Storage in visual working memory recruits a content-independent pointer system

William Thyer[1]

Kirsten C.S. Adam[2]

Gisella K. Diaz[1]

Itzel N. Velázquez Sánchez[1]

Edward K. Vogel[1]

Edward Awh[1]

University of Chicago[1]

University of California San Diego[2]

Running Head: POINTERS IN VISUAL WORKING MEMORY
Figures: 10
Tables: 0
Key words: visual working memory, spatiotemporal pointers, electroencephalogram
Contributions: W.T. and E.A. conceived and designed the experiments. W.T. and I.N.V. collected the data. W.T. and G.K.D. analysed the data. W.T. and E.A. drafted the manuscript. All authors read and assisted with revision of manuscript.
Funding: Research was supported by NIMH grant ROIMH087214 and Office of Naval Research grant N00014-12-1-0972.
Data availability: Datasets, analysis scripts, and task scripts will become available on Open Science Framework upon acceptance of the manuscript or upon reviewer request.
Conflicts of interest: none.
Correspondence to:
Edward Awh
University of Chicago
940 E. 57th St, Chicago, IL 60637
awh@uchicago.edu

Abstract

Past work has shown that storage in working memory (WM) 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 human subjects stored varying numbers of items in visual WM. 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 WM capacity. Critically, this signature of WM 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 online storage and attentive tracking.

Relevance statement

Because working memory plays a central role in intelligent behaviors, there is strong motivation to understand the classes of neural activity that support storage in this online memory system. This effort has been dominated by studies focused on neural activity that represents the properties of the stored items, but here we demonstrate that there is a distinct class of neural operation that indexes the number of stored items in a manner that is independent of the stored details. This provides evidence for a content-independent indexing or *pointer* operation that enables the continuous monitoring of items held in the focus of attention. Thus, this work suggests a critical distinction between the storage of information in working memory, and the indexing operations that track individuated items through time and space.

Introduction

A central goal of cognitive neuroscience has been to understand the neural underpinnings of working memory (WM), an online 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 working memory. 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., Fuster and Jervey, 1981; Funahashi et al., 1989; Goldman-Rakic, 1995; Serences et al, 2009; Harrison and Tong, 2009; Rademaker et al, 2019; D'Esposito and Postle, 2015). The motivation for these studies is clear, as they have the potential to elucidate the memory engrams (Poo et al., 2016) that allow us 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 working memory, rather than the content of those representations (e.g., Vogel and Machizawa, 2004; Todd and Marois, 2004; Xu and Chun, 2006; Adam et al, 2020). 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 working memory. This contralateral delay activity (CDA) 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 working memory (Luria et al, 2016). This kind of load-sensitive neural measure has provided insight into how observers control access to this limited online workspace (Vogel, McCollough & Machizawa, 2005; McNab and Klingberg, 2008), the role of working memory in complex tasks such as multiple object tracking (Drew and Vogel, 2008) and visual search (Carlisle and Woodman, 2011; Gunseli et al, 2014), and the relationship between working memory capacity and other cognitive abilities (Unsworth et al, 2014; Unsworth et al., 2015).

Although it is clear that load-sensitive neural signals have been potent tools for studying working memory, 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 longstanding 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" (FINSTs) 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, but that is separable from the maintenance of the specific features of the attended items.

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 towards *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., Xu and Chun, 2009; Swan and Wyble, 2014; Bouchacourt & Buschman, 2019; Balaban et al., 2019; Oberauer, 2019). For example, Swan and Wyble (2014) postulate 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 and Chun, 2005). Thus, there is clear motivation to postulate the existence of load-sensitive neural signals that index a content-independent aspect of working memory. 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 working memory (Adam et al., 2020). Although past work has found univariate signals that covary with the number of items stored in working memory, there are several reasons why multivariate load detection (mvLoad) provides a more powerful testbed 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 dataset 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 online 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, three studies using the mvLoad analytic approach confirm 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 working memory. We propose that this content-independent

signature of WM load indexes the deployment of spatiotemporal pointers (e.g., Pylyshyn, 2009; Khaneman et al., 1992) that enable the individuation, binding, and monitoring of attended objects.

Methods

Subjects

Experiments 1-3 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/hr). A total of 20 volunteers participated in all 3 studies, allowing us to implement cross-training analyses across experiments. 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.

Experiment 1

Our target sample was 30 subjects in Experiment 1. 42 volunteers participated in Experiment 1 (25 female; mean age = 23.8, SD = 4.5). 9 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 was unintentionally overwritten (n = 1), and too many trials were rejected due to eye movements (see Eye Movement Controls, 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 3 because we needed enough subjects to complete all 3 experiments and some could not return.

Experiment 2

Our target sample was 30 subjects in Experiment 1. 33 volunteers participated in Experiment 1 (18 female; mean age = 25.39, SD = 4.30). 2 subjects were excluded from the final sample for the following reasons: the subject did not complete enough blocks of the task (n = 2). The final sample size was 31 (18 female; mean age = 25.32 years, SD = 4.07). We overshot our target sample size by 1 because we needed enough subjects to complete all 3 experiments and some could not return.

Experiment 3

Our target sample was 20 subjects in Experiment 3. 20 volunteers participated in Experiment 3 (13 female; mean age = 25.45, 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" LCD monitor (refresh rate: 120 Hz, resolution 1080 x 1920 pixels) with their chin 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/m 2). Memory arrays included 1-4 to-be-remembered items. Ignored placeholder items also appeared in the

memory array such that each array had a total of 5 items. The placeholder items were a gray (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 1-4 colored items to be remembered, and the remainder of items were gray placeholder items to balance area and luminance across set size conditions (see Luminance-balanced displays for more detail). Memory and placeholder items were positioned with 1 item per quadrant, plus the 5th item which was placed in a randomly selected quadrant. 2 memory items never appeared in 1 quadrant together. All items were placed at least 4° apart. Participants viewed a memory array (250 ms), remembered the items across a delay (1000 ms), were probed on one item, and reported whether the probed item was the same as or different from the remembered item (unspeeded). Participants completed 14 blocks of 120 trials, for a total of 1680 trials per session (420 per set size). 2 subjects only completed 1,348 and 1,440 trials each. EEG acquisition duration was between 73 and 132 minutes, with an average of 105 minutes.

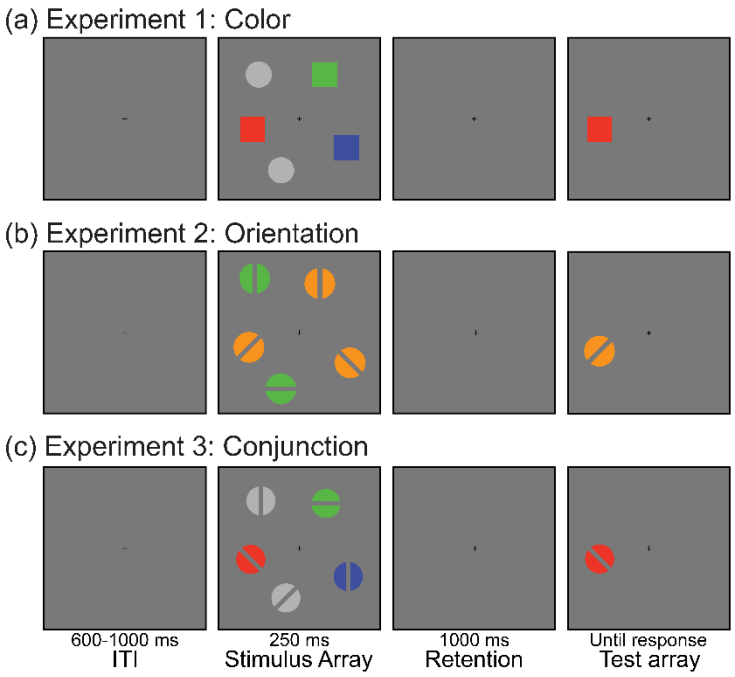

Figure 1 Task schematics for an example set size 3 trial in the whole-field change detection task used in all three experiments. (a) In Experiment 1. participants remembered the colored squares while ignoring the grey placeholders. (b) In Experiment 2. at the start of each blook, a color cue informed participants to attend and remember either the orange or the green orientations while ignoring the unced color. (c) In Experiment 3. participants 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.

Experiment 1: Color

In Experiment 1, the memory items were colored squares (width = 2°) (Figure 1a). The colors were randomly sampled without replacement from a set of 7 colors (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 such that each display contained 5 total objects.

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 = .5°) such that they were the same area as the items in Experiment 1 (Figure 1b). The possible orientations were 0, 90, 180, and 270 degrees and they were sampled without replacement for each trial. The placeholder items were the same shape. 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 (orange = 255, 155, 55; green = 75, 208, 75). Thus, luminance was perfectly balanced across set size conditions. For example, a trial that contains 1 orange and 4 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 (Figure 1c). The placeholders were the luminance-matched grey 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, 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 offline 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 was recorded from a bipolar pair of electrodes placed ~1 cm from the external canthus of each eye. Vertical EOG was recorded from a bipolar pair of electrodes placed above and below the right eye. Data were filtered online (low cut-off = .01 Hz, high cut-off = 80 Hz, slope from low- to high-cutoff = 12 dB/octave), and were digitized at 500 Hz using BrainVision Recorder (Brain Products, Munich, German) running on a PC. Impedance values were brought below 10 kΩ 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 .25-.50° of visual angle. We calibrated the eye tracker every 1-2 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 1000 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 discarded any epochs contaminated by artifacts. In all 3 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 which contained eye movements beyond a certain threshold (threshold = 1° of visual angle). For some subjects, eye tracking data was not available (Exp. 1, n = 2; Exp. 2, n = 3). In these cases, EOG data was used. We rejected trials which contained horizontal or vertical EOG values beyond a threshold of 50 μV.

Blinks

In addition to the threshold detection, blinks were detected by flagging trials with flatline data (no position data are recorded when the eye is 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 where a line fit to the EEG data has a slope greater than a certain threshold (slope = 10, minimal r^2 = .3). We checked for muscle artifacts with the pop_artmwppth function in ERPLAB (Lopez-Calderon, 2014). We excluded trials with peak to peak activity greater than 100 μV within a 200 ms window with 100 ms steps. We also excluded trials with any value beyond a threshold of 80 μV.

Classification Procedure

Binned trial classification (within-subject & within-experiment)

Classification was performed within-subject on baselined EEG data that had been band-pass filtered between .1 and 30 hz. Although our approach allowed robust above-chance performance with single trials (see SOM), 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 each electrode in the window. Classification was performed using an ordinal logistic regression model (Pedregosa-Izquierdo, 2016) 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, 2011). This cross-validation procedure splits the data in 80% training and 20% testing sets, while ensuring a balanced number of trials for each load condition. This split was repeated 1000 times and results for each subject and time point were averaged across these repetitions. Training data was standardized using the StandardScaler Scikit-Learn function, and test data was standardized using the mean and standard deviation of the training set.

Binned trial classification (within-subject & across-experiment)

Cross-training classification was used to test for generalization between the color (Exp 1) and orientation (Exp 2) conditions, and between the single-feature (Exps 1 and 2) and conjunction (Exp 3) conditions. These analyses followed the same procedures as the within-experiment classification except 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 Experiment 1 and 2 and tested on data from Experiment 3. All of these analyses were done within-subject, using the subset of subjects who completed all of the experiments involved in the analysis (Exp. 1 & 2, n = 24; Exp. 1, 2, & 3, n = 20).

Significance Testing

In all classification analyses, we tested if classification accuracy was significantly above chance at each time point using a paired, one-tailed *t*-test. Classification accuracy was compared to empirical chance accuracy, defined by testing the trained model on randomly shuffled trial labels (for more detail see Luck, 2021). Because we tested for significance at each timepoint (48 time bins between 0 and 1250 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 above chance (Figure 2, range of condition accuracies: .72-.97). In a one-way ANOVA within each experiment, there was a significant main effect of set size (Experiment 1, $F(3, 128) = 90.19$, $p < .001$; Experiment 2, $F(3, 120) = 32.93$, $p < .001$; Experiment 3, $F(3, 76) = 60.27$, $p < .001$). T-tests revealed no significant differences in accuracy across the experiments. Experiment 1 accuracy (M = .85, SD = .07) was not significantly different than Experiment 2 (M = .87, SD = .12), $t(62) = -1.43$, $p = .153$, or Experiment 3 (M = .85, SD = .11), $t(51) = -1.13$, $p = .262$. Experiment 2 was also not significantly different than Experiment 3, $t(49) = .105$, $p = .916$. Because there were twice as many features to remember with the conjunction stimuli in Experiment 3, near-equivalent performance between the dual- and single-feature conditions provides another example of “object-based benefits” for storage in visual WM (Olson and Jiang, 2002). That is, a larger number of feature values were stored in the conjunction condition, when the number of features per item was doubled compared to the single-feature conditions.

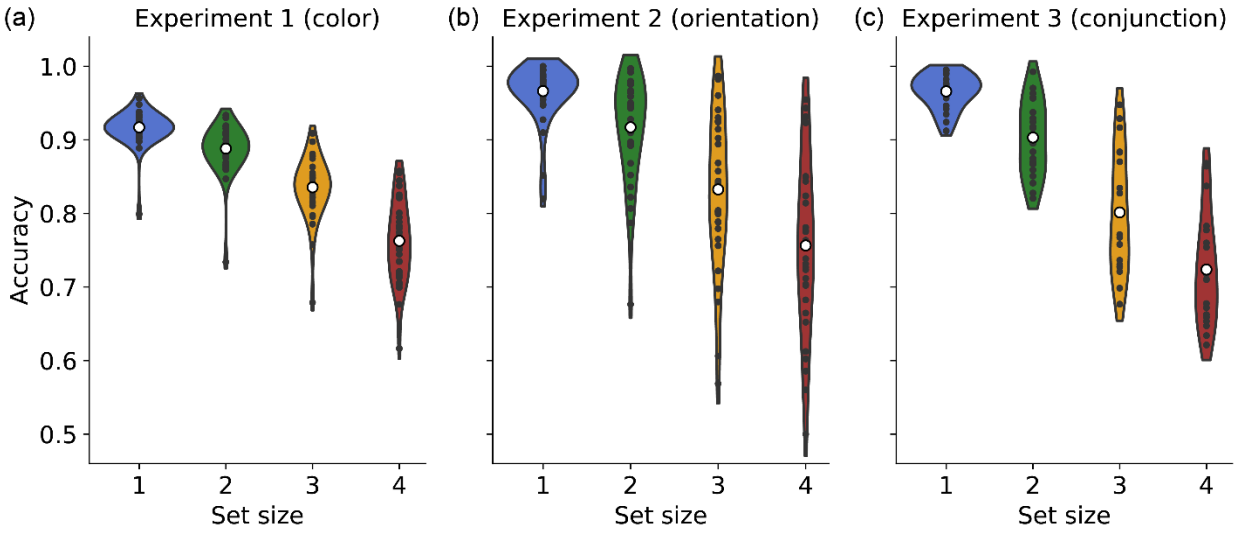

Figure 2 Change detection accuracy for each set size in each of the three experiments.

Precise classification of load while controlling for stimulus energy

The first key result was that the mvLoad analysis precisely classified working memory load, despite the use of stimulus displays that controlled for stimulus energy across all load conditions. For each experiment (Exp. 1, n = 33; Exp. 2, n = 31; Exp. 3, n = 20), we used an ordinal logistic regression classifier on raw EEG amplitudes from binned trials within-subject (20 trials per bin) at each time bin (50 ms window). We could classify working memory load (set size 1 vs. 2 vs. 3 vs. 4) during the stimulus presentation and throughout the delay period (Figure 3a, 3b, 3c; red squares indicate corrected $p < .05$, FDR-controlled at .05 with Benjamini-Hochberg procedure with 48 time-bins tested).

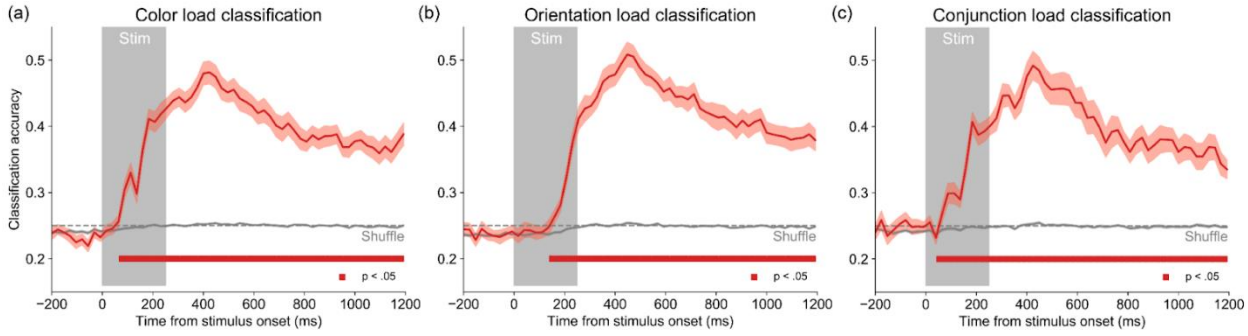

Figure 3 Classification accuracy over time for each experiment. Classification accuracy is indicated with a red line. Shaded error bars indicate ± SEM. Red squares indicate timepoints with classification significantly above chance (corrected p < .05, FDR = .05 with Benjamini-Hochberg procedure). Grey line indicates chance classification accuracy. Grey rectangle indicates time period where memory array is displayed.

Above-chance classification was observed starting in early time bins in each experiment (Exp. 1, 88-136 ms timebin; Exp. 2, 160-208 ms timebin; Exp. 3, 64-112 ms timebin). 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 .41 (SD = .04); Experiment 2, .43 (SD = .04); Experiment 3, .40 (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 sizes 1 vs. 2, 2 vs. 3, and 3 vs. 4. In all cases, accuracy was sustained above chance throughout the entire delay period (corrected $p < .05$, *FDR* = .05 with 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 working memory capacity (r^2 = .26, $p < .001$) (Figure 5). This correlation may be caused by

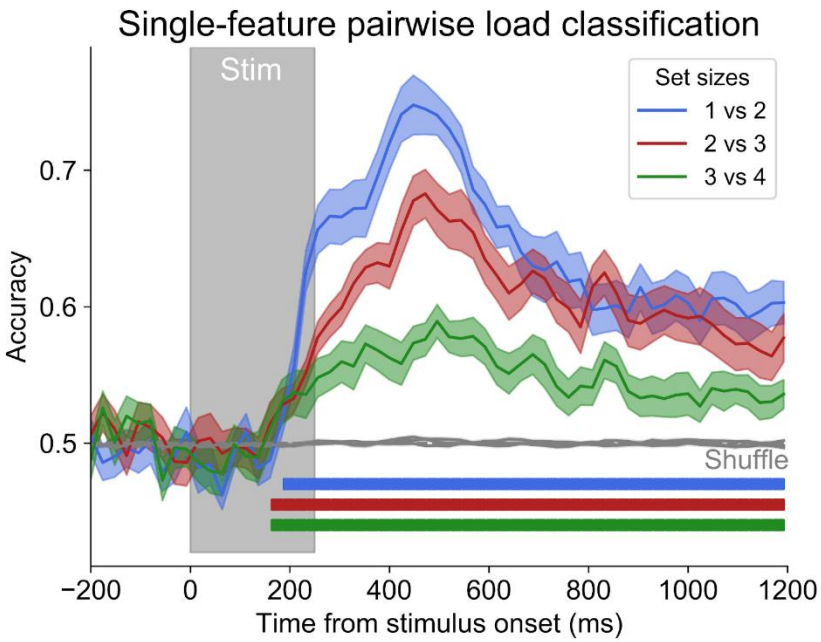

Figure 4 Classification accuracy of single-feature load (experiments 1 and 2 mixed together) for set sizes 1 vs. 2, 2 vs. 3, and 3 vs. 4. Colored lines indicate classification accuracy. Shaded error bars indicate ± *SEM.* Color-matched squares indicate timepoints with classification significantly above chance (corrected $p < .05$, *FDR* = .05 with Benjamini-Hochberg procedure). Grey line indicates chance classification accuracy. Grey rectangle indicates time period where memory array is displayed.

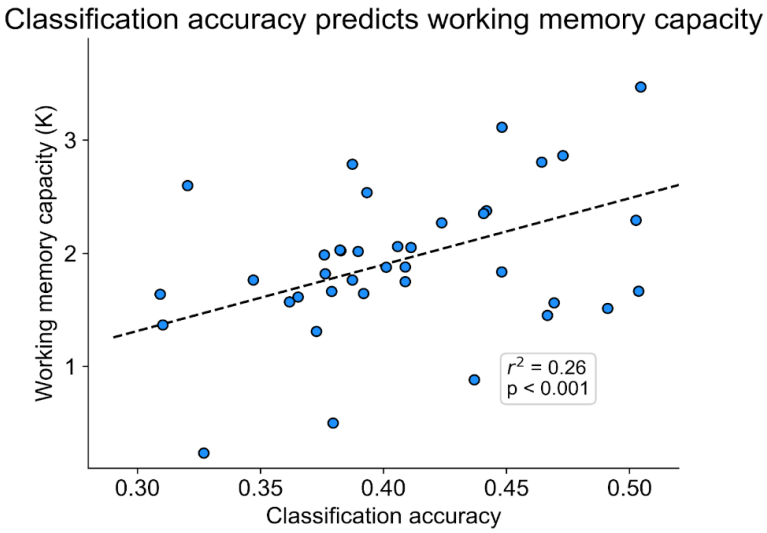

Figure 5 Individual differences in classification accuracy are positively correlated with working memory capacity. Classification accuracy is average delay period accuracy from all unique subjects across all three experiments; all data was used from each unique subject. Working

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 reinforces the earlier evidence that the mvLoad analysis taps into a neural operation that is relevant for understanding capacity limits in visual working memory.

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 (Figure 6). Mean classification accuracy during the delay period for color to orientation was .33 (SD = .03); orientation to color, .33 (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 working memory.

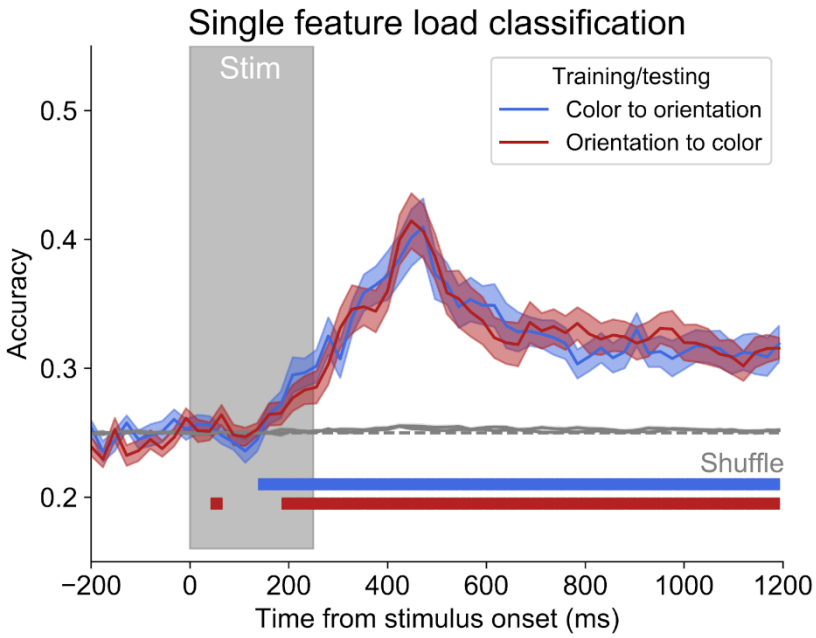

Figure 6 Accuracy for single-feature load classification. Blue line is classification accuracy when trained on data from experiment 1 (color) and tested on experiment 2 (orientation). Red line is accuracy when trained on experiment 2 and tested on experiment 1. Color-matched squares indicate timepoints with classification significantly above chance (corrected $p < .05$, *FDR* = .05 with Benjamini-Hochberg procedure). Grey line indicates chance classification accuracy. Grey rectangle indicates time period where memory array is displayed.

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 working memory. 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 Experiment 3 in which the number of relevant features per item was doubled (i.e., color and orientation) (Figure 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). Thus, 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 featural details.

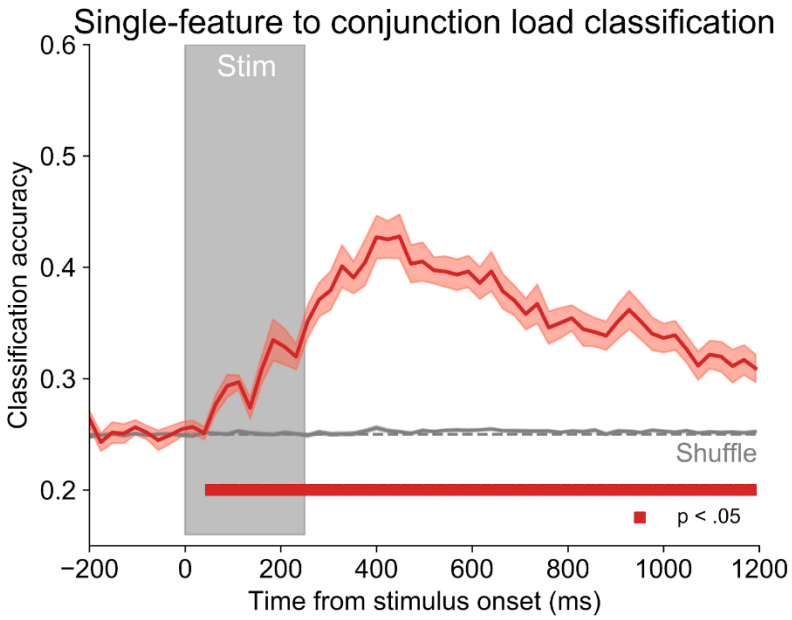

Figure 7 Accuracy for classification trained on data from experiments 1 and 2 (single-feature items, color *or* orientation) and tested on data from experiment 3 (conjunction items, color *and* orientation). Red squares indicate timepoints with classification significantly above chance (corrected $p < .05$, *FDR* = .05 with Benjamini-Hochberg procedure). Grey line indicates chance classification accuracy. Grey rectangle indicates time period where memory array is displayed.

While robust cross-training between single feature and conjunction conditions suggests a common load signature between the conditions, a more incisive test is to scrutinize the specific pattern of errors within each classification analysis. First, recall from Experiments 1 and 2 (Figure 3) that the mvLoad analysis robustly detected the difference between one and two single-feature items, 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 the mvLoad analysis decoding load based on the number of features stored, then a classifier trained with single feature stimuli should be biased to detect higher loads when it is tested with conjunction stimuli that contain twice as many stored features. Specifically, we should see that classification of set size 1 conjunction displays should be biased towards set size 2, and classification of set size 2 displays should be biased towards load 4. By contrast, if the classifier is instead detecting the number of *items* that are stored rather than the number of feature values, then we should see generalization from single feature to conjunction stimuli without any such bias. In line with the latter prediction, we found near-perfect generalization of the classifier when it was trained on a mixture of single-feature color and orientation stimuli, and tested with conjunction stimuli that each required the storage of both color and orientation. As Figure 8 illustrates, when the classifier was trained with single-feature stimuli, performance with conjunction stimuli that contained twice as many stored features was not biased towards higher set sizes. Instead, when the classifier was trained to discriminate between set sizes 1 and 2, precisely the same proportion of set size 1 and set size 2 trials were accurately classified as containing one and two items, respectively; a paired one-way t-test showed

that the percent of set size 1 trials mis-classified as set size 2 was not significantly greater in the single-feature to conjunction output (M = .34, SD = .12) than the single-feature output (M = .36, SD = .08), $t(19) = .584$, $p = .717$. The same result was observed with set size 2 trials; when the classifier was trained to discriminate between set sizes 2 and 4, the percentage of set size 2 trials mis-classified as set size 4 was not significantly greater in the single-feature to conjunction output (M = .32, SD = .16) than the single-feature output (M = .34, SD = .08), $t(19) = .367$, $p = .641$. Thus, the load signature identified by the mvLoad analysis was equivalent for single-feature and dual-feature objects.

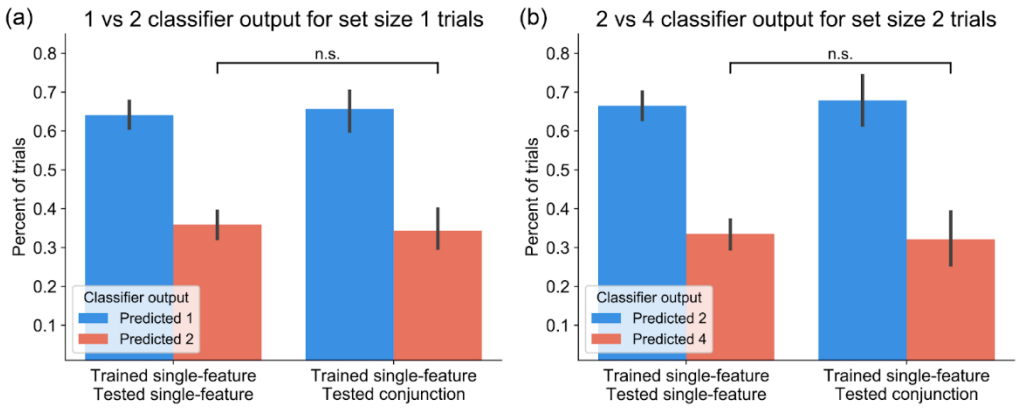

Figure 8 Classification predictions for (a) set size 1 trials when trained on set size 1 versus 2 and (b) set size 2 trials when trained on set size 2 versus 4. Left bars are trained and tested on a mixture of data from Experiments 1 and 2 (single-feature items). Right bars are trained on data from experiments 1 and 2 (single-feature items) and tested on Experiment 3 (conjunction items). Blue bars indicate correct classifications while red bars indicate mis-classifications (predicted as set size 2). There is no increase in the percent of trials mis-classified as (a) set size 2 or (b) set size 4 in the single-feature to conjunction output.

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

Although the results of the mvLoad analysis point 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 re-analyzed 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 working memory. In the *grouped* condition, the discs were arranged so that

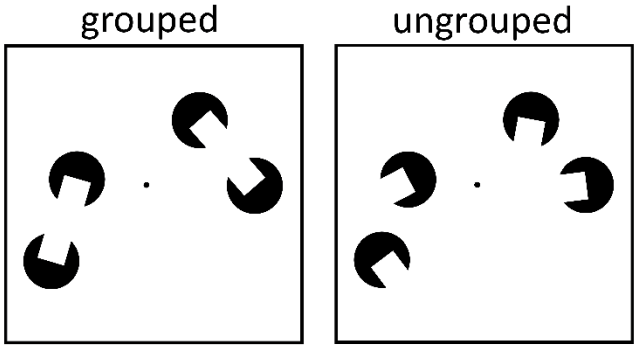

Figure 9 Examples of a set size 4 grouped and ungrouped memory arrays. In the grouped condition, collinearity between the notches yields the percept of a single oriented rectangle for each grouped pair.

collinearity between the notches in pairs of discs elicited the percept of a single illusory rectangle (Figure 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

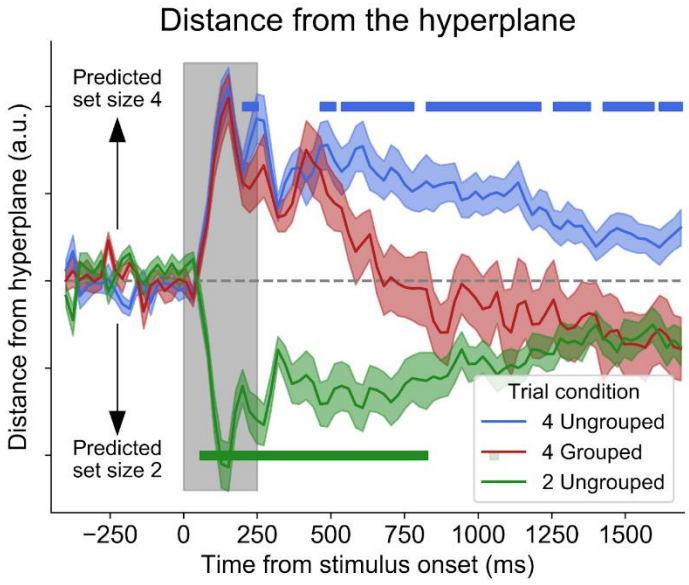

Figure 10 Distance from the classification hyperplane for set size 2 ungrouped, set size 4 grouped, and set size 4 ungrouped trials across time. Classification is trained on set size 2 and 4 ungrouped trials and tested on all three conditions. Hyperplane is indicated with the dashed grey line. Trials above the hyperplane are classified as set size 4 while trials below it are classified as set size 2. Distance from the hyperplane for each trial condition at each time point. Blue squares indicate timepoints where 4 ungrouped is significantly greater than 4 grouped (corrected $p < .05$, *FDR* = .05 with Benjamini-Hochberg procedure). Green squares indicate where 2 ungrouped is significantly less than 4 grouped (corrected $p < .05$, *FDR* = .05 with Benjamini-Hochberg procedure).

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. However, by the 488-512 ms time bin, set size 4 grouped had diverged away 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 time bin 824-848 ms, set size 4 grouped had crossed the hyperplane and was no longer consistently reliably different from set size 2 ungrouped. Thus, while 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 working memory serves as a cornerstone for intelligent behaviors, there is strong motivation to build a taxonomy of the neural operations that support online memory storage. The dominant strain of this work has focused on stimulus-specific neural activity that represents the stored content (e.g., Fuster and Jervey, 1981; Funahashi et al., 1989; Goldman-Rakic, 1995; Serences et al, 2009; Rademaker et al, 2019; D'Esposito and Postle, 2015), and great progress has been made in understanding the format and anatomical locus of this class of neural activity. By contrast, here 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 (Khaneman et al, 1992; Pylyshyn, 2009). Using a multivariate analytic approach (Adam et al., 2020), we show 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 robust 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 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., Vogel and Machizawa, 2004; Todd and Marois, 2004; Xu and Chun, 2006). But while 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" wherein 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 that isn't affected by the complexity of the memoranda (Xu and Chun, 2005; Woodman and Vogel, 2008). This empirical pattern suggests a neural operation that indexes the number of individuated representations stored in working memory 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 extremely robust with 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 working memory.

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., Pylyshyn, 2009; Khaneman et al., 1992; Xu and Chun, 2009). To study this cognitive process, Pylyshyn and Storm (1988) introduced “multiple object tracking” (MOT), a task that requires the observer to keep track of varying numbers of targets that move randomly amongst 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 (2009) used a lateralized version of the MOT task to show that CDA 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 CDA and mvLoad classifiers are picking up on a content-independent indexing operation that is required during tracking and visual working memory tasks (Tsubomi, Watanabe and Vogel, 2013; Balaban et al., 2019; Hakim et al., 2019).

In combination with stimulus-selective neural activity that supports the maintenance of precise memories (e.g., D’Esposito and Postle), 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 working memory. 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 while 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 and Chun, 2005). 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 multi-feature compared to 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 working memory, 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 online mental workspace.

Citations

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), e13691.

Anaconda Software Distribution. Computer software. Vers. 2-2.4.0. Anaconda, Nov. 2016. Web. Available at https://anaconda.com

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.

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.

Buschman, T. J. (2021). Balancing Flexibility and Interference in Working Memory. *Annual Review of Vision Science*, *7*.

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *Journal of Neuroscience*, *31*(25), 9315-9322.

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. *Journal of Neuroscience*, *28*(16), 4183-4191.

Feldmann-Wüstefeld, T. Neural measures of working memory in a bilateral change detection task. Psychophysiology. 2021; 58:e13683.

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.

Fukuda, K., Mance, I., & Vogel, E. K. (2015). α power modulation and event-related slow wave provide dissociable correlates of visual working memory. *Journal of Neuroscience*, 35(41), 14009-14016.)

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.

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.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., . . . Oliphant, T. E. (2020). Array programming with NumPy, *Nature*, 585, 357–362, DOI:10.1038/s41586-020-2649-2

Harrison S. A., Tong F. Decoding reveals the contents of visual working memory in early visual areas. Nature. 2009 Apr 2;458(7238):632-5. doi: 10.1038/nature07832. Epub 2009 Feb 18. PMID: 19225460; PMCID: PMC2709809.

Hunter, J. D (2007). Matplotlib: A 2D Graphics Environment, *Computing in Science & Engineering*, 9, 90-95, DOI:10.1109/MCSE.2007.55

Ikkai, A., McCollough, A. W., & Vogel, E. K. (2010). Contralateral delay activity provides a neural measure of the number of representations in visual working memory. *Journal of neurophysiology*, *103*(4), 1963-1968.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive psychology*, *24*(2), 175-219.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, 8, 213.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279-281.

McKinney, W. (2010). Data Structures for Statistical Computing in Python, *Proceedings of the 9th Python in Science Conference*, 51-56

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).

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., . . . 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*. Université Pierre et Marie Curie - Paris VI. English. ⟨NNT : 2015PA066015⟩. ⟨tel-01100921v2⟩
Peirce, J., Gray, J.R., Simpson, S. et al. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods,* 51, 195–203 (2019). https://doi.org/10.3758/s13428-018-01193-y

Poo, M. M., Pignatelli, M., Ryan, T. J., Tonegawa, S., Bonhoeffer, T., Martin, K. C., . . . & Stevens, C. (2016). What is memory? The present state of the engram. *BMC biology*, *14*(1), 1-18.

Pylyshyn, Z. (2009). Perception, representation, and the world: The FINST that binds. *Computation, cognition, and Pylyshyn*, 3-48.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, *3*(3), 179-197.

Python Software Foundation. Python Language Reference. Vers. 2.8.10. Available at http://www.python.org

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.

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature human behaviour*, *4*(11), 1156-1172.

Seabold, S. & Perktold, J. (2010). “statsmodels: Econometric and statistical modeling with python.” *Proceedings of the 9th Python in Science Conference*.

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. *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.

Virtanen, P., Gommers, R, Oliphant, T. E., Haberland, M, Reddy, T., Cournapeau, D., . . . van Mulbregt, P., and SciPy 1.0 Contributors. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3), 261-272.

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.

Waskom, M. L., (2021). seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021, https://doi.org/10.21105/joss.03021

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.

Figures

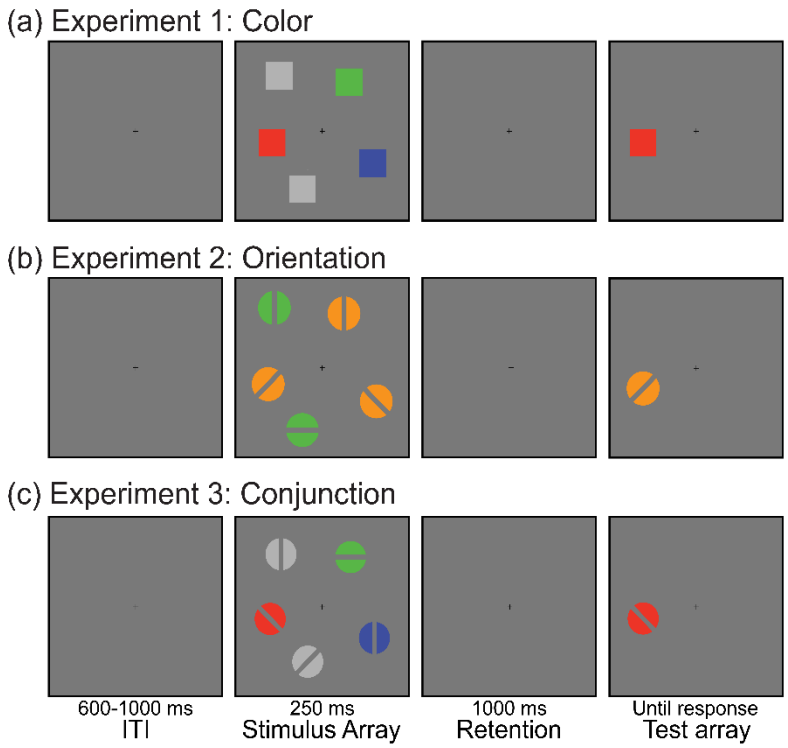

Figure 2 Task schematics for an example set size 3 trial in the whole-field change detection task used in all three experiments. (a) In Experiment 1. participants remembered the colored squares while ignoring the grey placeholders. (b) In Experiment 2. at the start of each blook, a color cue informed participants to attend and remember either the orange or the green orientations while ignoring the unced color. (c) In Experiment 3. participants 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.

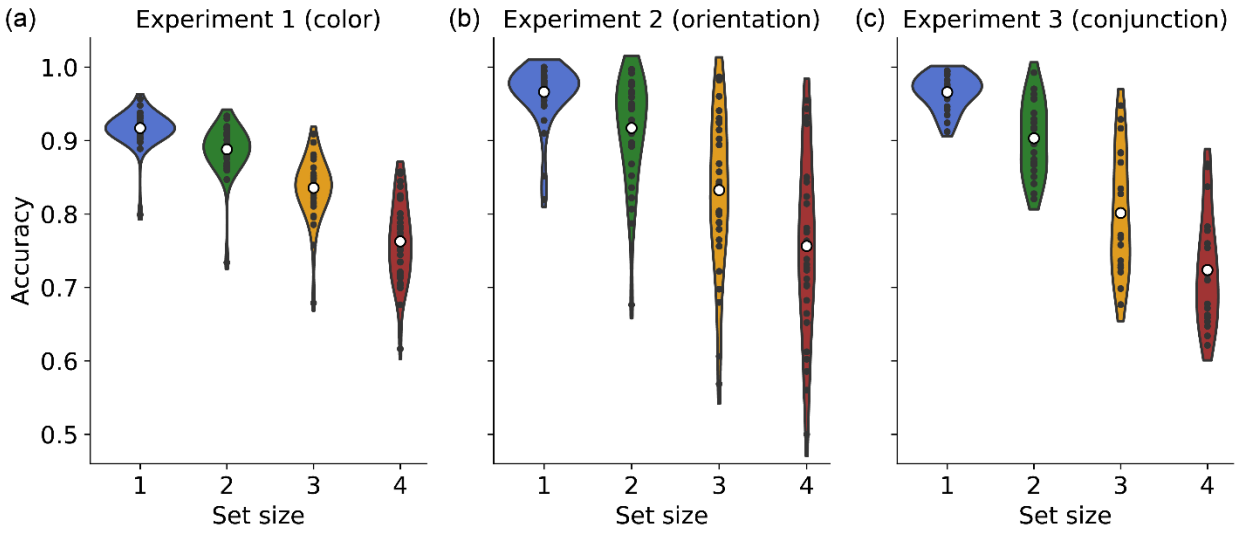

Figure 2 Change detection accuracy for each set size in each of the three experiments.

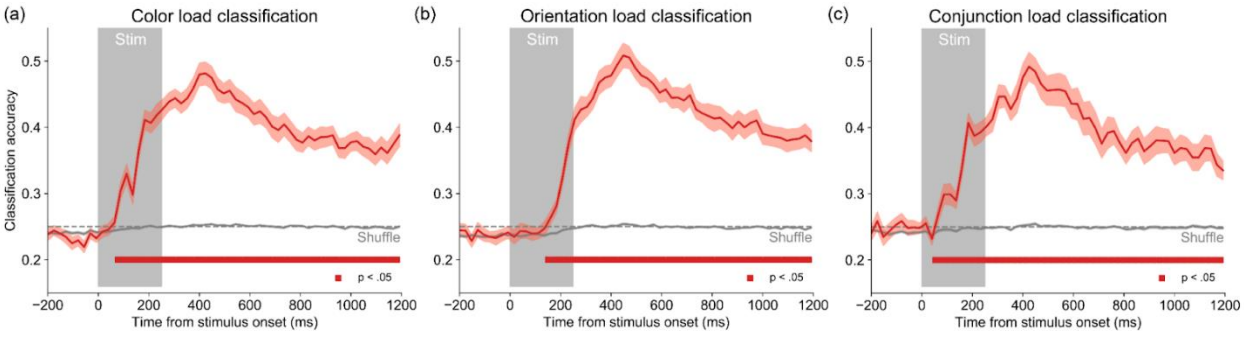

Figure 3 Classification accuracy over time for each experiment. Classification accuracy is indicated with a red line. Shaded error bars indicate ± SEM. Red squares indicate timepoints with classification significantly above chance (corrected $p < .05$, FDR = .05 with Benjamini-Hochberg procedure). Grey line indicates chance classification accuracy. Grey rectangle indicates time period where memory array is displayed.

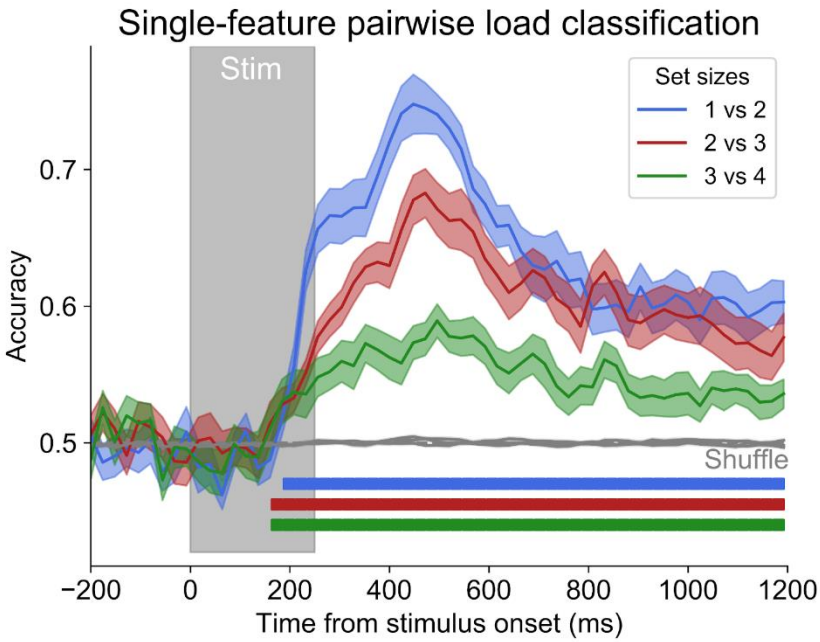

Figure 4 Classification accuracy of single-feature load (experiments 1 and 2 mixed together) for set sizes 1 vs. 2, 2 vs. 3, and 3 vs. 4. Colored lines indicate classification accuracy. Shaded error bars indicate ± *SEM.* Color-matched squares indicate timepoints with classification significantly above chance (corrected $p < .05$, *FDR* = .05 with Benjamini-Hochberg procedure). Grey line indicates chance classification accuracy. Grey rectangle indicates time period where memory array is displayed.

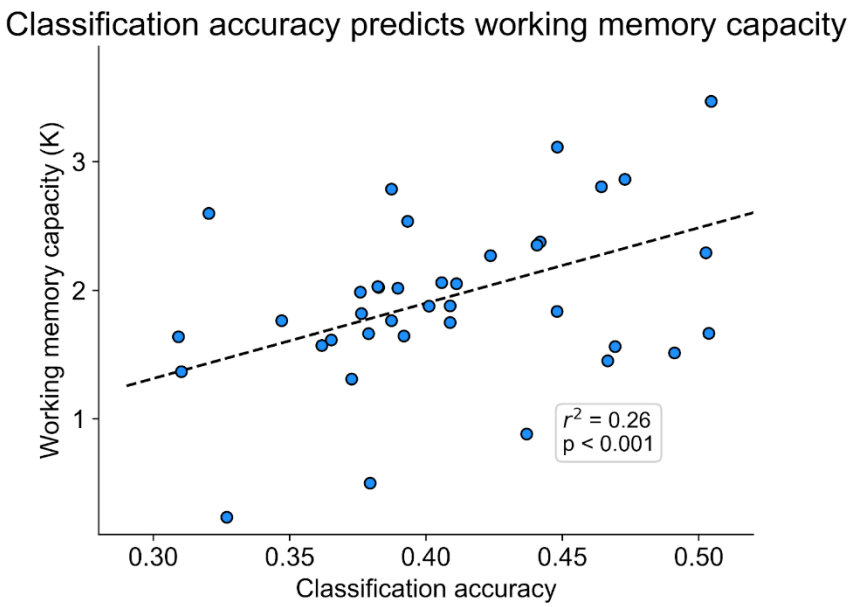

Figure 5 Individual differences in classification accuracy are positively correlated with working memory capacity. Classification accuracy is average delay period accuracy from all unique subjects across all three experiments; all data was used from each unique subject. Working memory capacity was measured using set size 4 trials.

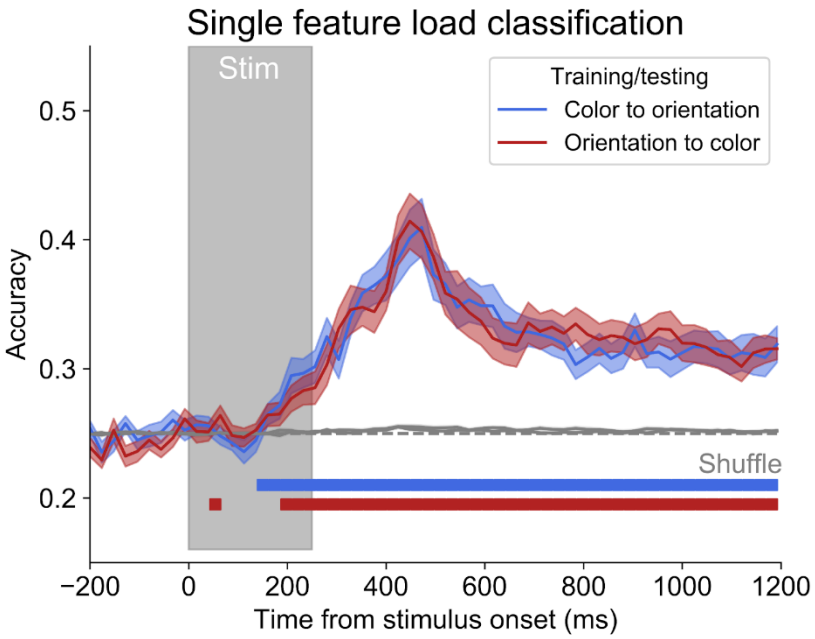

Figure 6 Accuracy for single-feature load classification. Blue line is classification accuracy when trained on data from experiment 1 (color) and tested on experiment 2 (orientation). Red line is accuracy when trained on experiment 2 and tested on experiment 1. Color-matched squares indicate timepoints with classification significantly above chance (corrected $p < .05$, *FDR* = .05 with Benjamini-Hochberg procedure). Grey line indicates chance classification accuracy. Grey rectangle indicates time period where memory array is displayed.

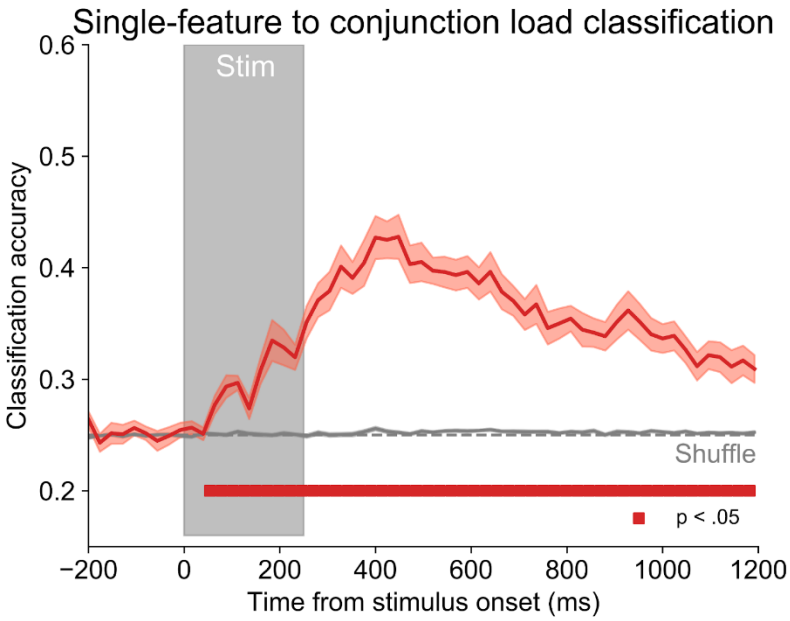

Figure 7 Accuracy for classification trained on data from experiments 1 and 2 (single-feature items, color *or* orientation) and tested on data from experiment 3 (conjunction items, color *and* orientation). Red squares indicate timepoints with classification significantly above chance (corrected $p < .05$, *FDR* = .05 with Benjamini-Hochberg procedure). Grey line indicates chance classification accuracy. Grey rectangle indicates time period where memory array is displayed.

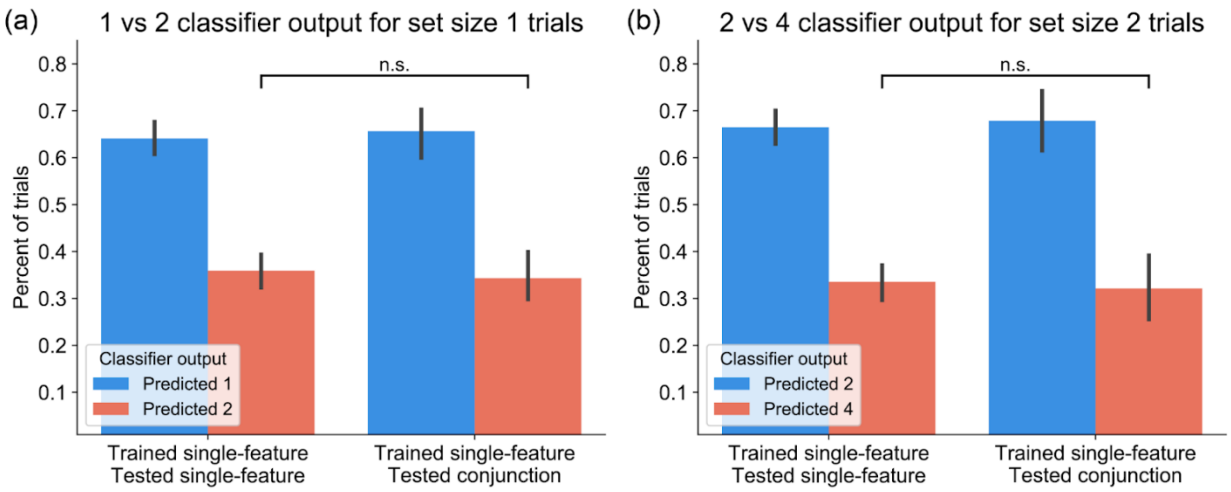

Figure 8 Classification predictions for (a) set size 1 trials when trained on set size 1 versus 2 and (b) set size 2 trials when trained on set size 2 versus 4. Left bars are trained and tested on a mixture of data from Experiments 1 and 2 (single-feature items). Right bars are trained on data from experiments 1 and 2 (single-feature items) and tested on Experiment 3 (conjunction items). Blue bars indicate correct classifications while red bars indicate mis-classifications (predicted as set size 2). There is no increase in the percent of trials mis-classified as (a) set size 2 or (b) set size 4 in the single-feature to conjunction output.

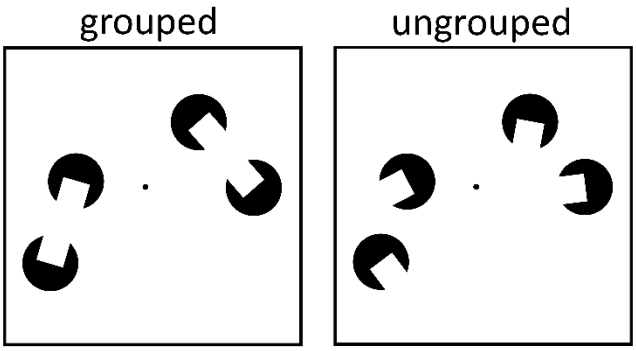

Figure 9 Examples of a set size 4 grouped and ungrouped memory arrays. In the grouped condition, collinearity between the notches yields the percept of a single oriented rectangle for each grouped pair.

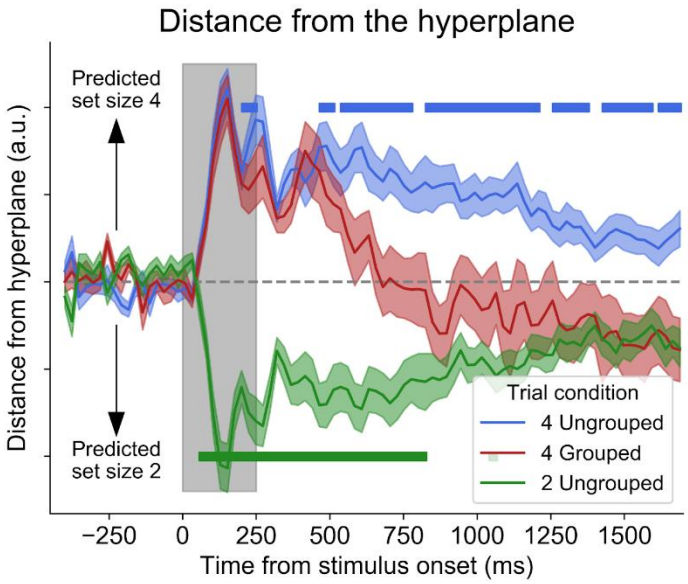

Figure 10 Distance from the classification hyperplane for set size 2 ungrouped, set size 4 grouped, and set size 4 ungrouped trials across time. Classification is trained on set size 2 and 4 ungrouped trials and tested on all three conditions. Hyperplane is indicated with the dashed grey line. Trials above the hyperplane are classified as set size 4 while trials below it are classified as set size 2. Distance from the hyperplane for each trial condition at each time point. Blue squares indicate timepoints where 4 ungrouped is significantly greater than 4 grouped (corrected $p < .05$, *FDR* = .05 with Benjamini-Hochberg procedure). Green squares indicate where 2 ungrouped is significantly less than 4 grouped (corrected $p < .05$, *FDR* = .05 with Benjamini-Hochberg procedure).