

Storage in Visual Working Memory Recruits a Content-Independent Pointer System



William Thyer^{1,2}, Kirsten C. S. Adam^{3,4} , Gisella K. Diaz^{1,2},
 Itzel N. Velázquez Sánchez^{1,2} , Edward K. Vogel^{1,2}, and
 Edward Awh^{1,2}

¹Department of Psychology, The University of Chicago; ²Institute for Mind and Biology, The University of Chicago;
³Department of Psychology, University of California San Diego; and ⁴Institute for Neural Computation, University of
 California San Diego

Psychological Science
 1–15

© The Author(s) 2022

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09567976221090923

www.psychologicalscience.org/PS



Abstract

Past work has shown that storage in working memory elicits stimulus-specific neural activity that tracks the stored content. Here, we present evidence for a distinct class of load-sensitive neural activity that indexes items without representing their contents per se. We recorded electroencephalogram (EEG) activity while adult human subjects stored varying numbers of items in visual working memory. Multivariate analysis of the scalp topography of EEG voltage enabled precise tracking of the number of individuated items stored and robustly predicted individual differences in working memory capacity. Critically, this signature of working memory load generalized across variations in both the type and number of visual features stored about each item, suggesting that it tracked the number of individuated memory representations and not the content of those memories. We hypothesize that these findings reflect the operation of a capacity-limited pointer system that supports on-line storage and attentive tracking.

Keywords

visual working memory, spatiotemporal pointers, electroencephalogram, open data, open materials

Received 7/26/21; Revision accepted 2/24/22

A central goal of cognitive neuroscience has been to understand the neural underpinnings of working memory (WM), an on-line memory system that is thought to be critical for virtually all forms of intelligent behavior. Significant progress has been made by focusing on stimulus-specific neural activity that tracks the features of the items stored in WM. In both animal and human subjects, WM storage has been shown to elicit sustained activity in neural units or brain regions that are selective for the particular items held in mind (e.g., D'Esposito & Postle, 2015; Funahashi et al., 1989; Fuster & Jervey, 1981; Goldman-Rakic, 1995; Harrison & Tong, 2009; Rademaker et al., 2019; Serences et al., 2009). The motivation for these studies is clear, as they have the potential to elucidate the memory engrams (Poo et al., 2016) that allow people to hold specific ideas in mind.

Nevertheless, a distinct category of studies has focused instead on neural signals that track the number of items stored in WM, rather than the content of those

representations (e.g., Adam et al., 2020; Todd & Marois, 2004; Vogel & Machizawa, 2004; Xu & Chun, 2006). For example, Vogel and Machizawa (2004) used scalp electroencephalogram (EEG) recordings to observe a sustained negative slow wave in posterior electrodes contralateral to the items stored in WM. This contralateral delay activity persists throughout the delay period, reaches a plateau when behavioral estimates of memory capacity are exceeded, and is a robust predictor of individual differences in the capacity of visual WM (Luria et al., 2016). This kind of load-sensitive neural measure has provided insight into how observers control access to this limited on-line workspace (McNab & Klingberg, 2008; Vogel et al., 2005), the role of WM in complex tasks such as multiple-object tracking (Drew

Corresponding Author:

Edward Awh, The University of Chicago, Department of Psychology
 Email: awh@uchicago.edu

& Vogel, 2008) and visual search (Carlisle & Woodman, 2011; Gunseli et al., 2014), and the relationship between WM capacity and other cognitive abilities (Unsworth et al., 2014, 2015).

Although it is clear that load-sensitive neural signals have been potent tools for studying WM, important questions remain regarding the computational role of this class of neural activity. Given past evidence for sustained stimulus-specific neural activity during WM storage, one possibility is that load-sensitive signals index the feature-selective neural activity required for storage. Here, however, we present evidence for neural activity that indexes a qualitatively different cognitive operation from the representation of content per se. There has been long-standing interest in the cognitive operations that support object individuation—the segmentation of objects from the background and from other objects—and the binding of an item’s features into an integrated percept that can be tracked in a dynamic visual scene. Kahneman et al. (1992) proposed the “object file” as a mechanism for registering specific tokens in the visual field to support the continuous tracking of those items through time and space. Likewise, Pylyshyn (2009) described “fingers of instantiation” as a mechanism for indexing visual tokens, thereby enabling perception to unfold over time despite changes in appearance or spatial position. Thus, both theories describe a kind of spatiotemporal pointer system that supports the apprehension and tracking of individuated items while the stored content about each item in memory is maintained via parallel but distinct mechanisms that support the maintenance of each item’s attended features.

Our hypothesis is that load-sensitive neural signals reflect the deployment of these spatiotemporal pointers. Although the pointer construct was developed in the context of attentional tracking tasks, WM storage can also be construed as the sustained deployment of attention toward internal representations (Awh & Jonides, 2001; Chun et al., 2011). Indeed, multiple models of visual WM have embraced the idea of separable neural processes for the storage of content on the one hand and the individuation and binding of those representations on the other (e.g., Balaban et al., 2019; Bouchacourt & Buschman, 2019; Oberauer, 2019; Swan & Wyble, 2014; Xu & Chun, 2009). For example, Swan and Wyble (2014) postulated a neural “binding pool” that serves to link together the multiple features of stored items, supporting their representation as individuated tokens. Likewise, Xu and Chun (2009) argued that object individuation and object identification are realized in independent stages of processing, with distinct cortical regions supporting each function (Xu & Chun, 2009). Thus, there is clear motivation to postulate the existence of load-sensitive neural signals that index a

Statement of Relevance

Working memory is an on-line memory system that is essential for almost all intelligent behaviors. Here, we examined patterns in brain activity that track the number of things that a person is holding in working memory in a given moment. The key insight from this work is that this neural “load signal” is unaffected by changes in both the type of information that is maintained in working memory, as well as the total quantity of information that is contained within each stored item. These findings show that one key limiting factor for storage in working memory is based entirely on the number of items that are maintained in this on-line memory system rather than on the specific details that are stored about those things. Our hypothesis is that the number of items matters because of a limit in the number of distinct entities that can be simultaneously tracked through time and space.

content-independent aspect of WM. Our primary conclusion is that EEG activity measured during WM storage provides evidence of precisely this kind of neural operation.

We used a recently developed multivariate approach that uses the scalp topography of EEG activity to decode the number of individuated items held in visual WM (Adam et al., 2020). Although past work has found univariate signals that index the number of items stored in WM, there are several reasons why multivariate load detection (mvLoad) provides a more powerful test bed for characterizing the properties of load-sensitive neural activity. First, mvLoad is far more sensitive, enabling above-chance tracking of the number of items stored even with single trials of EEG activity. Second, mvLoad analyses reveal a multivariate signature of WM storage that generalizes from the trained data set to novel human observers and across significant variations in task design (e.g., lateralized versus whole-field memory displays); thus, the method is able to isolate load-sensitive activity more decisively than prior approaches. Finally, mvLoad accuracy robustly predicts individual differences in WM capacity, showing that it taps into an integral aspect of this on-line memory system.

We focused on three clear predictions for the properties of load-sensitive neural activity that is separable from the maintenance of specific visual details. First, the activity should precisely track the number of individuated representations that are encoded into memory, independent of variations in stimulus-driven activity. Second, the activity should generate a load signature

that generalizes across the storage of distinct classes of visual information. Third, that signature should generalize across strong variations in the amount of information stored about each item, establishing that it tracks the number of individuated representations rather than the total amount of information stored. To anticipate the results, we found that three experiments using the mvLoad analytic approach confirmed all of these predictions, thereby providing critical new evidence for theories of WM capacity that distinguish between the storage of featural details and the indexing of individuated items within visual WM. We proposed that this content-independent signature of WM load indexes the deployment of spatiotemporal pointers (e.g., Kahneman et al., 1992; Pylyshyn, 2009) that enable the individuation, binding, and monitoring of attended objects.

Method

Subjects

Experiments 1, 2, and 3 included 95 separate data-collection sessions (42 in Experiment 1, 33 in Experiment 2, and 21 in Experiment 3), with 50 unique volunteers participating for monetary compensation (\$15 per hr). A total of 20 volunteers participated in all three experiments, allowing us to implement cross-training analyses across experiments. For subjects who completed multiple experiments, each experiment was done in a separate EEG session. Subjects were between 18 and 35 years old, reported normal or corrected-to-normal visual acuity, and provided informed consent according to procedures approved by The University of Chicago Institutional Review Board. Subjects were recruited via online advertisements and fliers posted on the university campus.

Experiment 1. Our target sample in Experiment 1 was 30 subjects. Forty-two volunteers participated in Experiment 1 (25 female; mean age = 23.8 years, $SD = 4.5$). Nine subjects were excluded from the final sample for the following reasons: We were unable to prepare the subject for EEG ($n = 2$); the subject did not complete enough blocks of the task ($n = 5$); the subject's data were unintentionally overwritten ($n = 1$); or too many trials were rejected because of eye movements (see Eye Movement section, $n = 1$). The final sample size was 33 (20 female; mean age = 24.33 years, $SD = 4.76$). We overshot our target sample size by three because we needed enough subjects to complete all three experiments, and some could not return.

Experiment 2. Our target sample in Experiment 2 was 30 subjects. Thirty-three volunteers participated in Experiment 2 (18 female; mean age = 25.39 years, $SD = 4.30$).

Two subjects were excluded from the final sample because the subject did not complete enough blocks of the task. The final sample size was 31 (18 female; mean age = 25.32 years, $SD = 4.07$). We overshot our target sample size by one because we needed enough subjects to complete all three experiments, and some could not return.

Experiment 3. Our target sample in Experiment 3 was 20 subjects. Twenty volunteers participated in Experiment 3 (13 female; mean age = 25.45 years, $SD = 4.07$). No subjects were excluded from the final sample.

Apparatus

We tested the subjects in a dimly lit, electrically shielded chamber. Stimuli were generated using *PsychoPy* (Peirce et al., 2019). Subjects viewed the stimuli on a gamma-corrected 24-in. LCD monitor (refresh rate = 120 Hz, resolution = 1,080 × 1,920 pixels) with their chins on a padded chin rest at a viewing distance of 75 cm.

Luminance-balanced displays

Stimuli were presented against a mid-gray background (~61 cd/m²). Memory arrays included one to four to-be-remembered items. Ignored placeholder items also appeared in the memory array, so each array had a total of five items. The placeholder items were shown in a shade of gray (red, green, blue [RGB] value = 166, 166, 166) that matched the average luminance of all possible colors in the color set.

Task procedures

All three experiments used a whole-field change-detection task. On each trial, a memory array appeared containing five total items. There were one to four colored items to be remembered, and the remainder of the items were gray placeholder items to balance area and luminance across set-size conditions (see the previous section for more detail). Memory and placeholder items were positioned with one item per quadrant plus the fifth item, which was placed in a randomly selected quadrant. Two memory items never appeared in one quadrant together, and all items were placed at least 4° apart. Subjects viewed a memory array (250 ms), remembered the items across a delay (1,000 ms), were probed on one item, and reported whether the probed item was the same as or different from the remembered item (unspedded). Subjects completed 14 blocks of 120 trials each, for a total of 1,680 trials per session (420 per set size). Two subjects completed only 1,348 and 1,440 trials each. EEG acquisition duration was between 73 and 132 min with an average of 105 min.

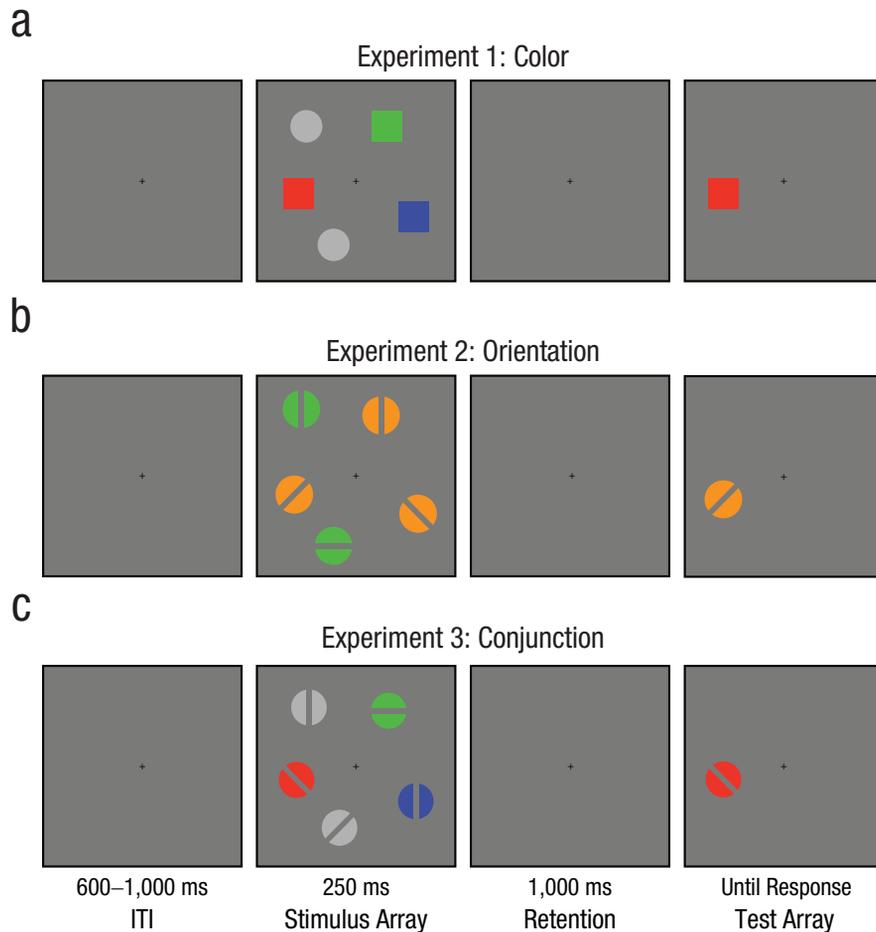


Fig. 1. Task schematics for an example Set Size 3 trial in the whole-field change-detection task used in all three experiments. In Experiment 1 (a), subjects remembered the colored squares while ignoring the gray placeholders. In Experiment 2 (b), at the start of each block, a color cue informed subjects to attend to and remember either the orange or the green orientations while ignoring the uncued color. In Experiment 3 (c), subjects remembered both the color and the orientation of each item. During change trials, one of the features (randomly selected) in the test item would change. ITI = intertrial interval.

Experiment 1: color. In Experiment 1, the memory items were colored squares (width = 2° ; see Fig. 1a). The colors were randomly sampled without replacement from a set of seven colors (RGB values: red = 255, 0, 0; green = 0, 255, 0; blue = 0, 0, 255; yellow = 255, 255, 0; purple = 255, 0, 255; teal = 0, 255, 255; orange = 255, 128, 0). Circular gray placeholders (radius = 1.13°) of the same area as the memory items also appeared during the memory array so that each display contained five total objects. One potential concern is that the spatial frequency of displays covaried with load because the colored squares had a higher spatial frequency than the circular placeholders. Although this raises a possible alternative explanation of load decoding in the color condition, there was no similar concern with the displays used in Experiments 2 and 3.

Experiment 2: orientation. In Experiment 2, the memory items were circles (radius = 1.3°) with oriented bars cut out of the middle (height = 2.6° , width = 0.5°) so that they

were the same area as the items in Experiment 1 (Fig. 1b). The possible orientations were 0° , 90° , 180° , and 270° , and they were sampled without replacement for each trial. The placeholder items were the same shape. In each block, either orange or green was indicated as the target color for that block. Subjects were instructed to remember the orientation of the stimuli presented in the target color and to ignore the stimuli presented in the other color. Both the orange and green were luminance matched to the average luminance of the color set in Experiment 1 (RGB values: orange = 255, 155, 55; green = 75, 208, 75). Thus, luminance was perfectly balanced across set-size conditions. For example, a trial that contained one orange and four green items would be a Set Size 1 trial in a “target orange” block but a Set Size 4 trial in a “target green” block.

Experiment 3: conjunction. In Experiment 3, each memory item included both an orientation and a color and was the same shape and size as items in Experiment

2. Both features were independently sampled without replacement from the same color and orientation values used in Experiments 1 and 2 (Fig. 1c). The placeholders were the luminance-matched gray from Experiment 1. In change trials, only one attribute (color or orientation) changed, with color and orientation changes occurring equally often.

EEG acquisition

We recorded EEG activity from 30 active Ag/AgCl electrodes mounted in an elastic cap (Brain Products actiCHamp, Munich, Germany). We recorded from international 10-20 sites Fp1, Fp2, F7, F3, Fz, F4, F8, FT9, FC5, FC1, FC2, FC6, FT10, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, Oz, and O2. Two additional electrodes were affixed with stickers to the left and right mastoids, and a ground electrode was placed in the elastic cap at position Fpz. All sites were recorded with a right-mastoid reference and were re-referenced off-line to the algebraic average of the left and right mastoids. We recorded electrooculogram (EOG) data using passive electrodes, with a ground electrode placed on the left cheek. Horizontal EOG data were recorded from a bipolar pair of electrodes placed ~1 cm from the external canthus of each eye. Vertical EOG data were recorded from a bipolar pair of electrodes placed above and below the right eye. Data were filtered on-line (low cut-off = 0.01 Hz, high cut-off = 80 Hz, slope from low to high cut-off = 12 dB/octave) and were digitized at 500 Hz using BrainVision Recorder (Brain Products, Munich, Germany) running on a PC. Impedance values were brought below 10 k Ω at the beginning of the session.

Eye tracking

We monitored gaze position using a desk-mounted EyeLink 1000 Plus infrared eye-tracking camera (SR Research, Ontario, Canada). Gaze position was sampled at 1000 Hz. According to the manufacturer, this system provides spatial resolution of .01° of visual angle and average accuracy of 0.25 to 0.50° of visual angle. We calibrated the eye tracker every one to two blocks of the task and between trials during the blocks if necessary. We drift-corrected the eye-tracking data for each trial by subtracting the mean gaze position measured during a 200-ms window immediately before the onset of the memory array.

Artifact rejection

We segmented the EEG data into epochs time-locked to the onset of the memory array (200 ms before until 1,000 ms after stimulus onset). We baseline-corrected

the EEG data by subtracting mean voltage during the 200-ms window immediately prior to stimulus onset. Eye movements, blinks, blocking, drift, and muscle artifacts were first detected by applying automatic criteria. After automatic detection, we visually inspected the segmented EEG data for artifacts (amplifier saturation, excessive muscle noise, and skin potentials) and the eye-tracking data for ocular artifacts (blinks, eye movements, and deviations in eye position from fixation), and we discarded any epochs contaminated by artifacts. In all three experiments, all subjects included in the final sample had at least 200 trials of each set-size condition (800 trials total).

Eye movements

For eye-tracking data, we rejected trials that contained eye movements beyond a certain threshold (threshold = 1° of visual angle). For some subjects, eye-tracking data were not available (Experiment 1, $n = 2$; Experiment 2, $n = 3$). In these cases, EOG data were used. We rejected trials that contained horizontal or vertical EOG values beyond a threshold of 50 μ V.

Blinks

In addition to the threshold detection, blinks were detected by flagging trials with flatline data (no position data were recorded when the eye was closed). Additionally, we visually inspected the eye-tracking data for trial segments with missing data points.

Drift, muscle artifacts, and blocking

We checked for drift (e.g., skin potentials) with the `pop_rejtrend` function in *ERPLAB*. We excluded trials in which a line fitted to the EEG data had a slope greater than a certain threshold (slope = 10, minimal $r^2 = .3$). We checked for muscle artifacts with the `pop_artm-wppth` function in *ERPLAB* (Lopez-Calderon & Luck, 2014). We excluded trials with peak-to-peak activity greater than 100 μ V within a 200-ms window with 100-ms steps. We also excluded trials with any value beyond a threshold of 80 μ V.

mvLoad procedure

Binned trial classification (within subjects and within experiments). The mvLoad analysis is within-subjects classification of WM load on baselined EEG. Although our approach allowed robust above-chance performance with single trials, we used randomly chosen groups of 20 trials within each set size to increase signal-to-noise ratio. We divided each trial into 50-ms windows with 25-ms steps and calculated the average voltage for

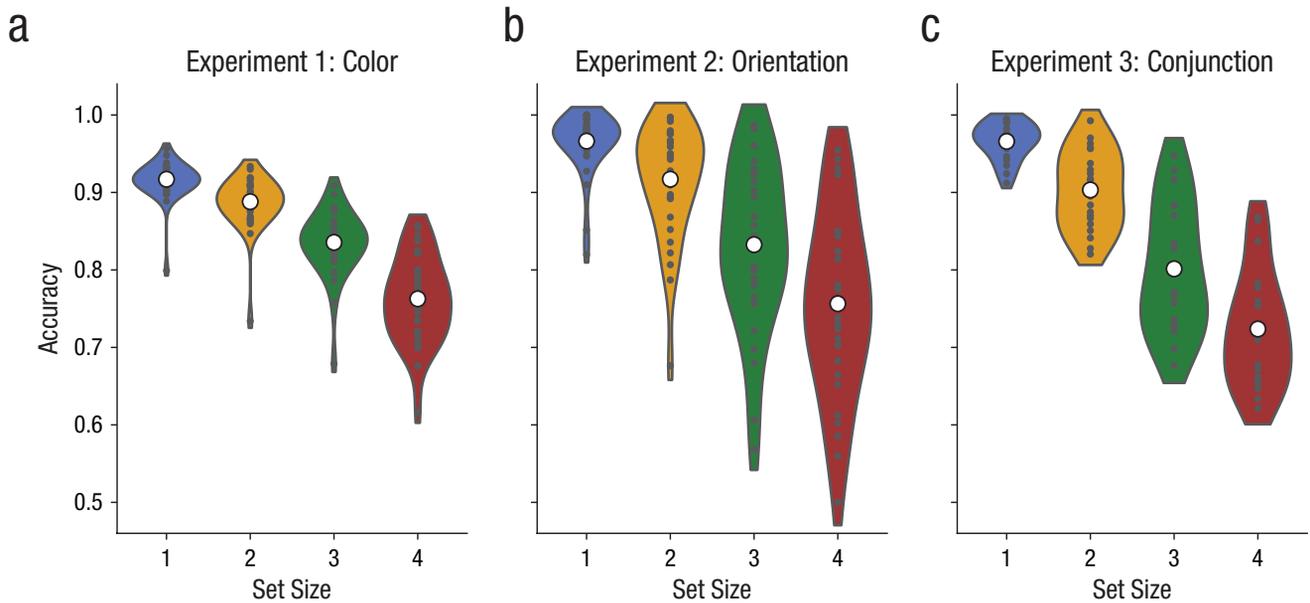


Fig. 2. Change-detection accuracy for each set size in each of the three experiments. Black dots indicate individual data, white dots indicate means, and shaded regions indicate the density of the data. See Figure S2 in the Supplemental Material for K estimates.

each electrode in the window. Classification was performed using an ordinal logistic regression model (Pedregosa-Izquierdo, 2015). The classifier was trained to discriminate between Load Conditions 1, 2, 3, and 4, giving a chance-level classification of 25%. Classification was tested on a held-out set of data using the StratifiedShuffleSplit function from *Scikit-Learn* (Pedregosa et al., 2011). This cross-validation procedure splits the data in 80% training and 20% testing sets while preserving the percentage of samples for each load condition. This split was repeated 1,000 times, and results for each subject and time point were averaged across these repetitions. Training data were standardized at each time point using the StandardScaler Scikit-Learn function, and test data were standardized using the mean and standard deviation of the training set.

Binned trial classification (within subjects and across experiments). Cross-training classification was used to test for generalization between the color (Experiment 1) and orientation (Experiment 2) conditions and between the single-feature (Experiments 1 and 2) and conjunction (Experiment 3) conditions. These analyses followed the same procedures as the within-experiment classification except that the testing was done on EEG data from a different experiment. For the single-feature generalizability analysis, the classifier was trained on data from Experiment 1 and tested on data from Experiment 2 and vice versa. For the single-feature-to-conjunction generalizability analysis, the classifiers were trained on a mixture of data from Experiments 1 and 2 and tested on data

from Experiment 3. All of these analyses were done within subjects, using the subset of subjects who completed all of the experiments involved in the analysis (Experiments 1 and 2: $n = 24$; Experiments 1, 2, and 3: $n = 20$).

Significance testing

In all classification analyses, we tested whether classification accuracy was significantly above chance at each time point using a paired-samples, one-tailed t test. Classification accuracy was compared with empirical chance accuracy, defined by testing the trained model on randomly shuffled trial labels (for more details, see Kappenman et al., 2021). Because we tested for significance at each time point (48 time bins between 0 ms and 1,250 ms), we used the Benjamini-Hochberg procedure to control the false-discovery rate (FDR) at .05.

Results

Behavioral

Across all experiments and conditions, subjects performed the change-detection task with above-chance accuracy (see Fig. 2; range of condition accuracies = .72–.97). In each experiment, a one-way analysis of variance (ANOVA) revealed a significant main effect of set size, indicating that accuracy declined as set size increased—Experiment 1: $F(3, 128) = 90.19, p < .001$; Experiment 2: $F(3, 120) = 32.93, p < .001$; Experiment 3: $F(3, 76) = 60.27, p < .001$. To examine whether

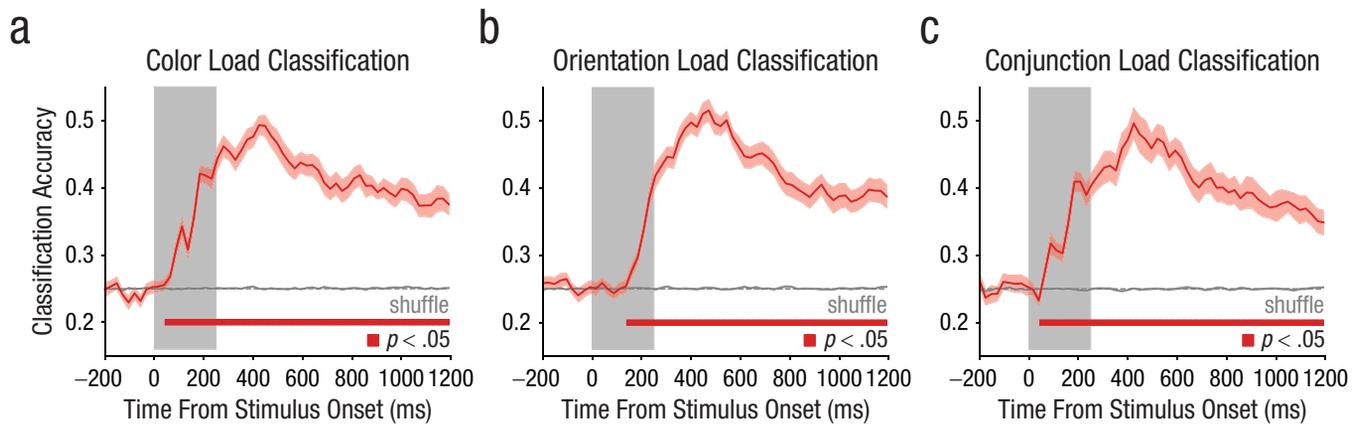


Fig. 3. Classification accuracy over time for (a) Experiment 1, (b) Experiment 2, and (c) Experiment 3. Classification accuracy is indicated with a red line. The shaded area around the red line indicates the standard error of the mean. Red squares indicate time points in which classification accuracy was significantly above chance (corrected $p < .05$, false-discovery rate = .05 with Benjamini-Hochberg procedure). The gray line indicates chance classification accuracy. The vertical gray rectangle indicates the time period during which the memory array was displayed. The shuffle condition reveals empirical chance accuracy, obtained by training the model on nonpermuted data then testing on data with permuted trial labels.

behavioral performance varied across the three experiments, we carried out a within-subjects analysis using only the 20 observers who completed all three experiments. We combined data from Experiments 1 and 2 (single-feature items) to compare with Experiment 3 (conjunction items). In a two-way repeated measures ANOVA on accuracy, there was no significant main effect of feature, $F(1, 19) = 4.09, p = .057$, but there was a significant main effect of set size, $F(3, 57) = 213.30, p < .001$, and a significant interaction of feature and set size, $F(3, 57) = 17.69, p < .001$. To characterize the significant interaction, we conducted four paired-samples t tests between the single-feature and conjunction conditions at each set size (corrected p , FDR = .05, with Benjamini-Hochberg procedure). Set Size 1 single-feature accuracy ($M = .94, SD = .03$) was significantly lower than conjunction ($M = .97, SD = .02$), $t(19) = -5.23, p < .001, d = 0.955$; Set Size 2 single-feature accuracy ($M = .90, SD = .04$) was not significantly different from conjunction ($M = .90, SD = .05$), $t(19) = 0.001, p = .993, d = 0.050$; Set Size 3 single-feature accuracy ($M = .84, SD = .06$) was significantly higher than conjunction ($M = .80, SD = .08$), $t(19) = 3.18, p = .007, d = 0.514$; and Set Size 4 single-feature accuracy ($M = .76, SD = .08$) was significantly higher than conjunction ($M = .72, SD = .07$), $t(19) = 3.81, p = .002, d = 0.651$. Despite revealing reliably worse performance in the conjunction experiment, this still provides evidence for object-based benefits for storage in visual WM (Olson & Jiang, 2002). That is, a larger number of feature values were stored in the conjunction condition than in the single-feature condition.

Precise classification of load while controlling for stimulus energy

The first key result was that the mvLoad analysis precisely classified WM load, despite the use of stimulus displays that controlled for stimulus energy across all load conditions. For each experiment (Experiment 1: $n = 33$; Experiment 2: $n = 31$; Experiment 3: $n = 20$), we used an ordinal logistic regression classifier on raw EEG amplitudes (see Fig. S3 in the Supplemental Material available online for event-related potentials) from binned trials within subjects (20 trials per bin) at each time bin (50-ms window). We could classify WM load (Set Size 1 vs. Set Size 2 vs. Set Size 3 vs. Set Size 4) during the stimulus presentation and throughout the delay period (Figs. 3a–3c; red squares indicate corrected $p < .05$, FDR-controlled at .05 with Benjamini-Hochberg procedure with 48 time bins tested). Above-chance classification was observed starting in early time bins in each experiment (Experiment 1: 64-ms to 88-ms time bin; Experiment 2: 160-ms to 208-ms time bin; Experiment 3: 64-ms to 88-ms time bin). Classification was sustained throughout the entire delay period for all three experiments. Mean classification accuracy (with chance at .25) during the delay period for Experiment 1 was .42 ($SD = .03$), for Experiment 2 was .43 ($SD = .04$), and for Experiment 3 was .41 ($SD = .04$). We also confirmed that the classifier was sensitive to single-item increments in the number of stored items. Figure 4 shows classification accuracy for Set Size 1 versus Set Size 2, Set Size 2 versus Set Size 3, and Set Size 3 versus Set Size 4. For Set Size 1 versus Set Size 2 and Set Size 2 versus Set

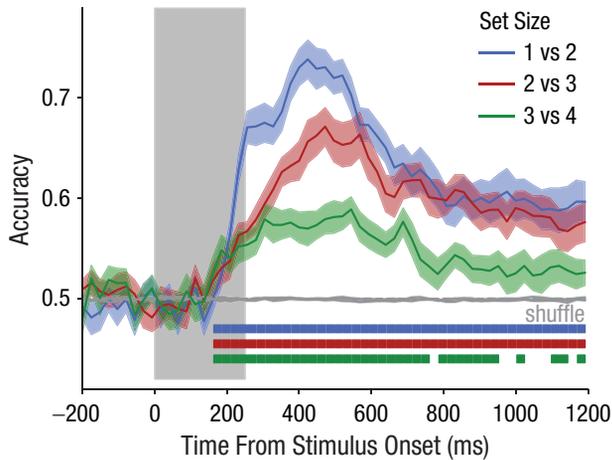


Fig. 4. Classification accuracy of single-feature load (Experiments 1 and 2 mixed together) for Set Size 1 versus Set Size 2, Set Size 2 versus Set Size 3, and Set Size 3 versus Set Size 4. Colored lines indicate classification accuracy. The shaded areas around the colored lines represent standard errors of the mean. Color-matched squares indicate time points in which classification accuracy was significantly above chance (corrected $p < .05$, false-discovery rate = .05 with Benjamini-Hochberg procedure). The gray line indicates chance classification accuracy. The vertical gray rectangle indicates the time period during which the memory array was displayed. The shuffle condition reveals empirical chance accuracy, obtained by training the model on nonpermuted data then testing on data with permuted trial labels.

Size 3, accuracy was sustained above chance throughout the entire delay period (corrected $p < .05$, FDR = .05 with the Benjamini-Hochberg procedure). Using behavioral (Set Size 4) and EEG data from each unique subject across the three experiments ($N = 40$), we replicated the finding from Adam et al. (2020) and Feldmann-Wüstefeld (2021) that classification accuracy was positively correlated with individual differences in WM capacity ($r^2 = .24$, $p = .001$; Fig. 5). Further analysis showed that this relationship was consistent across nearly all time points in the delay period (see Fig. S4 in the Supplemental Material). This correlation may be caused by the greater reliability with which higher-capacity individuals achieve the storage of all relevant items (e.g., Adam et al., 2015), which would in turn yield more discriminable patterns of activity for each set size. This finding reinforced the earlier evidence that the mvLoad analysis taps into a neural operation that is relevant for understanding capacity limits in visual WM.

A load signature that generalizes across distinct feature values

The second key analysis examined whether the load signatures revealed by mvLoad generalized across distinct feature values (i.e., color and orientation). Using data from subjects who had participated in both Experiments 1 and 2 ($n = 24$), we trained the classifier using

the color trials from Experiment 1 and tested it on the orientation trials from Experiment 2. We also trained on orientation trials and tested on color trials. In both directions of training and testing, robust classification was sustained throughout the entire delay period (Fig. 6). Mean classification accuracy during the delay period for color to orientation was .33 ($SD = .03$) and for orientation to color was .34 ($SD = .03$). Thus, the same multivariate pattern classified load precisely for memoranda with distinct relevant features, revealing a load-sensitive signal that is separable from the specific content stored in WM. These decoding accuracies are lower than those we saw with within-experiment analyses (see Figs. S5a and S5b in the Supplemental Material). Although this could reflect nongeneralizable aspects of the load signal, it could also reflect methodological noise across sessions, such as small differences in electrode placement or impedance. Thus, even if precisely the same load pattern were present in each EEG session, some drop in decoding accuracy would be expected for across-session relative to within-session training.

A signature of load that is independent of total amount of information stored

The third key analysis examined whether the load-sensitive activity revealed by the mvLoad analysis was independent of the total amount of feature information maintained about each item stored in WM. To this end, we trained the classifier on the combined data from Experiments 1 and 2, in which each item contained one relevant feature to be stored (i.e., either color or orientation), and we tested this model using data from

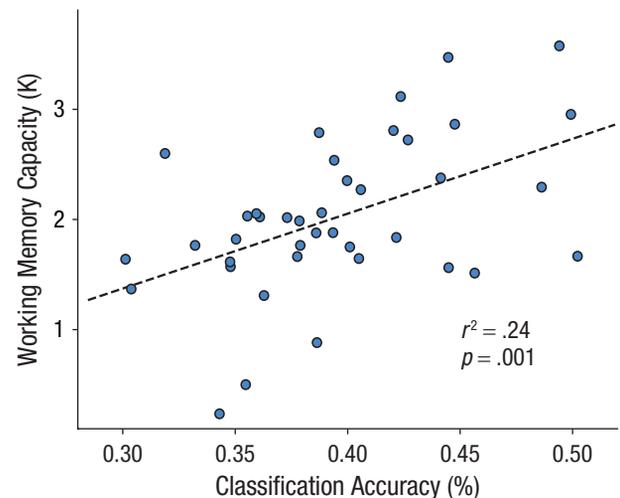


Fig. 5. Scatterplot (with best-fitting regression line) showing the relation between working memory capacity and classification accuracy. Classification accuracy is the average delay-period accuracy from all unique subjects across all three experiments; all data were used from each unique subject. Working memory capacity was measured using Set Size 4 trials.

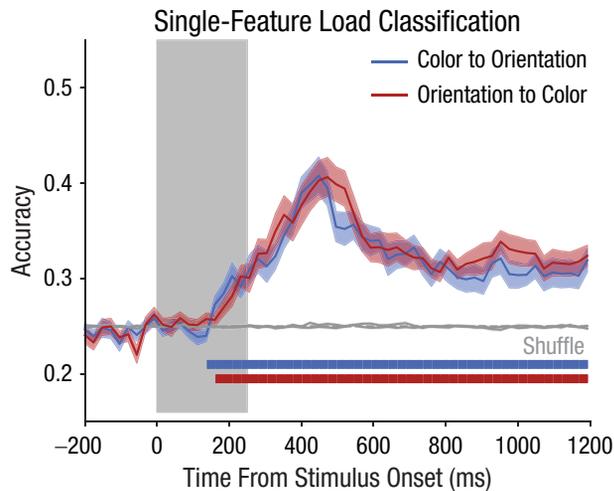


Fig. 6. Accuracy for single-feature load classification. The blue line represents the classifier's accuracy when trained on data from Experiment 1 (color) and tested on Experiment 2 (orientation). The red line represents the classifier's accuracy when trained on Experiment 2 and tested on Experiment 1. Color-matched squares indicate time points during which classification accuracy was significantly above chance (corrected $p < .05$, false-discovery rate = .05 with Benjamini-Hochberg procedure). The gray line indicates chance classification accuracy. The gray vertical rectangle indicates the time period during which the memory array was displayed. The shuffle condition reveals empirical chance accuracy, obtained by training the model on nonpermuted data then testing on data with permuted trial labels.

Experiment 3 in which the number of relevant features per item was doubled (i.e., both color and orientation; Fig. 7). This analysis included a group of 20 subjects who had participated in all three experiments. Classification accuracy was robustly above chance throughout the entire delay period with a mean delay-period accuracy of .36 ($SD = .03$). Again, the across-experiment decoding accuracy was lower than in the within-experiment analyses (see Fig. S5c in the Supplemental Material). Nevertheless, the same signature of load identified with single-feature stimuli was observed with conjunction stimuli that contained twice as many relevant features per item, in line with a load-sensitive cognitive operation that is separable from the maintenance of specific features.

Although robust cross-training between the single-feature and conjunction conditions suggests that they evoked a common load signature, further analyses provided more incisive evidence for the content-independent character of this load-sensitive neural activity. First, recall from Experiments 1 and 2 that the mvLoad analysis robustly detected the difference between one and two single-feature items (Fig. 3) and between two and three single-feature items, showing that the analysis is sensitive to the addition of a single item with one relevant feature. Thus, if load decoding with single-feature items was based on the number of color or orientation values stored, then one conjunction item should be

classified as the same load as two single-feature items. Alternatively, if load decoding was based on the number of feature-independent pointers stored, then one conjunction item should be classified as the same load as one single-feature item. To test this prediction, we trained the mvLoad classifier with single-feature stimuli and examined performance across three key conditions: (a) Set Size 1 single feature, (b) Set Size 2 single feature, and (c) Set Size 1 conjunction. The divergent predictions of the feature-load and pointer explanations are illustrated in Figures 8a and 8b, along with the observed data in 8c.

Visual inspection reveals that our findings fell directly in line with the pointer hypothesis in that a one-conjunction item was equivalent to one single-feature item. We tested the reliability of this pattern with two planned comparisons. First, we found a reliable difference between the predicted load for a single conjunction item and two single-feature items. A Bayesian paired-samples t test revealed strong evidence for a difference between these conditions, $t(19) = -9.01$, $p < .001$, $d = 3.09$, Bayes factor favoring the alternative over the null hypothesis ($BF_{10} > 100$), showing that a single conjunction item had a higher probability of being classified as Load 1 ($M = .65$, $SD = .12$) than Set Size 2 single-feature items ($M = .35$, $SD = .07$). Second, we examined the prediction that one conjunction item should have the same load as one single-feature item ($M = .65$, $SD = .06$), using a Bayesian paired-samples t test. This revealed substantial evidence for the null

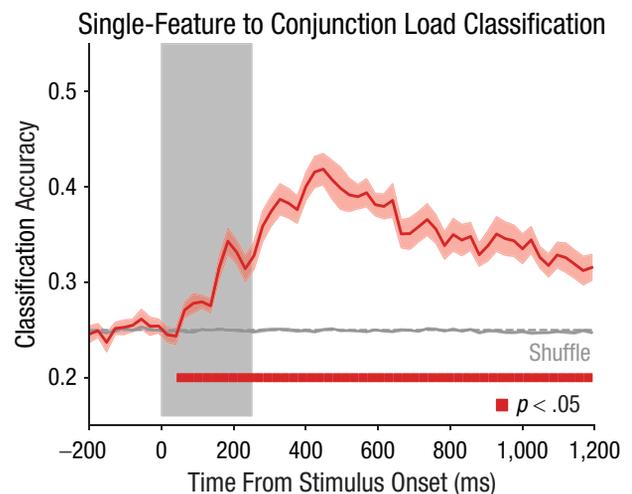


Fig. 7. Accuracy for classifier trained on data from Experiments 1 and 2 (single-feature items, color or orientation) and tested on data from Experiment 3 (conjunction items, color and orientation). Red squares indicate time points during which classification accuracy was significantly above chance (corrected $p < .05$, false-discovery rate = .05 with Benjamini-Hochberg procedure). The gray line indicates chance classification accuracy. The gray rectangle indicates the time period during which the memory array was displayed. The shuffle condition reveals empirical chance accuracy, obtained by training the model on nonpermuted data then testing on data with permuted trial labels.

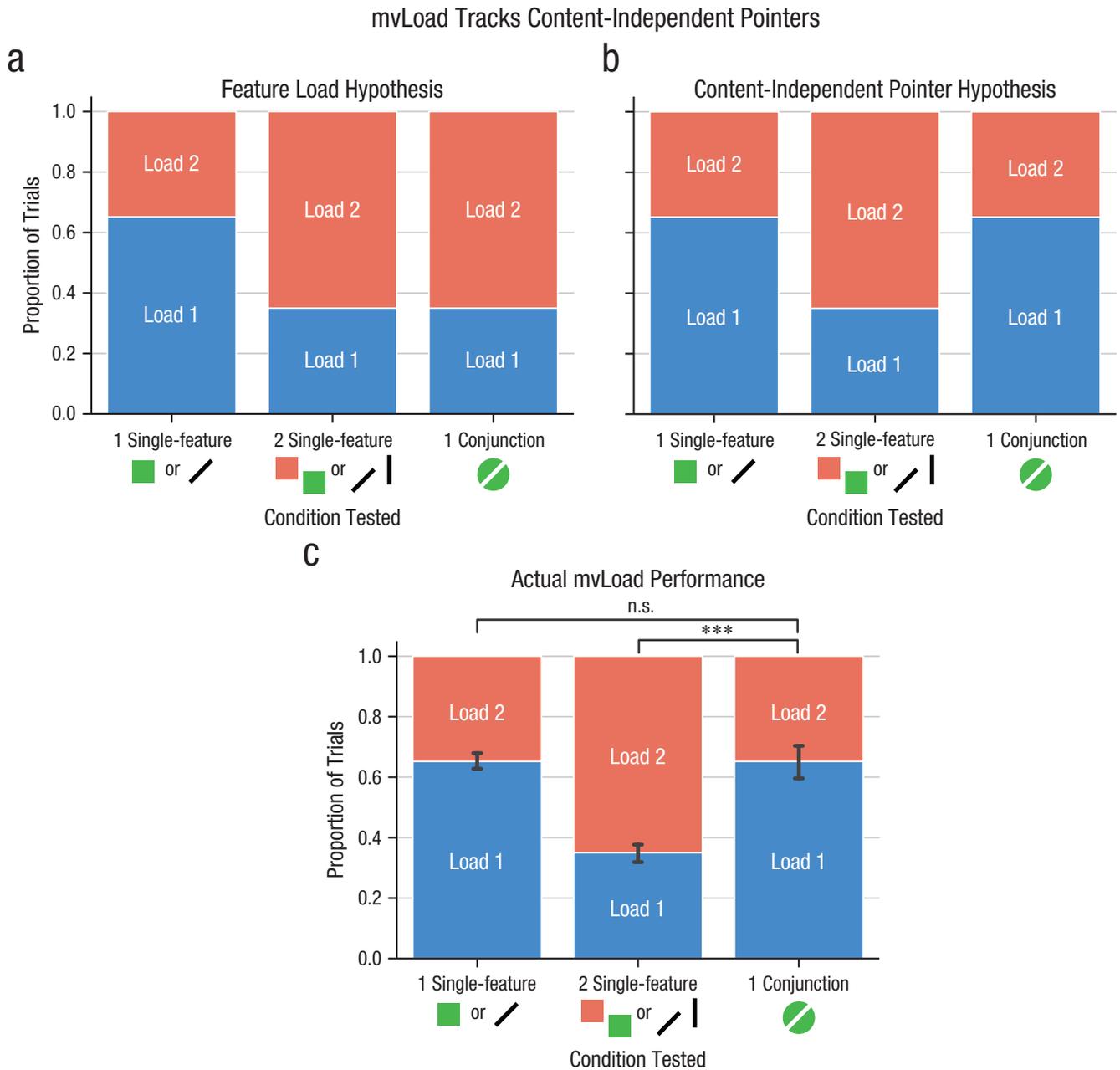


Fig. 8. Multivariate load detection (mvLoad) classifier output (trained on single-feature items) compared with the predicted output of the feature-load and pointer hypotheses. Within each graph, each bar represents 100% of trials classified as Load 1 (blue) or Load 2 (red) in each condition. The feature-load hypothesis (a) predicted that Load 1 conjunction items will be classified the same way as Load 2 single-feature items. The pointer hypothesis (b) predicts that Load 1 conjunction items will be classified the same way as Load 1 single-feature items. The actual mvLoad classifier performance is shown in (c). Asterisks indicate significant differences ($p < .001$).

hypothesis, suggesting that both had the same probability of being classified as Load 1, $t(19) = 0.00$, $p = .999$, $d = 0.00$, $BF_{10} = 0.232$). An analogous analysis of Set Size 2 and 4 trials revealed precisely the same empirical pattern, showing that two conjunction items ($M = .68$ classified as Load 2, $SD = .15$) were predicted as a lower load than four single-feature items ($M = .34$, $SD = .08$), $t(19) = -8.06$, $p < .001$, $d = 2.76$, $BF_{10} > 100$, and that two

conjunction items were predicted as the same load as two single-feature items ($M = .66$, $SD = .08$), $t(19) = -0.652$, $p = .522$, $d = 0.185$, $BF_{10} = 0.281$ (see Fig. S6 in the Supplemental Material). Thus, our findings strongly suggest that there is a common load signature for single-feature and conjunction stimuli that is determined by the number of individuated items stored, rather than by the number of feature values stored.

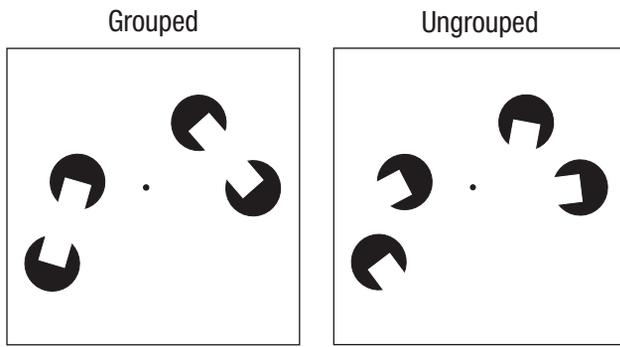


Fig. 9. Examples of Set Size 4 grouped and ungrouped memory arrays. In the grouped condition, collinearity between the notches yields the percept of a single oriented rectangle for each grouped pair.

Ruling out the size of the attended region as the driver of load-sensitive neural activity

Although the results of the mvLoad analysis pointed to load-sensitive neural activity that is separable from the quantity and type of content stored about each item, we noted that the spatial extent of the attended region in the display was confounded with the number of stored items. Thus, we examined whether the classifier was indexing the area of the attended regions on the screen, rather than the number of individuated items per se. To this end, we reanalyzed data from an EEG study of perceptual grouping by Diaz et al. (2021) in which subjects stored the orientation of two or four notched discs in visual WM. In the grouped condition, the discs were arranged so that collinearity between the notches in pairs of discs elicited the percept of a single illusory rectangle (Fig. 9). Thus, in the Set Size 4 grouped condition, perceptual grouping encouraged the perception of two individuated orientation values, whereas in the Set Size 4 ungrouped condition observers perceived four individuated orientation values. Critically, the number of relevant elements and their spatial extent were matched between the grouped and ungrouped displays. Diaz et al. (2021) reinforced this point by showing that the power of alpha oscillations in occipitoparietal electrodes, a neural signal that has been shown to track the number of attended locations (Fukuda et al., 2015), tracked the number of elements on the screen but was unaffected by the grouping manipulation. Thus, the key question for the present study was whether the mvLoad classifier would register the difference between the grouped and ungrouped displays. If load classification is based on the spatial extent of the attended locations, then it should return the same load value for the grouped and ungrouped

conditions, in line with the posterior alpha power signal examined by Diaz et al. (2021). By contrast, if load classification is based on the number of individuated items stored, then a lower load should be detected in the grouped relative to the ungrouped condition.

Figure 10 illustrates the output over time of a classifier that was trained exclusively on ungrouped displays (Set Size 2 or 4) and then tested on both the ungrouped and grouped displays. The output here is from the classifier's `decision_function` method, which returns the confidence score of the sample. This score is proportional to the signed distance of that sample to the hyperplane. In Figure 10, stronger evidence for Set Size 4 is plotted in the positive direction, whereas stronger evidence for Set Size 2 is plotted in the negative direction. When trained and tested on the ungrouped trials, the classifier exhibited sustained above-chance performance throughout the delay period (i.e., sustained positive values for Set Size 4 and sustained negative values for Set Size 2). However, when the same classifier was tested with Set Size 4 grouped trials, classification evolved over time. Set Size 4 grouped trials were initially classified the same as Set Size 4 ungrouped.

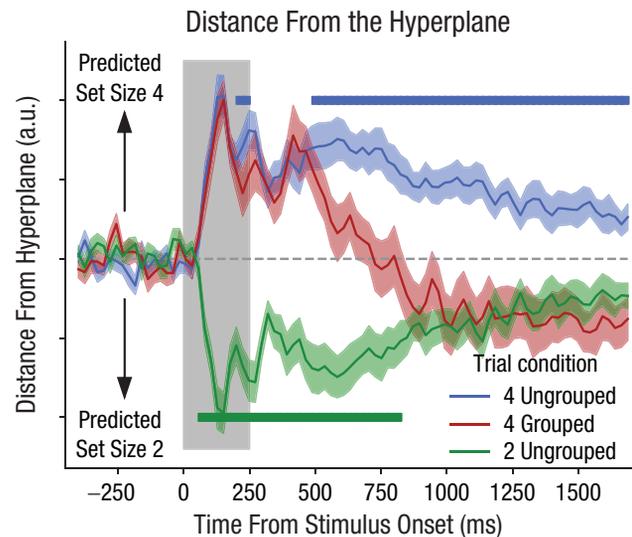


Fig. 10. Distance from the classification hyperplane for Set Size 2 ungrouped, Set Size 4 grouped, and Set Size 4 ungrouped trials across time. Classifiers were trained on Set Size 2 and Set Size 4 ungrouped trials and tested on all three conditions. The hyperplane is indicated with the dashed gray line. Trials above the hyperplane are classified as Set Size 4, whereas trials below it are classified as Set Size 2. The colored lines show distance from the hyperplane for each trial condition at each time point. Blue squares indicate time points during which four ungrouped was significantly greater than four grouped (corrected $p < .05$, false-discovery rate = .05 with Benjamini-Hochberg procedure). Green squares indicate time points during which two ungrouped was significantly less than four grouped (corrected $p < .05$, false-discovery rate = .05 with Benjamini-Hochberg procedure). a.u. = arbitrary units.

However, by the 512-ms to 536-ms time bin, Set Size 4 grouped diverged from Set Size 4 ungrouped and was reliably closer to the hyperplane. Set Size 4 grouped was also reliably different from Set Size 2 ungrouped at the start of the trial. However, by the 848-ms to 872-ms time bin, Set Size 4 grouped had crossed the hyperplane and was no longer reliably different from Set Size 2 ungrouped. Thus, although perceptual grouping did not affect the spatial extent of the attended region (Diaz et al., 2021), the mvLoad classifier indexed a lower number of stored items in the grouped condition, showing that the classifier indexes the number of individuated items stored in memory, not the spatial extent of covert attention.

Discussion

Given that WM serves as a cornerstone for intelligent behaviors, there is strong motivation to build a taxonomy of the neural operations that support on-line memory storage. The dominant strain of this work has focused on stimulus-specific neural activity that represents the stored content (e.g., D'Esposito & Postle, 2015; Funahashi et al., 1989; Fuster & Jervey, 1981; Goldman-Rakic, 1995; Rademaker et al., 2019; Serences et al., 2009), and great progress has been made in understanding the format and anatomical locus of this class of neural activity. By contrast, we highlight evidence for a qualitatively different neural operation that is integral to WM function but separable from the maintenance of stored content. Specifically, we refer to a spatiotemporal pointer operation that supports the segmentation of visual scenes into individuated representations that can be tracked through time and space (Kahneman et al., 1992; Pylyshyn, 2009). Using a multivariate analytic approach (Adam et al., 2020), we found that the scalp topography of EEG voltage precisely tracks the number of individuated representations stored in visual WM, while generalizing across variations in both the type and number of relevant features per item. Thus, although this neural operation is hypothesized to track the spatiotemporal coordinates of stored objects, it operates in a fashion that is insensitive to the contents of the tracked memory representations. Moreover, the fidelity of this load-sensitive neural activity is a predictor of individual differences in WM capacity, emphasizing its importance for understanding why WM capacity is limited.

The present findings provide a critical complement to past work that has sought to determine the computational role of load-sensitive neural activity. For instance, multiple studies have reported EEG and blood oxygen level dependent (BOLD) activity patterns that rise with each additional item stored and reach an

apparent plateau at set sizes that exceed behavioral estimates of capacity in visual WM (e.g., Todd & Marois, 2004; Vogel & Machizawa, 2004; Xu & Chun, 2006). But although this empirical pattern is consistent with a neural operation that tracks number per se, it can also be modeled using a biophysically plausible saturation model in which stimulus-specific neural activity follows an exponential function (Bays, 2018). There have also been reports of neural activity that rises with the number of items but is not affected by the complexity of the memoranda (Woodman & Vogel, 2008; Xu & Chun, 2006). This empirical pattern suggests a neural operation that indexes the number of individuated representations stored in WM rather than the total amount of visual information. That said, these conclusions are based on an intriguing null result: the absence of a difference in mean activity levels across distinct types of stimuli. By contrast, our findings provide positive evidence for a common neural index of the number of stored items when the type and number of visual features per item is varied: a multivariate signature of load that robustly generalizes across three distinct types of memoranda, demonstrating a content-independent aspect of storage-related neural activity. Moreover, our findings were supported by 84 separate EEG sessions across 40 unique observers that yielded above-chance decoding in every session for every observer tested. Thus, our findings provide compelling positive evidence for an item-based, content-independent aspect of storage in visual WM. These positive features notwithstanding, two limitations of our work include the use of a relatively limited set of visual stimuli, and a subject population that was dominated by people in and around our university community. It will be valuable for future work to examine how and whether our conclusions generalize to different stimuli and subject populations.

Our working hypothesis is that this load-sensitive neural activity reflects the deployment of spatiotemporal pointers or indexes that support *object individuation*—the segmentation of objects from the background and from other objects—and the continuous tracking of items through time and space (e.g., Kahneman et al., 1992; Pylyshyn, 2009; Xu & Chun, 2009). To study this cognitive process, Pylyshyn and Storm (1988) introduced *multiple-object tracking*, a task that requires the observer to keep track of varying numbers of targets that move randomly among a group of identical distractors. Their behavioral data indicated a relatively sharp capacity limit that they attributed to a limit on the number of pointers that could be concurrently deployed. Interestingly, Drew and Vogel (2008) used a lateralized version of the multiple-object-tracking task to show that contralateral delay

activity rises with the number of targets that are tracked, predicts individual differences in tracking ability, and reaches an apparent plateau after three targets are selected. Thus, it may be that both the contralateral delay activity and mvLoad classifiers are picking up on a content-independent indexing operation that is required during tracking and visual WM tasks (Balaban et al., 2019; Hakim et al., 2019; Tsubomi et al., 2013).

In combination with stimulus-selective neural activity that supports the maintenance of precise memories (e.g., D'Esposito and Postle, 2015), evidence for a content-independent pointer operation falls in line with various proposals for a separation between the precise maintenance of content and the number of representations maintained in WM. For example, if WM storage is limited by the deployment of content-independent pointers, this could explain why the maximum number of items an individual can store is uncorrelated with the precision of those representations (Awh et al., 2007) and why number exhibits a strong correlation with fluid intelligence whereas precision does not (Fukuda et al., 2010). Likewise, this separation may explain why different regions of visual cortex appear to track the number and complexity of the memoranda stored in WM (e.g., Xu & Chun, 2006). In addition, if storage in visual WM is contingent on the assignment of a pointer, this could explain why many studies have documented an object-based benefit in which a larger number of features can be maintained within multifeature objects compared with single-feature objects (e.g., Luck & Vogel, 1997; Olson & Jiang, 2002). Specifically, if each individuated object stored requires one of a limited number of pointers, then single-feature items would be the least efficient way to store the largest number of features.

In conclusion, multivariate analysis of the topography of EEG voltage reveals a load-sensitive neural operation that tracks the number of individuated items stored in WM while generalizing across variations in the type and number of visual features. This empirical pattern provides critical new evidence for a distinction between the maintenance of visual features and the discrete indexing of the items that contain those features. These findings help to clarify the taxonomy of neural operations that support storage in this on-line mental workspace.

Transparency

Action Editor: Sachiko Kinoshita

Editor: Patricia J. Bauer

Author Contributions

W. Thyer and E. Awh conceived and designed the experiments. W. Thyer and I. N. Velázquez Sánchez collected the

data. W. Thyer and G. K. Diaz analyzed the data. W. Thyer and E. Awh drafted the manuscript. All authors revised the manuscript and approved the final version for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

Research was supported by National Institute of Mental Health Grant No. RO1MH087214 and Office of Naval Research Grant No. N00014-12-1-0972.

Open Practices

All data, analysis scripts, and task scripts have been made publicly available via OSF and can be accessed at <https://osf.io/uuhbx5/>. The design and analysis plans for the experiments were not preregistered. This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Kirsten C. S. Adam  <https://orcid.org/0000-0002-4950-327X>
Itzel N. Velázquez Sánchez  <https://orcid.org/0000-0001-6357-6566>

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/09567976221090923>

References

- Adam, K. C. S., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The contribution of attentional lapses to individual differences in visual working memory capacity. *Journal of Cognitive Neuroscience*, 27(8), 1601–1616.
- Adam, K. C. S., Vogel, E. K., & Awh, E. (2020). Multivariate analysis reveals a generalizable human electrophysiological signature of working memory load. *Psychophysiology*, 57(12), Article e13691.
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, 18(7), 622–628.
- Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, 5(3), 119–126.
- Balaban, H., Drew, T., & Luria, R. (2019). Neural evidence for an object-based pointer system underlying working memory. *Cortex*, 119, 362–372.
- Bays, P. M. (2018). Reassessing the evidence for capacity limits in neural signals related to working memory. *Cerebral Cortex*, 28(4), 1432–1438.
- Bouchacourt, F., & Buschman, T. J. (2019). A flexible model of working memory. *Neuron*, 103(1), 147–160.

- Carlisle, N. B., & Woodman, G. F. (2011). Automatic and strategic effects in the guidance of attention by working memory representations. *Acta psychologica, 137*(2), 217–225.
- Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia, 49*(6), 1407–1409.
- D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology, 66*, 115–142.
- Diaz, G. K., Vogel, E. K., & Awh, E. (2021). Perceptual grouping reveals distinct roles for sustained slow wave activity and alpha oscillations in working memory. *Journal of Cognitive Neuroscience, 33*(7), 1354–1364.
- Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *The Journal of Neuroscience, 28*(16), 4183–4191.
- Feldmann-Wüstefeld, T. (2021). Neural measures of working memory in a bilateral change detection task. *Psychophysiology, 58*, Article e13683.
- Fukuda, K., Mance, I., & Vogel, E. K. (2015). α power modulation and event-related slow wave provide dissociable correlates of visual working memory. *The Journal of Neuroscience, 35*(41), 14009–14016. <https://doi.org/10.1523/JNEUROSCI.5003-14.2015>
- Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review, 17*(5), 673–679.
- Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology, 61*(2), 331–349.
- Fuster, J. M., & Jervey, J. P. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science, 212*(4497), 952–955.
- Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron, 14*(3), 477–485.
- Gunseli, E., Meeter, M., & Olivers, C. N. (2014). Is a search template an ordinary working memory? Comparing electrophysiological markers of working memory maintenance for visual search and recognition. *Neuropsychologia, 60*, 29–38.
- Hakim, N., Adam, K. C., Gunseli, E., Awh, E., & Vogel, E. K. (2019). Dissecting the neural focus of attention reveals distinct processes for spatial attention and object-based storage in visual working memory. *Psychological Science, 30*(4), 526–540.
- Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458*(7238), 632–635. <https://doi.org/10.1038/nature07832>
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*(2), 175–219.
- Kappenman, E. S., Farrens, J. L., Zhang, W., Stewart, A. X., & Luck, S. J. (2021). ERP CORE: An open resource for human event-related potential research. *NeuroImage, 225*, 117465.
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, Article 213. <https://doi.org/10.3389/fnhum.2014.00213>
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281.
- Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience & Biobehavioral Reviews, 62*, 100–108.
- McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience, 11*(1), 103–107.
- Oberauer, K. (2019). Working memory capacity limits memory for bindings. *Journal of Cognition, 2*(1), Article 40. <https://doi.org/10.5334/joc.86>
- Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Perception & Psychophysics, 64*(7), 1055–1067.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
- Pedregosa-Izquierdo, F. (2015). *Feature extraction and supervised learning on fMRI: From practice to theory* [Medical imaging] (NNT: 2015PA066015, tel-01100921v2). Université Pierre et Marie Curie–Paris VI.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Poo, M. M., Pignatelli, M., Ryan, T. J., Tonegawa, S., Bonhoeffer, T., Martin, K. C., Rudenko, A., Tsai, L.-H., Tsien, R. W., Fishell, G., Mullins, C., Gonçalves, J. T., Shtrahman, M., Johnston, S. T., Gage, F. H., Dan, Y., Long, J., Buzsáki, G., & Stevens, C. (2016). What is memory? The present state of the engram. *BMC Biology, 14*(1), 1–18.
- Pylshyn, Z. W. (2009). Perception, representation, and the world: The FINST that binds. In D. Dedrick & L. Trick (Eds.), *Computation, cognition, and Pylshyn* (pp. 3–48). MIT Press.
- Pylshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179–197.
- Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience, 22*(8), 1336–1344.
- Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science, 20*(2), 207–214.
- Swan, G., & Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception, & Psychophysics, 76*(7), 2136–2157.
- Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428*(6984), 751–754.

- Tsubomi, H., Fukuda, K., Watanabe, K., & Vogel, E. K. (2013). Neural limits to representing objects still within view. *The Journal of Neuroscience*, *33*(19), 8257–8263.
- 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.
- Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2015). Working memory delay activity predicts individual differences in cognitive abilities. *Journal of Cognitive Neuroscience*, *27*(5), 853–865.
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751.
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503.
- Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review*, *15*(1), 223–229.
- Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*(7080), 91–95.
- Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences*, *13*(4), 167–174.