Behavioral/Systems/Cognitive

Precision in Visual Working Memory Reaches a Stable Plateau When Individual Item Limits Are Exceeded

David E. Anderson, Edward K. Vogel, and Edward Awh

Department of Psychology, University of Oregon, Eugene, Oregon 97403

Multiple studies have demonstrated that resolution in working memory (WM) declines as the number of stored items increases. Discrete-resource models predict that this decline should reach a stable plateau at relatively small set sizes because item limits prevent additional information from being encoded into WM at larger set sizes. By contrast, flexible-resource models predict that the monotonic declines in precision will continue indefinitely as set size increases and resources are distributed without any fixed item limit. In the present work, we found that WM resolution exhibited monotonic declines until set size reached three items, after which resolution achieved a clear asymptote. Moreover, analyses of individual differences showed a strong correlation between each observer's item limit and the set size at which WM resolution achieved asymptote. These behavioral observations were corroborated by measurements of contralateral delay activity (CDA), an event-related potential waveform that tracks the number of items maintained during the delay period. CDA activity rose monotonically and achieved asymptote at a set size that predicted individual WM capacity. Moreover, this neural measure of on-line storage also predicted the set size at which mnemonic resolution reached a stable plateau for each observer. Thus, independent behavioral and neural measures of WM capacity support a clear prediction of discrete-resource models. Precision in visual WM reaches asymptote when individual item limits are exceeded.

Introduction

There is active debate regarding the nature of the resources that limit working memory (WM) capacity. Flexible-resource models suggest that WM capacity is determined by a central pool of resources that can be subdivided among an unlimited number of items (Wilken and Ma, 2004; Bays and Husain, 2008). By contrast, discrete-resource models suggest that WM resources are quantized such that any stored item must be assigned to one of a limited number of "slots" (Fukuda et al., 2010). The plausibility of the discrete-resource view is supported by computational modeling and neural studies that use asynchronous neural oscillations as a coding scheme for the storage of individuated representations in memory (Lisman and Idiart, 1995; Raffone and Wolters, 2001). Thus, discrete item limits in WM may result from a biophysical limit regarding the number of asynchronous oscillatory patterns that can be simultaneously maintained. Such a biophysical limit may underlie recent observations from EEG and functional magnetic resonance imaging studies that delay-specific activity in posterior parietal regions rises monotonically as the number of items to be stored increases, and reaches asymptote when individual memory capacity is exceeded (Vogel and Machizawa, 2004; Todd and Marois, 2005).

Here, we demonstrate a direct link between this asymptotic function in storage-related neural activity and the previously demonstrated inverse relationship between WM resolution and the number of items stored. As the number of items stored in WM increases, monotonic declines in WM resolution are observed (Bays and Husain, 2008; Zhang and Luck, 2008; Barton et al., 2009). These declines can be naturally explained by flexible-resource models that posit a finer division of WM resources as the number of stored items increases. At the same time, discrete-resource models can accommodate this finding by assuming a fixed set of slots that constrains the allocation of a separate resource that determines mnemonic resolution. Critically, however, flexible-resource models predict that resolution will decline monotonically across arbitrarily large set sizes, as resources are divided between larger numbers of representations. By contrast, discrete-resource models predict that WM resolution will achieve a stable lower bound at relatively low set sizes because no additional information is encoded into WM when putative item limits are exceeded. This prediction was supported by Zhang and Luck (2008) who found that precision was equivalent when the number of to-be-stored items varied from three to six. The present work extends this evidence in two important ways. First, we use fine-grained manipulations of set size to show that resolution achieves a stable lower bound at a set size that is strongly predicted by individual WM capacity. In line with discrete-resource models, this asymptotic function for mnemonic resolution strongly suggests that no additional items are added into WM once individual item limits are exceeded. Second, we corroborate this empirical pattern by using delay-specific electrophysiological activity to operationalize individual WM capacity. Thus, our findings demonstrate a direct link between delay-specific neural

activity and the dynamic shifts in WM resolution that occur when on-line storage loads change.

Materials and Methods

Overview. Multiple studies have demonstrated that resolution in WM declines as the number of stored items increases (Bays and Husain, 2008; Zhang and Luck, 2008; Barton et al., 2009). In the present work, a primary empirical pattern of interest was the shape of this resolution by set size function. On the one hand, discrete-resource models predict that resolution will reach a stable plateau when set size exceeds a fixed item limit because no additional information is added to WM at larger set sizes. On the other hand, flexible-resource models predict monotonic declines in mnemonic precision across much larger set sizes as resources are divided across larger numbers of items. Here, we used an orientation WM task to test these diverging predictions. A straightforward method for measuring the precision of orientation WM is to determine the angular deviation between the observer's recall of a probed stimulus and the true orientation of that item. A critical challenge for this approach, however, is that increases in response offsets with larger set sizes could be caused either by reduced precision of the stored orientations or by an increased probability that the critical item was not stored (which would elicit random responses with large mean deviations from the true value). To separate these two aspects of memory function, we relied on a recently developed analytic procedure that was used to test whether or not observers could store all items from a sample display in WM or whether WM capacity was constrained by a discrete item limit (Zhang and Luck, 2008). If there are discrete item limits in WM, then observers' responses with supraspan sample arrays should fall into two distinct categories. On some trials, the probed item will have been stored, and response offsets should fall into a Gaussian distribution centered on the angle of the sample stimulus (with the SD of that distribution monotonically related to the precision of the memory representation). On other trials, however, the critical item will not have been stored, such that responses will be random with respect to the critical item, creating a flat distribution across orientation space whose height provides a direct measure of the probability of storage failures. Thus, the overall distribution of response offsets should be a weighted sum of the Gaussian and flat distributions. Zhang and Luck (2008) tested these predictions with both color and shape stimuli and found that this mixture model provided an excellent description of the distribution of response errors. This mixture model was contrasted with a flexible-resource model that assumed successful storage of all items in the sample display, with decreasing precision as the number of stored items increased. The key finding was that the mixture model explained substantially more variance than the flexible-resource model, suggesting that WM capacity was indeed subject to fixed item limits.

Although the findings of Zhang and Luck (2008) provide strong evidence for discrete item limits in WM capacity, the present data seek to extend these results in two critical ways. First, we used a fine-grained manipulation of set sizes that allowed a more precise description of how resolution in WM changed as set size increased. As noted above, the discrete- and flexible-resource models make distinct predictions about the shape of this resolution by set size function because only the discrete-resource model predicts that resolution will arrive at a stable asymptote at relatively low set sizes. Moreover, we tested a relatively large number of observers, thereby enabling a sensitive test of whether estimates of item limits at the individual subject level would be a reliable predictor of the set size at which mnemonic resolution would reach asymptote. Second, we obtained strong converging evidence for these conclusions by relating this resolution by set size function to an on-line neural measure of how many items were stored in WM at each set size. Establishing this link between behavior and storage-related neural activity greatly strengthened the inferential power of our study because this approach addresses well known ambiguities in the interpretation of purely behavioral measures of capacity. For example, there is an important distinction between information that is currently active in working memory and that which is represented via latent episodic traces in long-term memory. Despite the relatively brief duration of the delay period in standard WM procedures, it is difficult to assess whether or not response accuracy is influenced by information retrieved from long-term memory rather than mnemonic
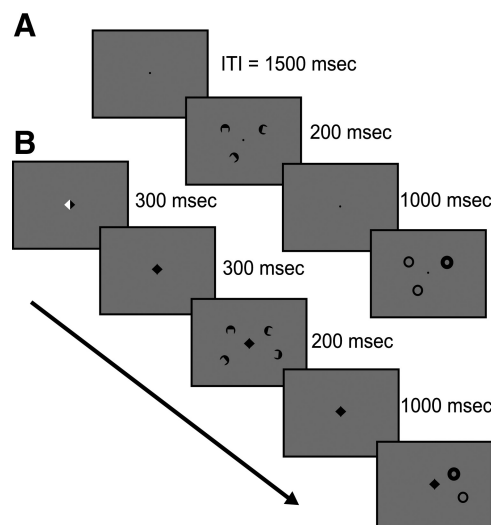
Figure 1. *A*, Visual working memory task used in experiment 1. Participants maintained fixation and were instructed to remember the orientations of all objects presented on the display. Set sizes used were one, two, three, four, six, and eight. After a short delay period, participants were probed to recall the orientation of one object presented in the memory display (demarcated with a thicker black ring). Participants responded by clicking on the location of the ring where they remembered the center of the gap being. *B*, ERP visual working memory task used in experiment 2. Participants were to attend to the hemifield indicated by one of two colored arrows (counterbalanced). Set sizes ranged from one to six. All other procedures were identical with those of experiment 1.

representations that were actively sustained in working memory. In addition, changes in behavioral performance across rising set sizes could potentially be influenced by changes in interference during the retrieval or execution of responses rather than changes in the quality of information stored in WM. The present work addresses these ambiguities by showing a direct link between the behavioral measures of mnemonic resolution across set sizes and on-line neural measures of WM storage that cannot be explained by the retrieval of episodic long-term memories, or the characteristics of later retrieval processes. Specifically, we measured contralateral delay activity (CDA) by recording event-related potentials (ERPs) over posterior scalp electrodes and measuring the difference in mean amplitude between electrodes ipsilateral and contralateral to the memory array. As in previous work, the amplitude of the CDA was a robust predictor of individual memory capacity, suggesting that this neural measure provides an on-line index of the number of items held active in WM during the 1 s delay period. Thus, our work provides a direct link between delay-specific neural activity that indexes storage in WM and the resolution by set size function at the core of this investigation.

Participants. A total of 77 (44 in experiment 1 and 33 in experiment 2) undergraduates at the University of Oregon completed the experiment for course credit (experiment 1) or monetary compensation (experiment 2). All had normal or corrected-to-normal visual acuity and gave informed consent according to procedures approved by the University of Oregon institutional review board.

Stimulus displays. Stimuli were generated in Matlab using the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997) and presented on a 17 inch flat cathode ray tube computer screen (refresh rate of 120 Hz). Viewing distances were ~77 cm (experiment 1) and 100 cm (experiment 2).

Our tasks required participants to remember the orientation of solid discs that contained a rectangular gap, which was randomly positioned at an angle varying across the full 360 space (Fig. 1). Objects had radii of 0.93° (experiment 1) or 0.86° of visual angle (experiment 2).

In experiment 1 (Fig. 1*A*), objects were presented within a square region subtending 10.7 × 10.7° of visual angle, and subjects fixated on a central fixation point that subtended 0.37 × 0.37°. In experiment 2 (Fig. 1*B*), objects were presented within two imaginary rectangles subtending

$4.0 \times 8.5°$ of visual angle, each centered $4.3°$ to the left or right of a central diamond fixation of $0.57 \times 0.57°$ visual angle. Object positions were randomly selected with the constraint that no two objects could fall within $1.43°$ of one another, resulting in a between-object separation of at least two-thirds of an object. In experiment 1, only one object was presented in each quadrant of the visual field when set sizes were less than or equal to four. When set sizes were greater than four, no more than two objects were presented within each quadrant. At the end of each trial, subjects were cued to recall the orientation of a single item. A specific object was probed by outlining its position with a thick ring, whereas the positions of the remaining items were outlined with a thin ring (Fig. 1). Ring color was black in both experiments. Each ring, with a radius of $0.93°$, had rim widths of 0.37 and $0.04°$ for the thicker and thinner rings, respectively.

Procedures. Experiment 1 took ~1 h to complete and was composed of 12 blocks of 60 trials each; experiment 2 took ~2.5 h to complete and was composed of 20 blocks of 60 trials each. The events in a single trial of experiment 1 went as follows. First, subjects saw a central fixation point, followed by the presentation of one, two, three, four, six, or eight discs with oriented gaps for 200 ms. A 1000 ms delay period followed the offset of the discs. After the delay, a probe display of rings appeared in the same positions as the sample items, with one ring highlighted by its extra thickness. Subject clicked on the perimeter of this ring in an unspeeded response to indicate the orientation of the sample item that had appeared in the same position. Each response was followed by a 1500 ms blank intertrial interval. The trial events in experiment 2 were identical to those in experiment 1, with the following exceptions: (1) The central fixation was blue on one side and yellow on the other. Subjects were instructed to pay attention only to the side of the screen indicated by one of the two fixation colors (counterbalanced across subjects). On the unattended side, an irrelevant display of the same number items appeared (to match visual stimulation across the two hemifields). (2) Set sizes varied from one to six.

Modeling response error distributions. Offset values were defined by the difference between the subjects' response and the angle of the probed sample stimulus (ranging from -180 to $180°$). Frequency histograms of response offsets for each set size were created to evaluate the predictions of the discrete- and flexible-resource models.

To test the predictions of the flexible-resource model, maximum-likelihood estimation (MLE) was used to fit the distribution of response errors with a von Mises distribution (the circular analog of a standard Gaussian) to both the group and individual subject data. The von Mises distribution is described by two parameters: μ, the mean of the distribution, which represents any systematic bias in subjects' responses relative to the actual target orientation, and SD, the width of the distribution. SD is inversely proportional to the precision of the representation; a wider distribution is equivalent to a less precise representation.

To test the predictions of the discrete-resource model, MLE was used to fit the distribution of response offsets. Three parameters were estimated: μ, the mean of a von Mises distribution corresponding to trials in which the subject had selected the target location; SD, the width of the same von Mises distribution (used to operationalize mnemonic resolution), and $p(failure)$, denoted P_f. This latter parameter corresponds to the height of a uniform distribution, corresponding to trials in which subjects failed to store the probed item. P_{mem} refers simply to the probability that the critical item was stored (1 minus P_f).

To compare the relative fits of the discrete- and flexible-resource models, we computed the adjusted R^2 statistic based on histograms of the data with 15 bins, each $24°$ wide. The adjusted R^2 statistic reflects the proportion of variance explained by a model weighted by its number of parameters. Models with a greater number of parameters are penalized relative to models with fewer parameters. Thus, this statistic ensures a fair comparison between the three-component discrete-resource model and the two-component flexible-resource model.

Electrophysiological recording and analysis. ERPs were recorded in each experiment using our standard recording and analysis procedures, including rejection of trials contaminated by blocking, blinks, or large ($>1°$) eye movements (Vogel et al., 1998; McCollough et al., 2007). We recorded from 22 tin electrodes mounted in an elastic cap (Electrocap

International) using the International 10/20 System. The 10/20 sites F3, FZ, F4, T3, C3, CZ, C4, T4, P3, PZ, P4, T5, T6, O1, and O2 were used along with five nonstandard sites: OL midway between T5 and O1; OR midway between T6 and O_2; PO3 midway between P3 and OL; PO4 midway between P4 and OR; and POz midway between PO3 and PO4. All sites were recoded with a left-mastoid reference, and the data were rereferenced off-line to the algebraic average of the left and right mastoids. Horizontal electrooculogram (EOG) was recorded from electrodes placed ~1 cm to the left and right of the external canthi of each eye to measure horizontal eye movements. To detect blinks, vertical EOG was recorded from an electrode mounted beneath the left eye and referenced to the left mastoid. Any trials containing either a blink or eye movement were excluded from additional analysis. Subjects with trial rejection rates $>20\%$ were excluded from the sample.

Contralateral waveforms were computed by averaging the activity recorded over the right hemisphere when subjects attended and remembered items in the array at the left side of screen. Contralateral working memory activity was measured at posterior parietal, lateral occipital, posterior temporal, parietal, and occipital electrode sites as the difference in mean amplitude between the ipsilateral and contralateral waveforms. We used two measurement windows: 200–275 ms after the onset of the memory display for the N2pc analyses and 400–800 ms for the memory analyses. The EEG and EOG were amplified with an SA Instrumentation amplifier with a bandpass of 0.01–80 Hz and were digitized at 250 Hz in LabView 6.1 running on a PC.

Results
Experiment 1: relating individual item limits to the resolution by set size function
Model fitting
The discrete-resource model and the flexible-resource model were fitted to the aggregate distributions of response offsets for each set size (Fig. 2). Both the mixture model, representing the predictions of the discrete-resource model (red), and the flexible-resource model were effective in explaining variance in response distributions for each set size [discrete-resource R^2 values (all values of $p < 0.0001$): 1.00 (SS1), 1.00 (SS2), 1.00 (SS3), 1.00 (SS4), 0.98 (SS6), 0.98 (SS8); flexible-resource R^2 values (all values of $p < 0.001$): 0.98 (SS1), 0.91 (SS2), 0.81 (SS3), 0.73 (SS4), 0.67 (SS6), 0.63 (SS8)]. Kolmogorov–Smirnov tests, however, revealed a significant difference between the predicted values of the flexible-resource model and the actual distributions at every set size, whereas the test revealed no difference between the predictions of the discrete-resource model and their fits to each distribution.

We also assessed whether the discrete- and flexible-resource models could account for the distribution of response errors in the individual subject data. Here again, the discrete-resource model was more effective on average than the flexible-resource model in characterizing the response distributions and explaining distribution variance for each subject at every set size. Dependent-samples t tests revealed significant differences between the two models in explaining variance (R^2) in the distributions for all set sizes (set size 1: $t_{(44)} = -3.71, p < 0.001$; set size 2: $t_{(44)} = -7.34, p < 0.001$; set size 3: $t_{(44)} = -10.69, p < 0.001$; set size 4: $t_{(44)} = -13.69, p < 0.001$; set size 6: $t_{(44)} = -11.06, p < 0.001$; set size 8: $t_{(44)} = -11.68, p < 0.001$). To summarize, at both the group and individual subject levels, the discrete-resource model was superior to the flexible-resource model in its ability to account for the distributions of response offsets. This replicates the findings of Zhang and Luck (2008) and suggests that observers could store a subset of the items in the sample array, while maintaining no information about the remaining items.
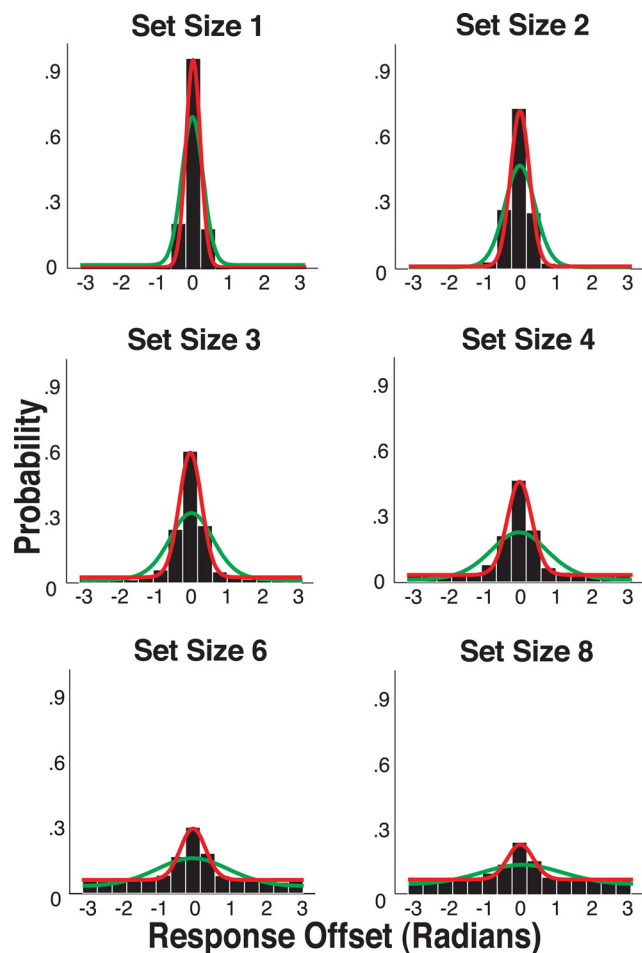
Figure 2. Response offset histograms for each set size in experiment 1. Response offsets were calculated as the deviation of the participants' response from the value of the target orientation. Each histogram was fitted using the parameters of the discrete-resource model (red) and the flexible-resource model (green).

Ruling out encoding as a limiting factor for performance
One essential design issue was to ensure that performance in this procedure was not limited by encoding time because this could provide an alternative explanation for putative item limits during the subsequent storage phase of the task. Given that no masks were presented, previous studies suggest that the 200 ms exposure duration for the sample arrays was more than enough time for encoding of the largest eight item arrays. Nevertheless, we performed a simple control experiment ($n = 8$) in which we contrasted performance with the eight item arrays in two conditions. In the simultaneous condition, all eight items were simultaneously presented for 200 ms just as in the primary design. In the sequential condition, the eight items were presented four at a time (200 ms exposure duration for each presentation) with a 100 ms interstimulus interval. We reasoned that, if 200 ms were not enough time to encode all eight items in the simultaneous condition, then performance should be superior in the sequential condition. Performance was equivalent in the two conditions. P_{mem} in the simultaneous (0.36) and sequential (0.31) was statistically equivalent ($t_{(7)} = 0.614$; $p = 0.56$). Likewise, SD in the simultaneous (23.44) and sequential (23.94) conditions was statistically equivalent ($t_{(7)} = -0.071$; $p = 0.945$). Thus, encoding was not a limiting factor for performance in this procedure.

Assessing the potential effects of mislocalized targets
Recently, Bays et al. (2009) have suggested an alternative explanation of the mixture distributions first reported by Zhang and Luck (2008). Bays et al. (2009) questioned whether the flat component of this mixture distribution was really caused by trials in which no information had been retained about the probed item. Instead, Bays et al. (2009) argued that the flat distribution may be the result of trials in which observers mislocalized the probed object and reported the value of a different item in the display. Because the orientations for each item varied randomly with respect to one another, this kind of mislocalization would yield a random (i.e., flat) distribution of response errors relative to the probed item. Indeed, Bays et al. (2009) found that a substantial proportion of responses in their study could be attributed to the erroneous report of nontarget values. To test this possibility in our own data, we calculated the response offset between each response and each of the nontarget values from the same trial. If subjects had consistently reported distractor values during the experiment, then this analysis should reveal a central tendency in the response error histogram, showing that subjects' reported nontarget values at greater than chance levels. Figure 3*A* shows the aggregate distribution of distractor response values for each set size. The mixture model did not provide a good fit of the data at any set size as revealed by Kolmogorov–Smirnov tests. Specifically, there was no evidence of central tendency in these error plots, suggesting that observers did not consistently report distractor values in this experiment. Recently, Umemoto et al. (2010) reported a similar empirical pattern in which there was little evidence for consistent reports of distractor values. Likewise, Gold et al. (2010) found a flat distribution when response offsets were calculated relative to distractor values (but see Bays, 2010). One concern, however, is that noise in the data may have made it difficult to detect the central tendency hypothesized by the mislocalization hypothesis. That is, given that each response was randomly related to all but one value from each display, it may have been difficult to detect the Gaussian component within the overall distribution. To assess the sensitivity of the analysis, we generated response histograms in which each response was related to all distractor values as well as the single probed target value from each trial (Fig. 3*B*). If the analysis is insensitive to the Gaussian component when there is a preponderance of responses that are randomly related to the target value, then the mixture model should provide a poor fit, especially at larger set sizes where the proportion of target-related responses is smallest. On the contrary, the mixture model (gray lines) provided an excellent fit of distributions, suggesting that the mixture model is sensitive to the nonrandom component of the distribution even when the distribution is composed predominately of random responses. This analysis confirms our conclusion that the uniform component of the target distribution is not attributable to mislocalizations, but rather to a failure to store the probed item.

The relationship between precision and on-line storage
Because discrete-resource models predict that no storage in WM is possible past a fixed item limit, they also predict that resources supporting resolution in WM cannot be distributed to any additional items once the item limit has been exceeded. Thus, discrete-resource models predict that WM resolution (operationalized by the SD parameter in the mixture model) will follow a bilinear function across set sizes, with resolution for the stored items reaching a stable plateau once the item limit has been exceeded. To test this prediction, we examined whether or not the resolution by set size function was well described by a bilinear

function at both the group and individual subject level. Such a bilinear function would be comprised of a linear function of positive slope at early set sizes (with slope corresponding to the rate at which SD increased with set size) that intersected with another linear function of zero slope that illustrated stable resolution across the larger set sizes. The set size over which these two lines intersected defined the inflection point of the function. This bilinear function was fitted to both the aggregate and individual subject data. The inflection point of the bilinear function was predicted to correlate with the individual item limit for each observer (i.e., P_{mem} multiplied by set size).

Figure 4A is a plot of the aggregate SD data collected in experiment 1. The fitted bilinear function (gray) provided an excellent description of the resolution by set size function (black) across set sizes ($R^2 = 0.996$; $p < 0.0001$) and revealed an asymptote at 3.20 objects. SD values rose monotonically as set size increased until set size 4, after which a stable asymptote was apparent. This impression was confirmed by conducting paired t tests on the SD values obtained from individual subject data (individual fits described below): (set size 1–2, $t_{(40)} = -9.63$, $p < 0.001$; set size 2–3, $t_{(40)} = -8.04$, $p < 0.001$; set size 3–4, $t_{(40)} = -1.95$, $p = 0.059$; set size 4–6, $t_{(40)} = 0.167$, $p = 0.869$; set size 6–8, $t_{(40)} = 0.724$, $p = 0.473$). Thus, the resolution by set size function derived from the aggregate data was well described by a bilinear function, as predicted by the discrete-resource model.

To extend this result, we examined whether variations in discrete item limits across observers would predict the specific set size at which mnemonic resolution achieved asymptote for each observer. This was accomplished by fitting a bilinear function to individual SD by set size plots and correlating individual SD inflections with P_{mem} (the probability that the critical item was stored, or $1 - P_{fail}$) for each observer. Four subjects were removed from the analysis because their inflection points fell >4 SDs above the mean of the remaining observations; nevertheless, the correlation reported below was also statistically reliable even with these outliers included. The bilinear function provided a strong fit to SD by set size functions for each individual observer (average $R^2 = 0.55$). In addition, the data confirmed the predicted correlation between individual item limits (estimated using P_{mem} for set size 8) and the set size at which SD reached asymptote (Fig. 4B) ($R^2 = 0.657$; $t_{(40)} = 8.76$; $p < 0.0001$). Strong correlations were also observed when item
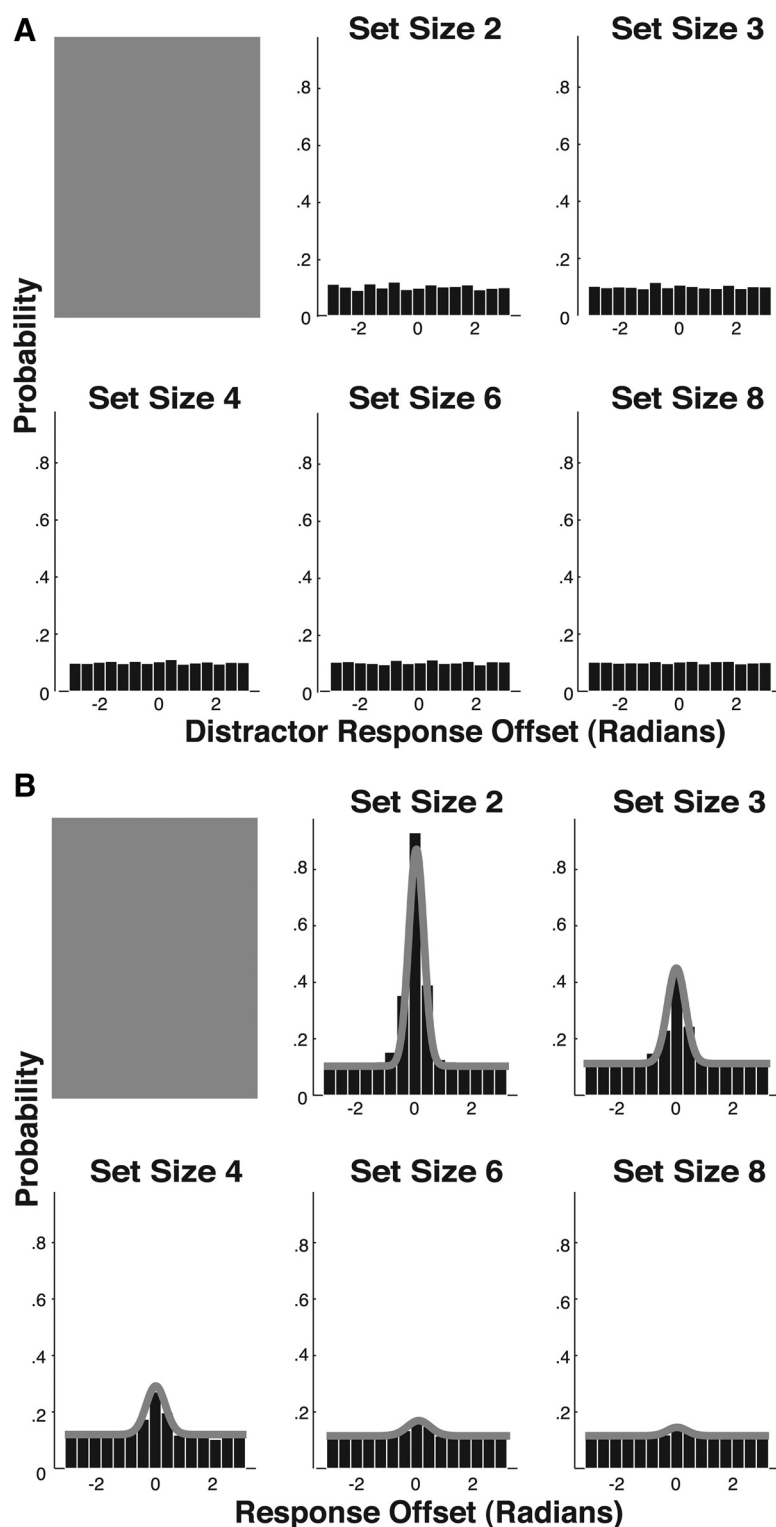
Figure 3. *A*, Distractor response offset histograms. Responses for each trial were subtracted from all distractor values within each display and binned according to set size. The absence of central tendency in the distractor offset distributions suggests that mislocalizations were not prevalent. *B*, To ensure that the mislocalization analysis was sensitive to the presence of a central tendency, indicative of nonzero information responses, the response offsets from target values were added to the distractor distribution in *A*. The mixture model provided a good fit for these data, suggesting that the model is sensitive to the presence of central tendency, albeit smaller than when target values alone are used.

individual item limits were estimated using the data from set sizes 4 and 6 ($R^2 = 0.365$, $p < 0.0001$, and 0.506, $p < 0.0001$, for set sizes 4 and 6, respectively). We note here that this correlation is also informative regarding the alternative "mislocalization" hy-
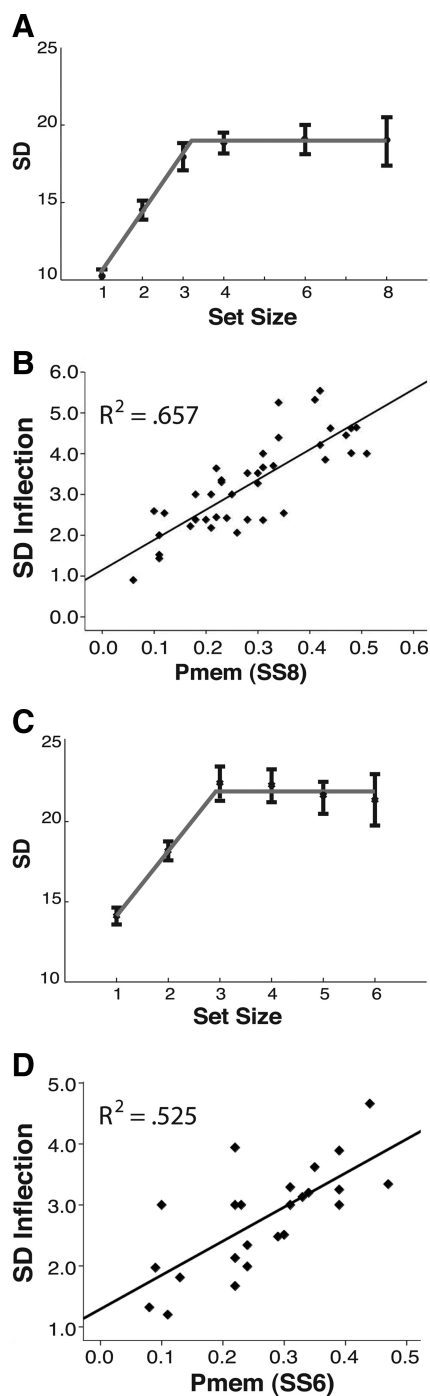
Figure 4. Bilinear fits and individual differences analysis. The resolution by set size functions (black) from experiment 1 (**A**) and experiment 2 (**C**) were fitted with a bilinear function (gray). The correlation ($p < 0.001$) between individual item limits and asymptotes in precision in experiments 1 (**B**) and 2 (**D**).

pothesis referenced above. If variations in P_{mem} simply reflected variations in the likelihood of mislocalization errors, as presumed by the interpretation offered in Bays et al. (2009), then there would be no reason to predict a correlation with between P_{mem} and the set size at which WM resolution reached asymptote. Thus, the discrete-resource model provides a more internally consistent interpretation of the overall findings.

To summarize, experiment 1 yielded three primary conclusions. First, the distributions of response errors in this orientation

WM task were well described by a mixture model that presumed a fixed item limit, such that observers stored a relatively small subset of the items in the sample array, while maintaining no information about the remaining items. This finding replicates the findings of Zhang and Luck (2008). Second, the fine-grained manipulation of set size enabled a sensitive test of an additional prediction of the discrete-resource model, that mnemonic resolution should achieve a stable asymptote once putative item limits are exceeded. This prediction was confirmed at both the aggregate and the individual subject levels. Third, the item limit determined for each observer was strongly predictive of the set size at which the resolution by set size function reached asymptote. This finding confirms a clear prediction of discrete-resource models, while also demonstrating the predictive validity of the parameters derived from the mixture model approach. When individual item limits are exceeded, mnemonic resolution reaches a stable plateau.

Experiment 2: relating the resolution by set size function to an on-line neural measure of WM storage

Our goal in experiment 2 was to examine the relationship between the empirical pattern observed in experiment 1 and an established neural index of storage in visual WM. Thus, we replicated experiment 1 while measuring CDA, a sustained electrophysiological response that is delay specific and strongly predictive of the number of representations held in visual WM. The CDA exhibits monotonic increases in amplitude as set size increases and plateaus at a set size that predicts individual WM capacity (Vogel and Machizawa, 2004; Vogel et al., 2005). Establishing a direct link between a delay-specific neural index of WM storage and the shape of the resolution by set size function addresses well known ambiguities in the interpretation of purely behavioral measures of WM storage. For example, the effect of set size on behavioral measures of WM storage could reflect differences in the level of interference between competing representations during the retrieval of the critical item, rather than true differences in the probability of successful storage. In addition, behavioral measures of WM storage could be contaminated by contributions from other memory systems (e.g., latent episodic traces in long-term memory) that are separable from WM but still capable of affecting the accuracy of behavioral reports. Thus, linking on-line neural markers of storage to the observed changes in WM resolution across set sizes would solidify the role of changes in on-line storage per se in determining the precision of representations in WM.

Model fitting

The discrete-resource model and the flexible-resource model were fitted to the aggregate distributions of response offsets for each set size (Fig. 5). Both the mixture model, representing the predictions of the discrete-resource model (red), and the flexible-resource model were effective in explaining variance in response distributions for each set size [discrete-resource R^2 values (all values of $p < 0.0001$): 0.988 (SS1), 0.988 (SS2), 0.976 (SS3), 0.976 (SS4), 0.952 (SS5), 0.893 (SS6); flexible-resource R^2 values (all values of $p < 0.001$): 0.92 (SS1), 0.80 (SS2), 0.74 (SS3), 0.67 (SS4), 0.66 (SS5), 0.61 (SS6)]. Kolmogorov–Smirnov tests, however, revealed a significant difference between the predicted values of the flexible-resource model and the actual distributions at every set size, whereas the test revealed no difference between the predictions of the discrete-resource model and their fits to each distribution.

The discrete-resource model was superior to the flexible-resource model in its ability to account for the distributions of
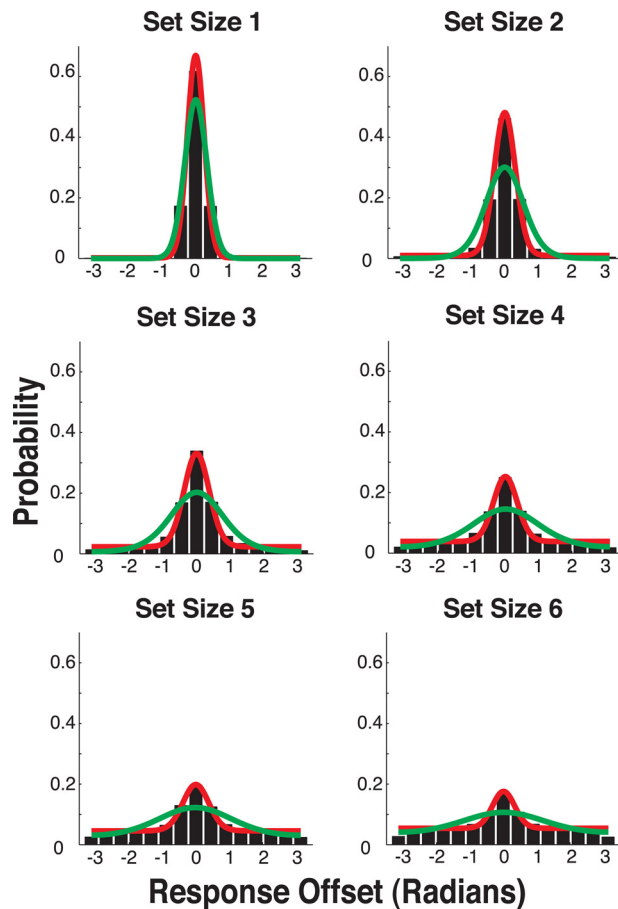
Figure 5. Response offset histograms for each set size in experiment 2. Response offsets were calculated as the deviation of the participants' response from the value of the target orientation. Each histogram was fitted with the parameters of the discrete-resource model (red) and the flexible-resource model (green).

response offsets. This replicates the findings of Zhang and Luck (2008) and experiment 1, and suggests that observers could store a subset of the items in the sample array, while maintaining no information about the remaining items.

Replication of the resolution by set size function observed in experiment 1
We used the same analytic procedure as in experiment 1 to determine whether an asymptote was observed in SD at a relatively low set size. In addition, we reexamined whether asymptotes in individual SD by set size functions predicted individual capacity estimates. Figure 4*C* is a plot of the aggregate SD data collected in experiment 2. The fitted bilinear function (gray) provided an excellent fit for the SD by set size function (black) across set sizes ($R^2 = 0.986$; $p < 0.0001$) and revealed an asymptote at 2.9 objects. Also replicating the findings of experiment 1, SD values rose monotonically from set size 1 to 3 and achieved asymptote at larger set sizes (set size 1–2, $t_{(29)} = -13.29$, $p < 0.001$; set size 2–3, $t_{(29)} = -4.94$, $p < 0.001$; set size 3–4, $t_{(29)} = 0.13$, $p = 0.45$; set size 4–5, $t_{(29)} = 0.72$, $p = 0.24$; set size 5–6, $t_{(29)} = 0.038$, $p = 0.49$). Finally, the key finding from experiment 1 was also replicated; estimates of the item limit for each observer strongly predicted the set size at which WM resolution reached asymptote for each subject (Fig. 4*D*) ($R^2 = 0.525$; $t_{(28)} = 5.554$; $p < 0.0001$) and the bilinear function again provided a strong fit to SD by set size functions for each individual observer (average $R^2 = 0.65$).
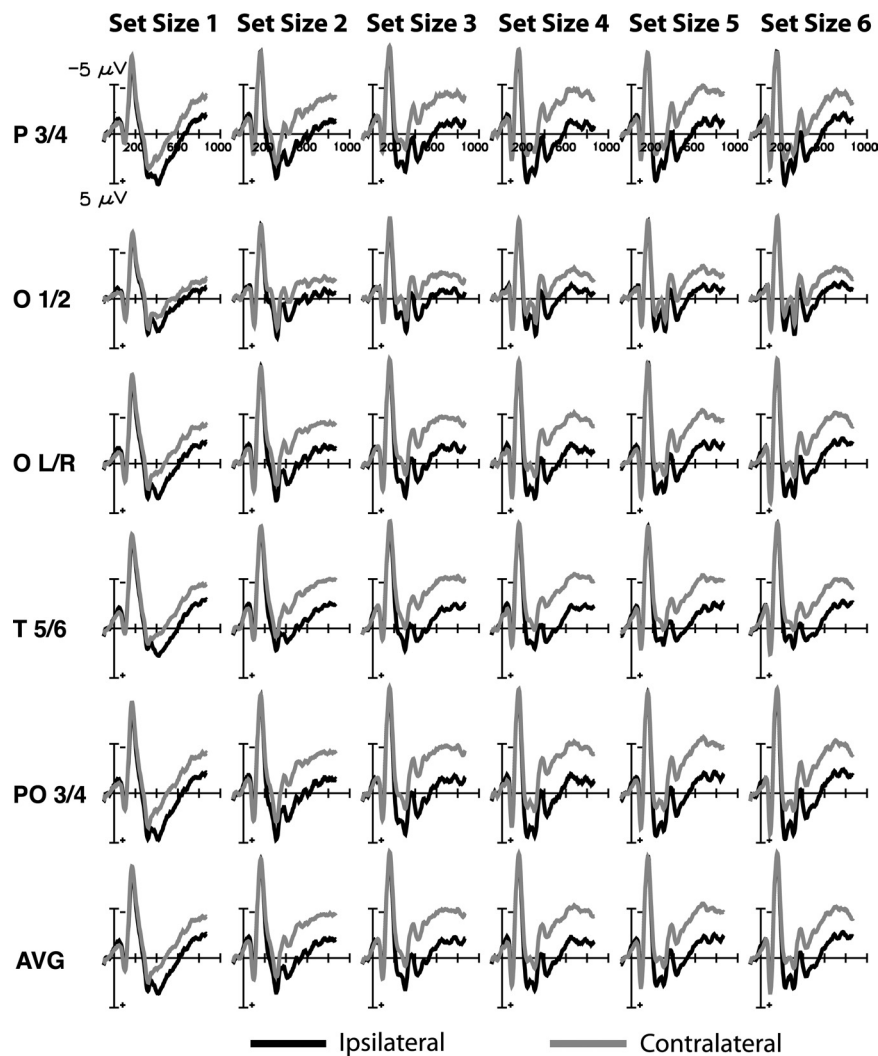
Linking delay-specific neural activity with mixture model parameters
CDA was apparent at electrodes P3/4, PO3/4, O1/2, OL/R, and T5/6 (Fig. 6). We plotted difference waves (subtracting ipsilateral amplitudes from contralateral amplitudes; see Fig. 7*A* for average difference waves) for each set size at every electrode and observed a monotonic increase in mean amplitude as a function of set size until set size 3, at which point amplitude reached an apparent asymptote (Fig. 8*A*) (set size 1–2, $t_{(29)} = 3.089$, $p < 0.01$; set size 2–3, $t_{(29)} = 2.372$, $p < 0.05$; set size 3–4, $t_{(29)} = 0.892$, $p = 0.38$; set size 4–5, $t_{(29)} = -1.017$, $p = 0.31$; set size 5–6, $t_{(29)} = -1.647$, $p = 0.11$). As the data will show, CDA amplitude across set sizes was well described by a bilinear function.

To determine whether CDA predicts individual WM capacity and the set size at which WM resolution reached asymptote (i.e., SD inflection), we calculated the difference in the mean amplitude of the CDA between set sizes 4 and 2 (Vogel and Machizawa, 2004; Vogel et al., 2005). The assumption for using this contrast is that individuals with lower capacities should have a relatively small difference between set sizes 2 and 4 (because set sizes 2 may approach their capacity limit), whereas individuals with higher capacities should show a larger difference between set sizes 2 and 4. Thus, this neural measure of WM storage should predict both individual WM capacity, as well as the set size at which WM resolution achieved asymptote for each observer. In line with these predictions, the difference in CDA amplitude at set sizes 4 and 2 (averaged across four electrodes, OL/R, P3/4, PO3/4, T5/6, from 400 to 800 ms after the onset of the sample array) reliably predicted both SD inflection ($R^2 = 0.44$; $p < 0.001$) (Fig. 7*B*) and P_{mem} ($R^2 = 0.22$; $p < 0.01$) (Fig. 7*C*). These correlations were also reliable at each individual electrode pair used to measure the CDA. We also tested for a relationship between the single parameter inferred from the flexible-resource model and the CDA. To operationalize capacity in the context of the flexible-resource model, we calculated the SD of a single Gaussian distribution that was fit to the response error histograms from each set size; this fits the assumption of the flexible-resource model that individuals with greater resources should have better precision (i.e., a lower SD) at any given set size. In terms of the CDA, individuals with more resources should show a larger change in amplitude from set size 2 to set size 4, given that those with more resources would have more left over to distribute when additional items were added into the sample array. We found no such relationship between precision at any set size and CDA ($p > 0.36$). Therefore, as we have already shown, the flexible-resource model is inferior to the discrete-resource model in explaining variations in the delay-specific neural response.

To corroborate our analysis of the CDA, we also fit a bilinear function to each individual's CDA data to examine whether or not the inflection point of this function would predict individual WM capacity and the set size at which WM resolution reached asymptote. The bilinear functions provided an excellent fit for the individual CDA data (mean $R^2 = 0.53$). Moreover, the inflection point of this bilinear function predicted WM capacity ($R^2 = 0.30$; $p < 0.01$) (Fig. 8*C*) and the inflection point in the SD by set size function ($R^2 = 0.45$; $p < 0.001$) (Fig. 8*B*). Thus, both methods for characterizing storage-related electrophysiological responses during the delay period led to the same conclusion: storage-related neural activity strongly predicted individual WM capacity as well as the set size at which WM resolution reached a stable plateau for each observer. One possible concern for this analysis is that the CDA by set size function is not perfectly bilinear; instead, there is a noticeable drop in CDA amplitude at the highest set

Figure 6. Contralateral and ipsilateral activity in response to each set size in experiment 2. Waveforms were time locked to the onset of the memory array and are presented from all parietal and occipital electrodes.

$t_{(29)} = 6.162$, $p < 0.001$; set size 2–3, $t_{(29)} = 2.372$, $p < 0.05$; set size 3–4, $t_{(29)} = 0.892$, $p = 0.38$; set size 4–5, $t_{(29)} = -1.017$, $p = 0.31$; set size 5–6, $t_{(29)} = -1.647$, $p = 0.11$); thus, the N2pc by set size function was well described by a bilinear function. To examine the relationship between the N2pc waveform and WM performance, we modeled each observer's N2pc by set size data with a bilinear function. Just as with the CDA, the N2pc by set size function showed an excellent fit with the bilinear model (mean $R^2 = 0.59$). Moreover, the inflection point of this function, estimated for each individual observer, was a strong predictor of both WM capacity estimates ($R^2 = 0.25$; $p < 0.01$) (Fig. 9C) and the set size at which WM resolution reached asymptote ($R^2 = 0.33$; $p < 0.01$) (Fig. 9B). Thus, this early marker of visual selection further corroborates the conclusions gleaned from the CDA data. Neural activity that strongly predicts individual WM capacity also predicts the set size at which WM resolution reaches asymptote.
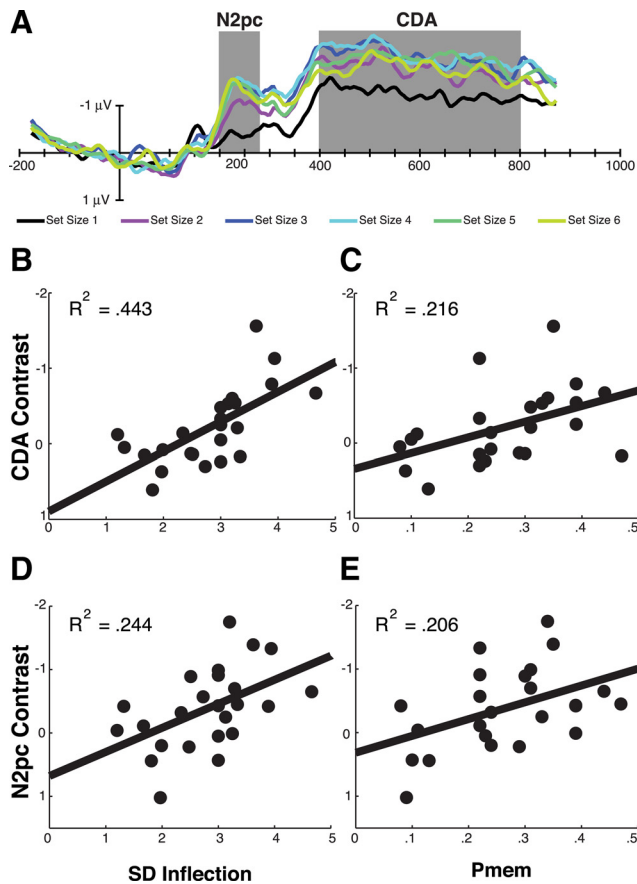
To summarize our findings, the behavioral and electrophysiological results provide strong converging evidence in support of discrete item limits in WM capacity. Analysis of response offsets during an orientation WM task replicated previous findings that observers were able to maintain nonzero information about only a few items from the sample display, while retaining no information about the remaining items (Zhang and Luck, 2008). We observed monotonic declines in WM resolution as the number of items to be stored increased. These declines in WM resolution, however, did not continue across the full range of set sizes. Instead, we found that WM resolution reached a stable asymptote after individual item limits had been exceeded. This resolution by set size function confirms a clear prediction of discrete-resource models of WM, in which it is assumed that increases in set size have no effect on what is stored in WM once a fixed item limit has been exceeded. In addition, this empirical pattern showed a tight link with a sensitive neural measure of on-line storage in WM. A contralateral electrophysiological response that was sustained throughout the delay period of the WM task predicted individual WM capacity as well as the specific set size at which WM resolution reached asymptote. Thus, our data provide a direct link between a well validated neural measure of WM storage and a resolution by set size function that confirms a clear prediction of discrete-resource models of WM capacity.

Discussion

Although a growing body of behavioral and neuroscience research has continued to hone our understanding of the neural systems that mediate storage in visual WM, fundamental debates remain regarding the nature of capacity limits in this on-line store. The present work focused on the core question of whether WM capacity is constrained by a discrete item limit, as postulated by discrete-resource models, or whether WM resources can be

sizes. Although we cannot fully explain the underlying reasons for this drop, it has also been reported in the behavioral literature when WM capacity is measured with large set sizes (Cusack et al., 2009). Although the precise reasons for this drop at larger set sizes have not yet been fully documented, our intent was to demonstrate how regularities in shape of the CDA by set size function are a robust predictor of the resolution by set size function as well as individual WM capacity.

Finally, as the waveforms in Figure 7A illustrate, a similar profile of set size effects was also apparent 200–275 ms after the onset of the target array. This transient negative wave appears to be the N2pc wave. Previous work has shown that the N2pc is evoked contralateral to the hemifield to which covert spatial attention is shifted (Luck et al., 1997). In addition, more recent studies have found that the amplitude of the N2pc is reliably correlated with the number of relevant targets that are selected from a visual array (Drew and Vogel, 2008). Indeed, the difference in N2pc amplitude at set sizes 4 and 2 (averaged across four electrodes, OL/R, P3/4, PO3/4, T5/6, from 200 to 275 ms after the onset of the sample array) reliably predicted both SD inflection ($R^2 = 0.24$; $p < 0.01$) (Fig. 7D) and P_{mem} ($R^2 = 0.21$; $p < 0.01$) (Fig. 7E). Here, we observed a monotonic increase in N2pc amplitude as set size increased from one to three items, and then a clear asymptote from set size 3 until set size 6 (Fig. 9A) (set size 1–2,
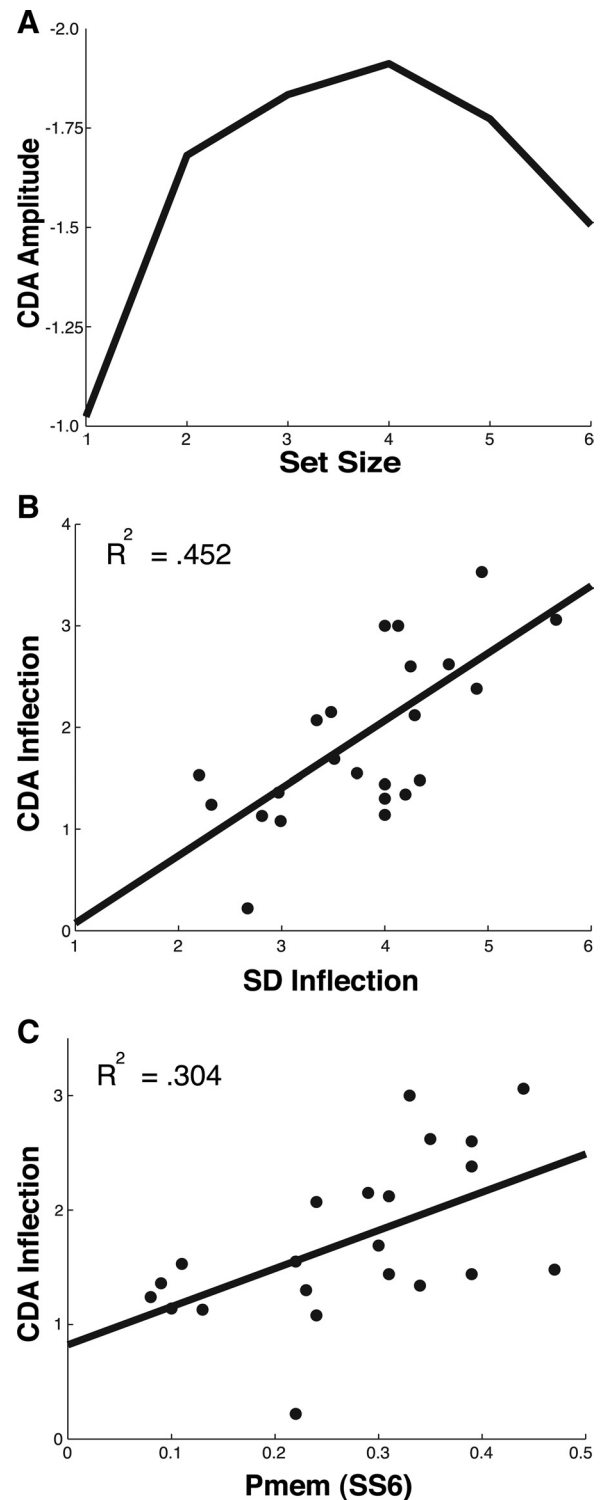
Figure 7. *A*, ERP grand averaged difference waves. The gray bars indicate the temporal windows used to measure N2pc and CDA amplitude. *B*–*E*, Individual differences analyses between behavioral measures (precision asymptotes and capacity estimates) and electrophysiological measures (N2pc and CDA). Our electrophysiological measure of on-line maintenance (CDA) strongly predicted both asymptotes in precision (*B*) ($p < 0.001$) and individual capacity estimates (*C*) ($p < 0.01$). N2pc amplitudes also predicted both asymptotes in precision (*D*) ($p < 0.05$) and individual capacity estimates (*E*) ($p < 0.05$).

Figure 8. *A*, Plot of average CDA amplitude for each set size. The inflection point of each observer's CDA by set size function was calculated to examine whether this neural measure of individual item limits would predict the inflection point in the resolution by set size function (*B*) ($p < 0.01$) and individual capacity estimates (*C*) ($p < 0.01$).

divided flexibly without any item limit, as asserted by flexible-resource models. To discriminate between these models, we acquired behavioral and electrophysiological measures of WM function as the number of items to be stored was manipulated. The rationale for this design was motivated by previous observations that WM resolution declines as the number of items stored in memory increases (Bays and Husain, 2008; Zhang and Luck, 2008; Barton et al., 2009). Although both discrete- and flexible-resource models can accommodate this empirical pattern, only discrete-resource models predict that WM resolution will reach a stable plateau at relatively low set sizes, when discrete item limits are exceeded.

In line with the discrete-resource model, the behavioral data from both experiments 1 and 2 revealed monotonic declines in precision from set size 1 up to 3, followed by a clear plateau across the remaining set sizes. Thus, when the putative item limit had been exceeded, WM resolution reached a stable asymptote. This plateau in resolution corroborates the assumption of an item limit because no additional change in on-line memory loads is postulated when putative item limits are exceeded. Experiment 2 provided additional converging evidence for this conclusion by providing a delay-specific neural index of on-line storage for each observer in the experiment. Here again, when the active content of WM was operationalized via sustained load-dependent neural

activity, we found robust correlations between individual capacity estimates and the specific set size at which WM resolution reached asymptote. Thus, experiment 2 provides strong evidence that the apparent plateau in WM resolution was a consequence of active storage per se. In addition, this study demonstrates a direct link between the two primary parameters of the mixture
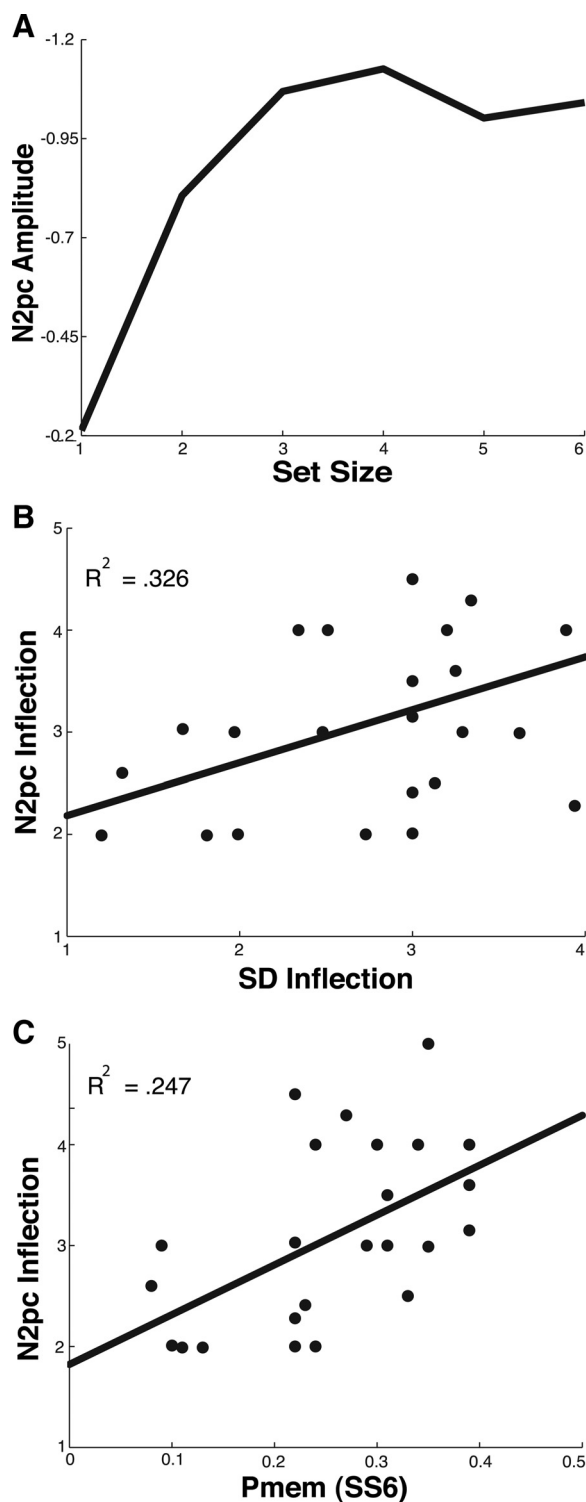
Figure 9. *A*, Plot of average N2pc amplitude for each set size. The inflection point in the N2pc by set size function was calculated for each observer, and this measure strongly predicted both asymptotes in precision (*B*) ($p < 0.01$) and individual capacity estimates (*C*) ($p < 0.01$).

model (P_{mem} and SD) and an established neural index of on-line storage; this convergence of results bolsters confidence in the interpretation of both the behavioral and neural measures of WM performance.

These findings fall in line with the hybrid model described in Introduction, in which a discrete set of slots constrains the allocation of mnemonic resources for maintaining high-fidelity rep-

resentations. Once the fixed slot limit has been exceeded, the resources for resolution cannot be allocated to any additional items; thus, resolution remains stable as set size grows larger because the content of memory does not change. Moreover, the hypothesis that fixed slots constrain the allocation of a separate resource that determines resolution dovetails with recent evidence that number and resolution in WM are implemented in distinct cortical regions and associated with distinct aspects of individual memory ability. Xu and Chun (2006) provided support for this dissociation by demonstrating that brain activation in the inferior intraparietal sulcus (IPS) tracked the number of objects stored in working memory regardless of complexity, whereas activity in the superior IPS and lateral occipital complex (LOC) was sensitive to the complexity of the stored objects. Given that visual details must be represented more precisely to support behavioral performance with complex objects, we postulate that the superior IPS and LOC may support the maintenance of the visual details that determine mnemonic resolution, whereas the inferior IPS determines how many of those representations can be stored. Finally, this inferior IPS locus corresponds well with the presumed source of the CDA measured in experiment 2 (Mc-Collough et al., 2007) and identified by previous imaging studies as sensitive to the number of items stored in WM (Todd and Marois, 2004, 2005).

The distinction between number and resolution in WM is also supported by studies that have examined individual differences in these two aspects of WM ability (Awh et al., 2007; Vogel and Awh, 2008; Fukuda et al., 2010). Awh et al. (2007) examined individual differences in the number and resolution of representations stored in WM and found no correlation between these behavioral scores (despite evidence showing the reliability of these measures). Likewise, Fukuda et al. (2010) performed a latent variable analysis of WM ability and found fully orthogonal factors corresponding to the number and resolution of the representations in visual WM. Moreover, Fukuda et al. (2010) found that, although the number factor exhibited a robust correlation with fluid intelligence ($r = 0.66$), there was no apparent predictive relationship between WM resolution and fluid intelligence ($r = -0.05$). Thus, both neural and behavioral data provide strong support for a two-factor model in which number and resolution correspond to distinct aspects of WM ability. In turn, this two-factor model fits naturally with our proposal that the allocation of resources supporting WM resolution may be constrained by a separate discrete resource that determines the number of items held on-line in WM.

The present results are also conducive to so-called sensory recruitment models of WM maintenance (for review, see Awh and Jonides, 2001; Jonides et al., 2005; Postle, 2006), which propose that active maintenance in WM is mediated in part by the same regions that support the sensory encoding of the stored items. This sensory recruitment hypothesis has gained new support from human imaging studies that have examined the pattern of activity in early visual cortex while observers held orientation values in visual WM. These studies reveal a sustained pattern of activity in the primary visual cortex that represents the attended aspects of multifeature stimuli (Serences et al., 2009) and is qualitatively similar to the patterns of activity observed during sensory encoding (Ester et al., 2009; Harrison and Tong, 2009; Serences et al., 2009). Our working hypothesis is that precision in visual WM is determined by the integrity of these sustained patterns of sensory activity. At the same time, activity in distinct cortical regions (e.g., inferior IPS) determines the number of representations that can be simultaneously maintained in WM

(Vogel and Machizawa, 2004; Todd and Marois, 2005; McCollough et al., 2007). Thus, the monotonic declines in resolution that are observed when additional items are stored in WM may reflect a basic limit in this collaboration between IPS and sensory regions, such that increasing storage loads compromise the integrity of the sustained sensory activity that represents each stored item. This hypothesis is compatible with computational models (Lisman and Idiart, 1995; Raffone and Wolters, 2001) in which WM storage entails the maintenance of asynchronous oscillatory patterns in the neural populations that represent each of the stored items. To avoid interference between individuated representations, it is necessary for the oscillatory activity associated with stored items to be held out of phase with one another. Thus, declines in WM resolution with increasing memory loads could reflect increased difficulty in maintaining desynchronized oscillatory patterns across different populations of sensory neurons. Once item limits are exceeded, however, no additional oscillatory patterns are added and no additional declines in WM resolution are observed.

To conclude, we examined the core claim of discrete-resource models that WM capacity is constrained by a fixed item limit (Luck and Vogel, 1997; Awh et al., 2007; Rouder et al., 2008; Zhang and Luck, 2008). In line with this hypothesis, dynamic changes in WM resolution across set sizes reached a stable plateau at the same set size that individual item limits were exceeded. These data provide direct links between psychophysical and neural measures of on-line storage capacity and demonstrate the predictive validity of these measures. Thus, although there is continued debate regarding the validity of the item limit assumption (Wilken and Ma, 2004; Bays and Husain, 2008), the empirical patterns documented here provide new evidence regarding the internal consistency and predictive power of discrete-resource models.

References

Awh E, Jonides J (2001) Overlapping mechanisms of attention and spatial working memory. Trends Cogn Sci 5:119–126.

Awh E, Barton B, Vogel EK (2007) Visual working memory represents a fixed number of items regardless of complexity. Psychol Sci 18:622– 628.

Barton B, Ester EF, Awh E (2009) Discrete resource allocation in visual working memory. J Exp Psychol Hum Percept Perform 35:1359–1367.

Bays P (2010) Precision versus capacity of working memory in schizophrenic and healthy individuals. Arch Gen Psychiatry Online 16.

Bays PM, Husain M (2008) Dynamic shifts of limited working memory resources in human vision. Science 321:851– 854.

Bays PM, Catalao RFG, Husain M (2009) The precision of visual working memory is set by allocation of a shared resource. J Vis 9:7.1–11.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433– 436.

Cusack R, Lehmann M, Veldsman M, Mitchell DJ (2009) Encoding strategy and not visual working memory capacity correlates with intelligence. Psychon Bull Rev 16: 641– 647.

Drew T, Vogel EK (2008) Neural measures of individual differences in selecting and tracking multiple moving objects. J Neurosci 28:4183– 4191.

Ester EF, Serences JT, Awh E (2009) Spatially global representations in human primary visual cortex during working memory maintenance. J Neurosci 29:15258 –15265.

Fukuda K, Vogel E, Mayr U, Awh E (2010) Quantity not quality: the relationship between fluid and working memory capacity. Psychon Bull Rev 17:673– 679.

Gold JM, Hahn B, Zhang WW, Robinson BM, Kappenman ES, Beck VM, Luck SJ (2010) Reduced capacity by spared precision and maintenance of working memory representations in schizophrenia. Arch Gen Psychiatry 67:570–577.

Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. Nature 458:632– 635.

Jonides J, Lacey SC, Nee DE (2005) Processes of working memory in mind and brain. Curr Dir Psychol Sci 14:2–5.

Lisman JE, Idiart MA (1995) Storage of 7 +/− 2 short-term memories in oscillatory subcycles. Science 267:1512–1515.

Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. Nature 390:279–281.

Luck SJ, Girelli M, McDermott MT, Ford MA (1997) Bridging the gap between monkey neurophysiology and human perception: an ambiguity resolution theory of visual selective attention. Cogn Psychol 33:64 – 87.

McCollough AW, Machizawa MG, Vogel EK (2007) Electrophysiological measures of maintaining representations in visual working memory. Cortex 43:77–94.

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437– 442.

Postle BR (2006) Working memory as an emergent property of the mind and brain. Neuroscience 139:23–38.

Raffone A, Wolters G (2001) A cortical mechanism for binding in visual working memory. J Cogn Neurosci 13:766–785.

Rouder JN, Morey RD, Cowan N, Zwilling CE, Morey CC, Pratte MS (2008) An assessment of fixed-capacity models of visual working memory. Proc Natl Acad Sci U S A 105:5975–5979.

Serences JT, Ester EF, Vogel EK, Awh E (2009) Stimulus-specific delay activity in human primary visual cortex. Psychol Sci 20:207–214.

Todd JJ, Marois R (2004) Capacity limit of visual short-term memory in human posterior parietal cortex. Nature 428:751–754.

Todd JJ, Marois R (2005) Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. Cogn Affect Behav Neurosci 5:144–155.

Umemoto A, Drew T, Ester EF, Awh E (2010) A bilateral advantage for storage in visual working memory. Cognition 117:69–79.

Vogel EK, Awh E (2008) How to exploit diversity for scientific gain: using individual differences to constrain cognitive theory. Curr Dir Psychol Sci 17:171–176.

Vogel EK, Machizawa MG (2004) Neural activity predicts individual differences in visual working memory capacity. Nature 428:748–751.

Vogel EK, Luck SJ, Shapiro KL (1998) Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. J Exp Psychol Hum Percept Perform 24:1656–1674.

Vogel EK, McCollough AW, Machizawa MG (2005) Neural measures reveal individual differences in controlling access to working memory. Nature 438:500–503.

Wilken P, Ma WJ (2004) A detection theory account of change detection. J Vis 4:1120–1135.

Xu Y, Chun MM (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. Nature 440:91–95.

Zhang W, Luck SJ (2008) Discrete fixed-resolution representations in visual working memory. Nature 453:233–235.