%	6	1	6	33	33	13
	6 cues – 100%	6	6	6	100	100	N/A
Experiment 1b							
	1 cue – 100%	4	1	1	100	100	0
	2 cues – 100%	4	2	2	100	50	0
	3 cues – 100%	4	3	3	100	33	0
	4 cues – 100%	4	4	4	100	25	0
	2 cues – 66%	4	2	4	66	33	16.5
	1 cue – 33%	4	1	4	33	33	22
	3 cues – 66%	4	3	4	66	22	33
	Load 1, 1 cue - 100%	1	1	1	100	16	N/A

Table 2. *t*-test statistics comparing SD_{MM} for neutral (3 cues – 100%) trials to uncued trials (uncorrected, one-tailed)

Experiment	Set Size	Condition	Probe Likelihood (uncued items) (%)	t (df = 19)	*p*	Cohen's *d*
1a						
	6	1 cue – 50%	10	-1.730	.05	-.537
	6	2 cues – 66%	8.25	-3.448	.001	-1.141
	6	1 cue – 33%	13	-1.835	.041	-0.557
1b						
	4	2 cues – 66%	16.5	-1.715	.051	-0.579
	4	3 cues – 66%	33	-0.558	.292	-1.175
	4	1 cue – 33%	22	-1.558	.068	-0.449

Supplementary Tables

Supplementary Table 1. Fits for validation data (uncued trials) for the power law model described in Figures 2 and 3.

Measure	Fit	RMSE
$SD_{response}$	Power (by load)	10.58
	Power (by Probe Likilihood)	15.78
Relative Precision	Power (by load)	0.048
	Power (by Probe Likelihood)	0.09
SD_{MM}	Power	11.08
G_P	Power	0.35
NT_P	Power	0.21

Supplementary Table 2. Means (and SDs) for response error ($SD_{response}$) and mixture model measures (SD_{MM}, G_p, NT_p), by condition.

		Condition	$SD_{response}$	SD_{MM}	Gp	NTp
Experiment 1a						
	Cued Trials					
		1 cue – 100%	16.2 (8.7)	15.9 (3.5)	0.04 (0.1)	0.01 (0.01)
		2 cues – 100%	27.4 (10.4)	20.6 (4.2)	0.08 (0.1)	0.04 (0.06)
		3 cues – 100%	40.0 (13.1)	21.5 (4.4)	0.19 (0.2)	0.1 (0.2)
		1 cue – 50%	30.8 (10.2)	20.1 (4.7)	0.1 (0.1)	0.04 (0.06)
		2 cues – 66%	37.5 (11.2)	20.8 (4.1)	0.17 (0.2)	0.1 (0.1)
		1 cue – 33%	35.2 (12.4)	21.4 (6.7)	0.15 (0.1)	0.06 (0.1)
		6 cues – 100%	51.1 (6.3)	28.2 (18.8)	0.3 (0.3)	0.3 (03.)
	Uncued Trials					
		1 cue – 50%	52.9 (2.0)	31.4 (25.7)	0.31 (0.3)	0.39 (0.3)
		2 cues – 66%	53.6 (3.4)	47.8 (32.3)	0.37 (0.3)	0.43 (0.4)
		1 cue – 33%	52.6 (3.6)	28.3 (14.5)	0.36 (0.2)	0.35 (0.2)
Experiment 1b						
	Cued Trials					
		1 cue – 100%	15.2 (7.3)	15.9 (2.7)	0.02 (0.03)	0.01 (0.01)
		2 cues – 100%	28.8 (13.7)	18.1 (4.0)	0.10 (0.2)	0.06 (0.1)
		3 cues – 100%	41.4 (10.8)	22.2 (5.3)	0.17 (0.1)	0.15 (0.2)
		4 cues – 100%	45.2 (9.5)	21.4 (5.2)	0.28 (0.2)	0.16 (0.2)
		2 cues – 66%	39.9 (10.8)	22.0 (4.2)	0.17 (0.3)	0.12 (0.1)
		1 cue – 33%	35.2 (8.1)	21.0 (3.6)	0.15 (0.1)	0.05 (0.04)
		3 cues – 66%	43.1 (9.6)	22.1 (4.1)	0.24 (0.3)	0.15 (0.1)
		Load 1, 1 cue - 100%	15.1 (6.6)	22.0 (4.1)	0.03 (0.03)	N/A
	Uncued Trials					
		2 cues – 66%	49.6 (5.8)	31.5 (22.2)	0.29 (0.3)	0.26 (0.3)
		1 cue – 33%	47.7 (6.4)	23.0 (3.9)	0.27 (0.1)	0.23 (0.2)
		3 cues – 66%	45.9 (5.7)	25.6 (9.5)	0.31 (0.2)	0.17 (0.22)

Supplementary Table 3. Maximum Likelihood Estimates for Each Model and Fit Type Experiment 1a

Model	Fit Type	Parameter	*M ± SEM*	*Mdn*
Slots+Averaging	Load	$\log J_1$	8.25 ± 7.46	1.99
		$\log \kappa_1$	10.25 ± 9.21	9.47
		K_{mean}	2.15 ± 0.13	2
	Probability	$\log J_1$	1.77 ± 0.33	1.67
		$\text{Log}\ \kappa_1$	11.37 ± 10.2	10.8
		K_{mean}	1.75 ± 0.1	2
Slots+Resources	Load	$\log J_1$	9.33 ± 8.9	2.5
		α	-0.86 ± 0.23	-0.67
		$\log \kappa_1$	9.41 ± 8.1	9.0
		K_{mean}	2.1 ± 0.12	2
	Probability	$\log J_1$	7.6 ± 7.05	2.51
		α	-1.25 ± 0.17	-1.8
		$\log \kappa_1$	9.67 ± 8.39	9.34
		K_{mean}	1.75 ± 0.1	2
Equal Precision	Load	$\log J_1$	2.12 ± 0.1	2.10
		α	-1.61 ± 0.06	-1.58
		$\log \kappa_1$	10.77 ± 9.0	10.62
	Probability	$\log J_1$	10.76 ± 10.62	2.44
		α	-3.25 ± 0.70	-2.16
		$\log \kappa_1$	10.70 ± 9.15	10.24
Variable Precision	Load	$\log \bar{J}_1$	5.6 ± 3.85	5.43
		α	-0.89 ± 0.043	-0.89
		$\log \tau$	6.40 ± 4.72	6.09
		$\log \kappa_1$	3.03 ± 0.53	2.99
	Probability	$\log \bar{J}_1$	6.23 ± 5.17	5.55
		α	-1.43 ± 0.07	-1.43
		$\log \tau$	6.63 ± 5.87	6.06
		$\log \kappa_1$	2.96 ± 0.45	2.96

Supplementary Table 4. Maximum Likelihood Estimates for Each Model and Fit Type Experiment 1b

Model	Fit Type	Parameter	*M* ± *SEM*	*Mdn*
Slots+Averaging	Load	log J_1	8.2 ± 7.8	1.11
		log κ_1	10.41 ± 8.01	10.09
		K_{mean}	2.55 ± 0.47	2
	Probability	log J_1	7.74 ± 6.93	2.23
		log κ_1	11.53 ± 11.14	9.14
		K_{mean}	1.85 ± 0.08	2
Slots+Resources	Load	log J_1	7.19 ± 6.89	0.96
		α	-1.08 ± 0.35	-0.58
		log κ_1	10.9 ± 9.83	9.95
		K_{mean}	3.6 ± 0.89	2
	Probability	log J_1	7.39 ± 7.38	2.33
		α	-1.01± 0.38	-0.79
		log κ_1	10.0 ± 8.77	8.99
		K_{mean}	1.85 ± 0.08	2
Equal Precision	Load	log J_1	9.35 ± 9.13	0.92
		α	-1.47 ± 0.43	-0.9
		log κ_1	11.43 ± 10.48	10.65
	Probability	log J_1	2.29 ± -.03	2.31
		α	-1.86 ± 0.17	-1.69
		log κ_1	10.75 ± 9.014	10.58
Variable Precision	Load	log $\bar{J}_1$	6.07 ± 5.23	4.88
		α	-0.71 ± 0.10	-0.61
		log τ	8.08 ± 7.49	5.84
		log κ_1	3.11 ± 0.9	2.97
	Probability	log $\bar{J}_1$	6.18 ± 5.51	4.95
		α	-1.2 ± 0.1	-1.04
		log τ	6.84 ± 6.43	4.98
		log κ_1	3.2 ± 0.99	3.05

Figure Captions.

Figure 1. A schematic of experimental trials from Experiment 1a. Participants were instructed to report the color of the probed item (as indicated by the bold outline). The number of items to be remember was determined by the validity of the cues, which was indicated to participants prior to each block. Experiment 1b used total set sizes of 4 or 1, otherwise the procedures were identical.

Figure 2. Overall response error ($SD_{response}$) and relative precision, plotted by total number of to-be-remembered items (A,C) or the likelihood of an item being probed (B,D). Data are fit to a power law function as in Bays & Husain (2008). Green and red items represent data from Experiments 1a and 1b, respectively. Closed circles represent cued items, open diamonds represent uncued items (validation data). Dashed lines indicated 95% confidence intervals of the fitted models. Full means and SDs are available in Supplementary Table 2.

Figure 3. Error (SD_{MM}), guess rate (G_p) and non-target errors (NT_p) obtained from the three-component mixture model (Bays et al., 2009). (Left) Data are plotted by total number of items to-be-remembered. (Right) Plotting the data by the likelihood of an item being probed and fitting it to a power law function is a strong predictor across all three measures. Green and red items represent data from Experiments 1a and 1b, respectively. Closed circles represent cued items, open diamonds represent uncued items (validation data). Dashed lines indicated 95% confidence intervals of the fitted models. Full means and SDs are available in Supplementary Table 2.

Figure 4. Change in AIC values relative to the best fitting model for Experiments 1a (left) and 1b (right). Black bars represent the model fits when data were modeled as a function of load (the total number of to-be-remembered items). Grey bars represent the model fits when data were modeled as a function of probe likelihood. Error bars represent 1 standard error of the mean across subjects.