Content-Independent Pointers Mediate Working Memory Storage for Both Visual and Verbal Stimuli

Woohyeuk Chang[1], Will Epstein[2], Henry Jones[1], William X. Q. Ngiam[3], and Edward Awh[1]

Abstract

■ Prominent models of working memory (WM) have long argued that visual and verbal WM rely on distinct storage buffers. Although this claim has been bolstered by both behavioral and neural studies that reveal important differences between visual and verbal storage in WM, recent work has also provided neural evidence for a domain-general signature of WM storage that is independent of the specific content stored. Our working hypothesis is that this content-independent load activity tracks the deployment of a "pointer" operation that supports contextual binding of the stored items. Here, we provide evidence that this pointer signal generalizes between simple colors and meaningful words. Across three data sets, including a reanalysis of previously published data and two novel EEG experiments, we demonstrate that multivariate EEG patterns track the number of items regardless of whether stimuli are visual (colored rectangles), verbal (letters or words), or conjunctions of both (colored words). Our findings demonstrate a robust load signal that remains stable across changes in perceptual format and stimulus type and is distinct from neural signals tracking spatial attention and cognitive effort. Critically, this pointer signal was observed in parallel with distinct EEG activity that tracked the stored content, suggesting a dissociation between the item-based indexing of stored items and the maintenance of their features. ■

INTRODUCTION

Working memory (WM) is a limited-capacity system that supports the temporary maintenance and manipulation of information. Prominent models of WM argue for separate storage subsystems for different types of information. For example, Baddeley and colleagues proposed a visuospatial sketchpad for visual and spatial information and a distinct phonological loop for verbal information (Baddeley, 2000; Baddeley & Hitch, 1974). Dual-task studies have provided support for this model by demonstrating selective patterns of interference with visuospatial and verbal WM tasks. Concurrent performance of two tasks within the same modality (e.g., two verbal tasks) leads to a significantly stronger decrease in performance compared to the performance of two tasks across verbal and visual modalities (e.g., Fougnie, Zughni, Godwin, & Marois, 2015; Logie, Zucco, & Baddeley, 1990). Moreover, neuroimaging studies have identified distinct brain networks for visual (Xu & Chun, 2006; Todd & Marois, 2004) and verbal (Rottschy et al., 2012; Awh et al., 1996; Paulesu, Frith, & Frackowiak, 1993) WM systems.

These findings notwithstanding, recent work has also highlighted important commonalities across visual and verbal WM. Rajsic, Burton, and Woodman (2019) reported that the contralateral delay activity (CDA), a well-characterized EEG signal of visual WM storage (Luria, Balaban, Awh, & Vogel, 2016; McCollough, Machizawa, & Vogel, 2007; Vogel & Machizawa, 2004), showed parallel amplitude increases that plateaued at similar set sizes as the number of colors, letters, and meaningful words increased in WM. This finding suggests that there may be a common neural signature of WM storage for verbal and visual materials. This falls in line with a growing body of evidence for common neural signatures of WM storage across distinct visual features (Thyer et al., 2022; Adam, Vogel, & Awh, 2020), even for cortically disparate features such as color and motion (Jones, Thyer, Suplica, & Awh, 2024). Critically, these studies have also shown that distinct variance in ongoing neural activity tracks the specific features that are stored, revealing the parallel operation of content-independent load signals, and stimulus-specific activity related to the features of the stored items.

Motivated by these findings, the current project had two central goals. First, we reanalyzed the Rajsic and colleagues (2019) data set using multivariate approaches that provided converging evidence for a common signature of storage across visual and verbal modalities while also documenting parallel but distinct EEG activity that tracked the content (visual or verbal) of the stored materials. Second, we conducted two novel studies that replicated these findings while providing more rigorous controls for bottom–up stimulus-evoked activity and ruling out

[1]University of Chicago, [2]National Institute of Mental Health, Bethesda, MD, [3]Adelaide University, Adelaide, South Australia

Journal of Cognitive Neuroscience 38:7, pp. 1419–1436
https://doi.org/10.1162/JOCN.a.2522

changes in covert spatial orienting as the source of content-independent load signals. Our findings suggest that a common neural operation mediates the WM storage of visual and verbal information, whereas parallel but separate operations maintain the specific features associated with the stored items. Moreover, further analyses ruled out cognitive effort as a source of content-independent neural activity. We interpret these findings in a broader framework in which domain-general process enables the contextual bindings of the items stored in WM, whereas feature-specific neural activity supports the maintenance of featural details.

RAJSIC ET AL. REANALYSIS

We reanalyzed EEG data from the original study by Rajsic and colleagues (2019), in which participants performed a lateralized change detection task using either colored rectangles or letters (Study 1) and colored rectangles or meaningful words (Study 2). This data set allowed us to examine whether WM load signals generalize across perceptually distinct stimulus categories, specifically across visual and verbal domains.

Methods

Participants and Procedures

Rajsic and colleagues compared CDA between colored squares and letters for Study 1 ($n = 12$) and between colored squares and words for Study 2 ($n = 12$). Both experiments shared the same procedure: an intertrial blank display for 2200 msec, ±200 msec of jitter, 500-msec fixation display, 100-msec cue display, 900-msec fixation display, 500-msec memory sample display, 1000-msec fixation display, and a memory test display that persisted until response (Figure 1A). Participants completed 1536 trials across four blocks.

All EEG data from both studies were baseline corrected by subtracting the mean voltage from the 200-msec interval preceding the memory array display. Participants were excluded if the cue-locked averaged horizontal EOGs exceeded ±3 μV or if more than 33% of their epochs contained artifacts (e.g., blinks, large saccades). ERPs were computed for each condition after artifact rejection, using MATLAB (The MathWorks). CDA amplitude was quantified as the mean voltage from 350 to 1500 msec.

Data from both studies were reanalyzed.

Multivariate Decoding

We used the multivariate load analysis framework (Thyer et al., 2022), a within-participant decoding method using logistic regression (*LogisticRegression* in Scikit-learn; Pedregosa et al., 2011), for our decoding analyses. A general workflow is as follows. From all available electrodes, EEG amplitudes were baseline-corrected at each time point by the mean amplitudes of EEG activities from 200 msec to 0 msec before the stimulus onset. For higher signal-to-noise ratio, we randomly selected 20 trials within each selected condition and averaged across the trials in each group. We then averaged the voltage for each electrode across 50-msec windows with 25-msec increments. The classifier was trained to discriminate between the selected conditions and then tested on held-out data (*StratifiedShuffleSplit* in Scikit-learn; Pedregosa et al., 2011) through cross-validation of 70% to 30% train–test split. Both training and test data were standardized using the mean and standard deviation of the training data at each time point. We repeated this cross-validation procedure 1000 times and averaged the results across all iterations.

For our decodability measure, we used the hyperplane contrast (Jones, Thyer, et al., 2024; Walther et al., 2016), where we recorded signed distance from the fitted model's decision boundary, also known as the hyperplane, for each test-trial bin and averaged within the same conditions. Greater contrast between the two conditions' distances from the decision boundary indicates better decodability.

For load decoding, we assessed the decodability within each condition as well as its generalizability to other conditions (i.e., how well would the model decode load on word condition when trained on color condition, and vice versa). We also examined whether we could decode the attended feature when load conditions were balanced across distinct features.

In all decoding analyses, we tested whether decodability, represented by hyperplane contrast, was significantly above chance at each time point using a paired-samples, one-tailed t test. Specifically, we tested whether the hyperplane contrast was significantly greater than zero, which served as the null hypothesis representing chance-level decoding. To test for imperfect generalization, we assessed whether the hyperplane contrast for one condition was significantly different from another condition using a paired-samples, two-tailed t test. For multiple testing correction, we used the Benjamini–Hochberg procedure to control the false discovery rate (FDR) at 0.05.

Representational Similarity Analysis

We used representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008) to complement our multivariate decoding analyses. Although multivariate decoding provides a powerful tool for detecting whether discriminative information is present in the data, the resulting classification performance can be challenging to interpret, because of models' tendency to exploit any available signals that help distinguish between given conditions. In contrast, RSA offers a more interpretable, hypothesis-driven approach by assessing the similarity structure of neural patterns to theoretical models. Here, we use RSA to test whether feature-specific signals are present and

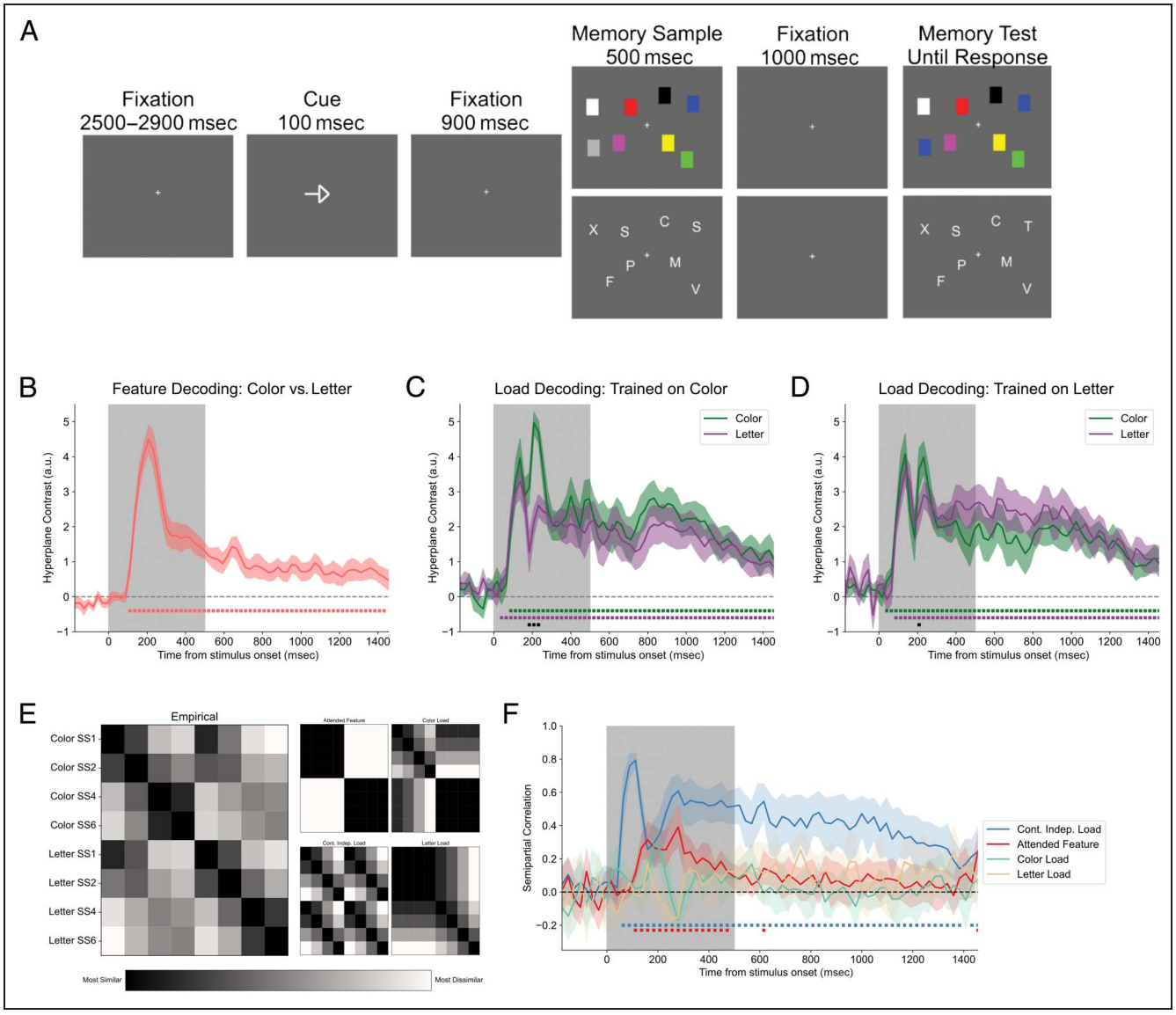

Figure 1. Rajsic et al. Study 1. (A) A single-trial design adopted from Rajsic and colleagues (2019). Main stimuli were colored rectangles and letters with set sizes 1, 2, 4, and 6. (B) Attended feature decodability. (C) Within-feature and across-feature load decodability when trained on color conditions. (D) Within-feature and across-feature load decodability when trained on letter conditions. The gray vertical rectangle indicates the time when the memory array was displayed. The colored squares represent time points where the hyperplane contrast values are significantly greater than zero (corrected $p < .05$, FDR = .05 with the Benjamini–Hochberg procedure). The black squares represent time points when decodability between two conditions significantly differs from one another (corrected $p < .05$, FDR = .05 with the Benjamini–Hochberg procedure). (E) Empirical RDM, averaged across the delay period and across participants, along with four theoretical RDMs (content independent [Cont. Indep.] load, attended feature, color load, letter load). All RDMs are organized by attended feature (color, letter) and set size (1, 2, 4, 6). (F) Semipartial correlations between empirical and theoretical RDMs estimated across time points, averaged across participants. The gray vertical rectangle indicates the time when the memory array was displayed. The colored lines represent semipartial correlations of a given theoretical model to the empirical data. The colored squares represent time points where semipartial correlations are significantly greater than zero (one-tailed Wilcoxon signed-rank test with FDR correction using the Benjamini–Hochberg procedure).

to examine whether a content-independent load signal persists even after accounting for potential confounds.

Our RSA approach closely followed the method described in Jones, Thyer, and colleagues (2024), which used time-resolved, within-participant RSA to assess the neural similarity across conditions. For each participant, we computed the empirical representational dissimilarity matrices (RDMs) between every pair of conditions at each time point via cross-validated Mahalanobis distance. We then compared the empirical RDMs with theoretical RDMs (e.g., attended feature, feature-specific load, content-independent load) using Spearman rank correlation. We tested the significance using a one-tailed Wilcoxon signed-rank test against zero for each time window with FDR correction using the Benjamini–Hochberg procedure, consistent with the procedure outlined by Jones, Thyer, and colleagues (2024).

Results

Study 1: Colored Rectangles and Letters

We assessed whether we could decode the number of items held in WM (Thyer et al., 2022) and whether this load signal was generalizable to other conditions of interest. Using set sizes 1 and 4 as representative conditions (Figure 1A), we trained feature-specific classifiers and tested them both within and across features. We found robust decoding of attended feature information across memory sample display and delay period, with a peak decodability around 200 msec from stimulus onset (Figure 1B). In addition, we were able to successfully decode the load (Figure 1C and D). In addition to the sustained within-feature decodability across the delay period for both features, we saw near-perfect generalization across features, except for a few time points during the memory sample display.

Next, we examined the neural similarity structure across time using RSA. Unlike multivariate decoding methods where we only used two conditions of interest, we utilized all available conditions (eight total conditions: 4 set sizes × 2 features; Figure 1E). We found evidence for two primary signals in the EEG data: a content-independent load signal and an attended feature signal (Figure 1F). We found a strong and sustained content-independent load signal shortly after the stimulus onset (window centered on 64 msec) that lasted throughout the end of the delay period. Next, we found a transient attended feature signal during the memory sample display period (start: window centered on 112 msec; end: window centered on 496 msec). We did not find evidence for both color- and letter-specific load signals (no significant time points across 1500 msec from stimulus onset).

Study 2: Colored Rectangles and Meaningful Words

We asked whether findings in Study 1 would extend to words (Figure 2A), stimuli with stronger semantic associations than letters. Using set sizes 1 and 3 as representative conditions, we again trained classifiers and tested within and across features. Similar to Study 1, we observed reliable decoding of attended feature information, with a peak decodability around 200 msec from stimulus onset (Figure 2B). Multivariate load decoding was again robust within each feature, as well as near-perfect generalization across features (Figure 2C and D).

RSA provided evidence for multiple signals (Figure 2E and F). First, we again found an evidence for a content-independent load signal shortly after the stimulus onset (start: window centered on 64 msec; end: window centered on 1456 msec) as well as a transient attended feature signal during the memory sample display period (start: window centered on 112 msec, end: window centered on 496 msec). Finally, we found evidence for transient color-specific load signals at two time windows: one briefly after the stimulus onset (start: window centered on 136 msec, end: window centered on 232 msec) and another during the start of the delay period (start: window centered on 616 msec, end: window centered on 880 msec). We saw no evidence for a word-specific load signal.

Discussion

In our reanalysis of Rajsic and colleagues (2019), we identified a robust content-independent load signal that generalized across both visual (e.g., colored rectangles) and verbal (e.g., letters and words) stimuli. This finding closely paralleled earlier results showing that letters and words evoke a capacity-limited CDA pattern that is comparable with that elicited by simple colored objects. Notably, although Rajsic and colleagues reported a main effect of stimulus type on CDA amplitude, with larger amplitudes for words and for colors, this stimulus effect did not interact with set size, such that CDA peaked at the same set sizes for both types of stimuli. Moreover, we found no difference in the decodability or cross-feature generalizability of our multivariate load signal across attended features. This suggests that greater CDA amplitude observed for words may not reflect increased storage demands per se but could instead stem from stimulus-driven differences unrelated to WM load. Employing the same multivariate decoding framework, we also found evidence for an attended feature signal, with peak decodability emerging around 200 msec after stimulus onset. This was expected given the perceptual differences between the conditions (e.g., colored rectangles vs. printed letters and words). Finally, RSA nicely captured the unique variance explained by the content-independent load signal and the attended feature signal, underscoring the functional dissociation between storage-related and feature-specific activity during WM storage (Suplica et al., 2025; Jones, Thyer, et al., 2024; Xu & Chun, 2009).

EXPERIMENT 1: COLORED WORDS

Despite the presence of content-independent load signals that generalized across visual and verbal stimuli, the original task design included several confounds that may challenge the interpretations of our decoding results. In Rajsic and colleagues (2019), visual and verbal conditions differed not only in the content but also in their perceptual features, with colored rectangles used for visual stimuli and letters and words used for verbal stimuli. Thus, although we observed robust decoding of the stored feature, independent of WM load, these findings do not establish whether this was because of goal-driven storage or bottom–up differences in stimulus-evoked activity. Likewise, although decoding of WM load was robust and generalizable across words, letters, and colors, WM load was confounded with total sensory energy and the total number of attended locations in the memory display. Although these confounds do not undermine measurement of

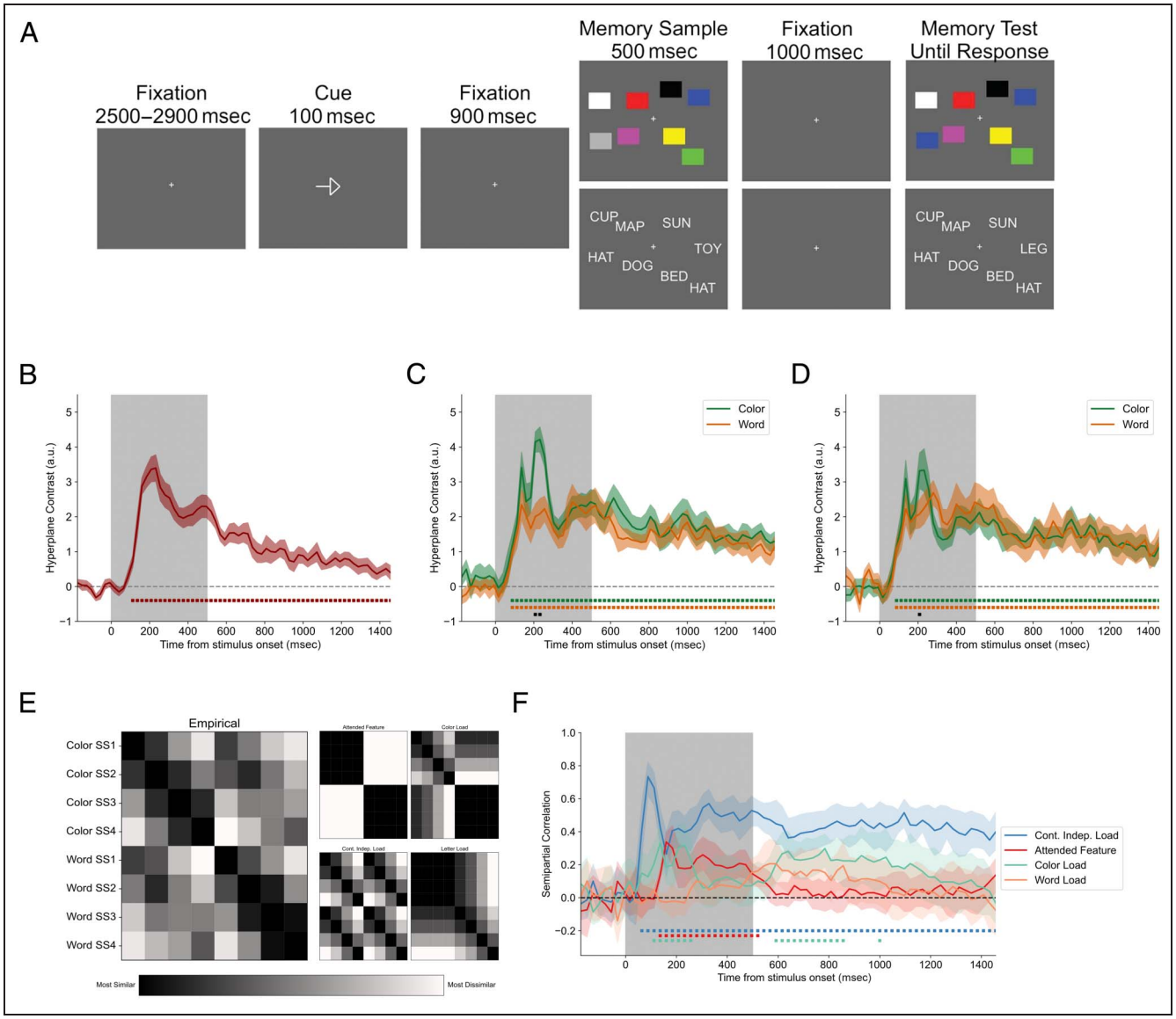

Figure 2. Rajsic et al. Study 2. (A) A single-trial design adopted and modified from Rajsic and colleagues (2019). Main stimuli were colored rectangles and meaningful words with set sizes 1, 2, 3, and 4. (B) Attended feature decodability. (C) Within-feature and across-feature load decodability when trained on color conditions. (D) Within-feature and across-feature load decodability when trained on word conditions. The gray vertical rectangle indicates the time when the memory array was displayed. The colored squares represent time points where the hyperplane contrast values are significantly greater than zero (corrected $p < .05$, FDR = .05 with the Benjamini–Hochberg procedure). The black squares represent time points when decodability between two conditions significantly differs from one another (corrected $p < .05$, FDR = .05 with the Benjamini–Hochberg procedure). (E) Empirical RDM, averaged across the delay period and across participants, along with four theoretical RDMs (content independent [Cont. Indep.] load, attended feature, color load, word load). All RDMs are organized by attended feature (color, word) and set size (1, 2, 3, 4). (F) Semipartial correlations between empirical and theoretical RDMs estimated across time points, averaged across participants. The gray vertical rectangle indicates the time when the memory array was displayed. The colored lines represent semipartial correlations of a given theoretical model to the empirical data. The colored squares represent time points where semipartial correlations are significantly greater than zero (one-tailed Wilcoxon signed-rank test with FDR correction using the Benjamini–Hochberg procedure).

lateralized CDA (Ikkai, McCollough, & Vogel, 2010), they do raise alternative explanations for neural decoding of load and attended features. Thus, the new studies we present were designed to address these limitations. To replicate and extend the findings from Rajsic and colleagues (2019), we used a design optimized for isolating content-independent neural signals while controlling for other variables that are often confounded with WM load. First, we used colored words as target stimuli and manipulated which features were attended; this eliminated the perceptual differences between visual and verbal memory conditions. Second, we included gray placeholder items to equate the stimulus energy across set sizes, eliminating sensory energy as a driver of the load classifier. Finally, we replaced the change-detection task with a lateralized two-alternative forced-choice (2AFC) task, which provides more sensitive and bias-free estimates of memory performance for individually probed items compared with

yes/no change-detection tasks (Brady, Robinson, Williams, & Wixted, 2023). Using the same multivariate EEG decoding and RSA as in the reanalysis of the original study, we asked whether content-independent load signals persisted under these more tightly controlled conditions.

Methods

Participants

Participants between 18 and 35 years old with normal or corrected-to-normal visual acuity were recruited from the University of Chicago and the surrounding community. Experimental procedures were approved by the institutional review board at the University of Chicago. All participants provided informed consent and were compensated for their participation at a rate of $20 per hour. The target sample size was 18 participants based on both empirical and methodological considerations. First, we consulted the publicly available CDA power calculator (Ngiam, Adam, Quirk, Vogel, & Awh, 2021), which indicated that a sample size of approximately 16–18 participants provides adequate power (≥.80) to detect load-dependent differences in CDA amplitude under conditions similar to our design. Second, we aligned our sample size with prior studies using same multivariate EEG decoding approaches that demonstrated reliable decoding performances with sample sizes in the range of 12–20 participants (Jones, Diaz, Ngiam, & Awh, 2024; Jones, Thyer, et al., 2024; Thyer et al., 2022). From 23 participants (14 female; mean age = 26.6 years, SD = 4.32), we excluded data from five participants because of the following reasons: system crashed during the experiment (n = 1), EEG data were missing for the first half of the experiment (n = 1), or participant's data had excessive EEG artifacts (n = 3).

Task Procedures

Participants performed a lateralized 2AFC task consisting of four sequential displays per trial: fixation, cue, memory sample, and memory test. Each trial began with a central black fixation cross (width = 0.8, height = 0.8) presented for a randomized duration between 1000 and 1500 msec against a mid-gray background (~61 cd/m^2). Participants were instructed to maintain central fixation and avoid blinking during this period. Next, a centrally presented black arrow (width = 0.8, height = 0.8) appeared for 400 msec, cueing attention to either the left or right visual hemifield. After the cue, the memory sample display was presented for 500 msec. It contained either one or three colored words (width = 2, height = 0.8) in each hemifield. The words were randomly selected from eight possible options ("BED," "CUP," "DOG," "HAT," "LEG," "MAP," "SUN," and "TOY") and colored using one of eight colors (RGB values: red = 170, 0, 0; orange = 255, 128, 0; yellow = 255, 255, 0; green = 0, 255, 0; blue = 0, 0, 255; purple = 128, 0, 128; black = 0, 0, 0; white = 255, 255, 255). Gray "XXX" placeholders (red, green, blue [RGB] value = 166, 166, 166) were used to control for visual stimulus energy across conditions. Items were displayed within a rectangular region (width = 9, height = 6), with a minimum spacing of 2.2 to prevent overlap. After a 1000-msec retention interval with central fixation, the memory test display appeared. A white outlined rectangle (width = 2, height = 0.8) indicated the location of the to-be-tested item, with the target and foil words (width = 2, height = 0.8) presented above and below the rectangle. Participants responded using the up or down arrow key with their right hand. The test display remained on screen until a response was made.

Experiment 1 followed a 2 × 2 within-participant design: Attended Feature (color or word) × Set Size (1 or 3). Each participant completed 1200 trials (300 trials per condition), divided into 20 blocks of 60 trials. Attended feature was blocked, whereas set size varied randomly within blocks. Participants were allowed to take breaks between blocks. The total session lasted approximately 3 hr.

EEG Acquisition

We recorded EEG activity from 30 active Ag/AgCl electrodes mounted in an elastic cap (actiCHamp, Brain Products GmbH). We recorded from International 10–20 sites Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, C3, Cz, C4, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO7, PO3, PO4, PO8, O1, Oz, and O2. Two additional electrodes were placed on the left and right mastoids, and a ground electrode was placed at position Fpz. All sites were recorded with a reference to the right mastoid and offline rereferenced to the algebraic average of the left and right mastoids. We recorded EOGs using passive electrodes with a ground electrode placed on the participant's left cheek. Horizontal EOG was recorded with a bipolar pair of electrodes placed ~1 cm from the external canthus of each eye. Vertical EOG was recorded with a bipolar pair of electrodes placed above and below the right eye. Data were filtered online (low cutoff = 0.01 Hz, high cutoff = 80 Hz, slope from low-to-high cutoff = 12 dB/octave) and were digitized at 500 Hz using BrainVision Recorder (Brain Products GmBH) running on a PC. The impedance values were set to be below 10 kΩ.

Eye Tracking

We recorded gaze position using a desk-mounted infrared eye-tracking system (EyeLink 1000 Plus, SR Research). Gaze position was sampled at 1000 Hz. We calibrated the eye tracker at the beginning of the experiment and recalibrated between the trials during the blocks if necessary. Drift correction was done every five trials.

Artifact Rejection

We segmented the EEG data into epochs time-locked to the onset of memory array (200 msec before and until 1500 msec after the stimulus onset). We used an automated pipeline to detect ocular (eye movements of over 1° visual angle away from fixation) or EEG artifacts (e.g., skin potentials, muscle artifacts, and excessive noise) and flag the corresponding trials for later manual inspection. We excluded trials with EEG activity greater than 80 μV or less than −80 μV and with peak-to-peak activity greater than 100 μV within a 200-msec window with 100-msec steps. Rejected trials were excluded from both behavioral and EEG analyses. For Experiment 1, participants were excluded from the final sample if the remaining trials per condition were fewer than 200 trials. For Experiment 2, participants were excluded if the remaining trials per condition were fewer than 150 trials.

CDA

Preprocessed EEG data from Experiment 1 were baselined from 200 msec to 0 msec before the stimulus onset. EEG data were then low-pass filtered at 25 Hz and averaged separately for each condition and separately for electrodes ipsilateral and contralateral to the attended side (left or right). The difference between contralateral and ipsilateral activity for the electrode pairs PO7/PO8, PO3/PO4, P7/P8, P3/P4, and O1/O2 was calculated, resulting in five CDA waveforms. The average CDA amplitude for each condition was calculated for time windows between 350 and 1500 msec after the stimulus onset and averaged across participants. Only the correct trial data were used to calculate CDA.

Data Analysis

Data were analyzed using the same two approaches (multivariate decoding and RSA), as described in the Methods section for the reanalysis of Rajsic and colleagues (2019). To further examine whether cognitive efforts could account for content-independent load signals, we trained the classifier on pupil size, a well-established proxy for effort in cognitive tasks (van der Wel & van Steenbergen, 2018).

Results

Behavioral

Across both conditions, participants performed the task with above-chance accuracy (Figure 3B). By conducting a repeated-measures ANOVA, we found a significant main effect of Set Size, indicating that accuracy declined as set size increased, $F(1, 17) = 107.53, p < .001$. We did not find a significant main effect of Attended Feature, $F(1, 17) = 3.13, p = .09$, or an interaction of Set Size and Attended Feature, $F(1, 17) = 1.09, p = .31$. For z-scored RT (-Figure 3C), we found a significant main effect of both Set Size, $F(1, 17) = 180.72, p < .001$, and Attended Feature, $F(1, 17) = 40.42, p < .001$, but no reliable interaction between the two, $F(1, 17) = 0.37, p = .55$.

CDA

CDA was computed as the mean voltage of the difference wave (contralateral − ipsilateral) between 350 and 1500 msec from stimulus onset. Repeated-measures ANOVA revealed a significant main effect of Set Size, $F(1, 17) = 25.35, p < .001$, a load-dependent CDA pattern with increasing CDA voltage as a function of set size (Figure 3D). However, unlike the previous study where they found higher CDA for words compared with colored rectangles (Rajsic et al., 2019), we did not see a main effect of Attended Feature, $F(1, 17) = 1.75, p = .20$, nor an interaction between the two, $F(1, 17) = 3.81, p = .07$. This difference is likely because of our use of matched stimulus displays across the color and word conditions, supporting the hypothesis that prior differences in CDA amplitudes for these stimuli were because of stimulus-driven activity rather than WM storage.

Multivariate Decoding

We assessed whether we could still decode WM load for conditions that had no perceptual differences. Following the same analytic approach described in the previous section, we investigated attended feature decoding as well as load decoding for both within-feature and across-features. First, we successfully decoded the attended feature for most of the time points (Figure 4A), starting from a window centered on 232–856 msec, as well as the latter half of the delay period (start: window centered on 1168 msec, end: window centered on 1408 msec). A breakdown by set size revealed different temporal dynamics (Figure 4B). For Set Size 1 conditions, we saw transient feature signals, starting from a window centered on 232–688 msec. For Set Size 3 conditions, we saw more sustained feature signal that started at a window centered on 256 msec and lasted till the end of the delay period. Next, we evaluated the decodability of WM load. We observed strong within-feature decodability, both when trained on the attend-color condition (Figure 4C) and when trained on the attend-word condition (Figure 4D). Finally, we saw essentially perfect across-feature generalizability of the load signal that persisted throughout the memory sample display period to the end of the delay period.

Could this shared signature of WM load across colors and words be explained by common changes in cognitive effort? To examine this question, we measured pupil size, a well-accepted index of effort within displays that are matched for stimulus energy. Using the same decoding procedure, we trained the classifiers on pupil size data from the attend-color and attend-word conditions (Figure 4E and F). Although both analyses yielded above-

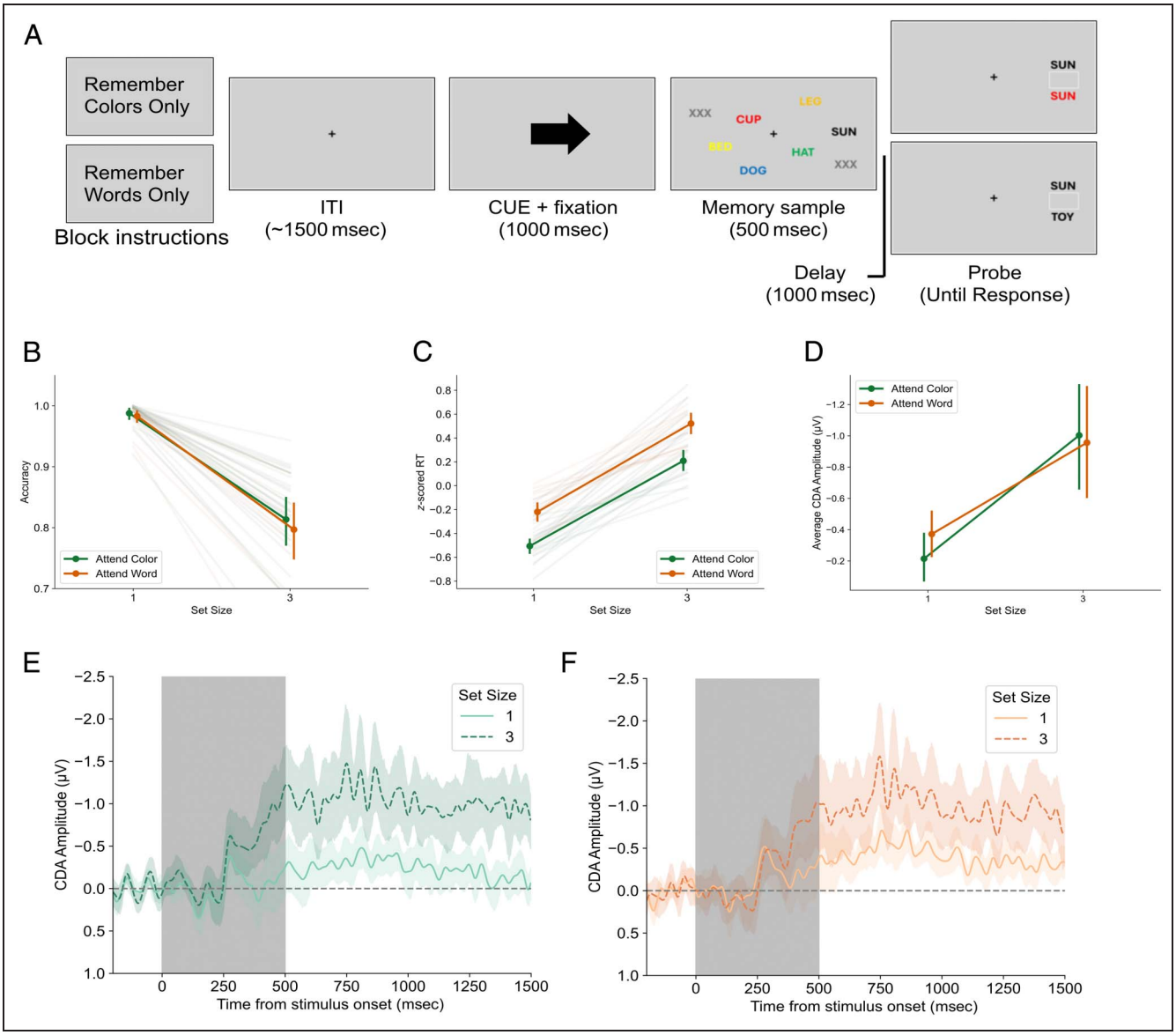

Figure 3. Experiment 1 task procedure, behavioral, and CDA results. (A) A single-trial design for Experiment 1. Participants were cued at the beginning of each block to remember either the color or word of the memory items while ignoring the gray placeholder items (e.g., "XXX" in gray). (B) 2AFC accuracy result for each set size. (C) 2AFC RT result for each set size, *z* scored within participant. (D) Mean CDA amplitude for the interval of 350–1500 msec. Dim colored lines indicate individual participant data. Error bars indicate 95% confidence intervals, calculated using 1000 iterations of bootstrapped resampling of the data. (E) CDA amplitude by set size for attend-color conditions. (F) CDA amplitude by set size for attend-word conditions.

chance decoding of load, the temporal profile of these effects diverged markedly from the EEG-based load dynamics, with pupil-based decoding emerging (window centered on 640 msec) after robust signatures of content-independent load were detected. These findings converge with multiple past studies that have revealed a clear disconnect between content-independent load patterns and various measures of effort (Jones, Thyer, et al., 2024).

RSA

We complemented the results found in multivariate decoding analyses with RSA. To improve the reliability of RSA while maintaining trial independence, we implemented an odd–even split approach (Kikumoto, Sameshima, & Mayr, 2022; Kikumoto & Mayr, 2020). Because RSA requires estimating RDM across multiple experimental conditions, the number of unique conditions can be limited in designs with relatively few stimulus or load combinations (four in our case). To increase the effective number of independent condition estimates without altering the underlying data, we divided the trials within each condition into two nonoverlapping subsets, odd trials and even trials, yielding a total of eight condition estimates (2 modalities × 2 set sizes × 2 trial parities). The RDM was then computed between

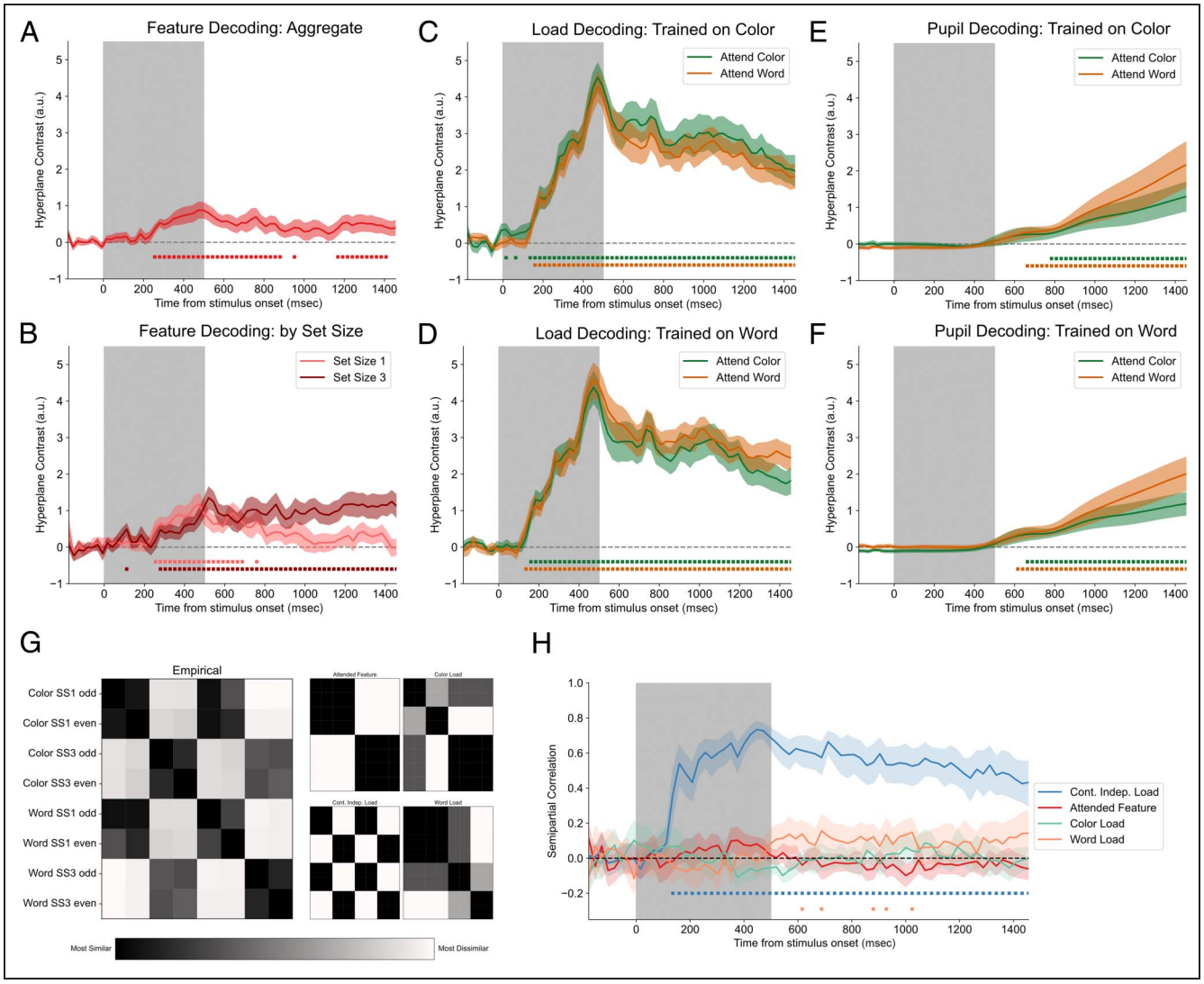

Figure 4. Experiment 1 multivariate decoding and RSA results. (A) Aggregate attended feature decodability. (B) Attended feature decodability for each set size conditions. (C) Within-feature and across-feature load decodability when trained on "attend-color" conditions. (D) Within-feature and across-feature load decodability when trained on attend-word conditions. (E) Within-feature and across-feature load decodability when trained on attend-color conditions using pupil data. (F) Within-feature and across-feature load decodability when trained on attend-word conditions using pupil data. The gray vertical rectangle indicates the time when the memory array was displayed. The colored squares represent time points where the hyperplane contrast values are significantly greater than zero (corrected $p < .05$, FDR corrected using the Benjamini–Hochberg procedure). (G) Empirical and four theoretical RDMs, organized by attended feature (color, word), set size (1, 3), and trial parity (odd, even). (H) Semipartial correlations between empirical and theoretical RDMs estimated across time points, averaged across participants. The colored lines represent semipartial correlations of a given theoretical model to the empirical data. The colored squares represent time points where semipartial correlations are significantly greater than zero (one-tailed Wilcoxon signed-rank test with FDR correction using the Benjamini–Hochberg procedure). Cont. Indep. = content independent.

these independent subsets, ensuring that all similarity estimates were based on independent data partitions (Figure 4G).

We found evidence for a content-independent load signal, which remained robust and consistent across most of the time points, right after the stimulus onset (Figure 4H). We did see several significant time points for word-specific load signal, but this pattern was not as robust and sustained as the content-independent load signal. Finally, we did not find evidence for attended feature and color-specific load signals.

Discussion

Experiment 1 replicated and extended the evidence for content-independent load signals using a more controlled design. By replacing visually dissimilar stimuli with colored words and equating stimulus energy across set sizes, we ruled out low-level perceptual confounds in Rajsic and colleagues (2019). Despite the removal of such perceptual differences, both multivariate decoding and RSA revealed a persistent, generalizable load signal that was not tied to a specific feature type. At the same time, the attended

feature was robustly decoded, ruling out the possibility that observers stored both features in every condition. Thus, these results provide an evidence for a domain-general aspect of WM storage.

Notably, although RT was somewhat faster for the color conditions, the CDA amplitude did not differ between color and word conditions. This suggests that the higher CDA amplitudes Rajsic and colleagues observed for words over colors were because of a stimulus-driven contralateral negativity that does not reflect WM load (e.g., Woodman & Vogel, 2008). We also observed reliable decoding of the attended feature, despite matched perceptual inputs, particularly at higher set sizes. RSA confirmed the presence of a content-independent load signal while revealing only weak and inconsistent signals for feature-specific representations, particularly when only one item was stored in WM. One possible explanation for this empirical pattern is that participants (despite our instructions) may have encoded both features, especially at lower set sizes, where automatic word reading could have led to incidental dual encoding of features (Stroop, 1935). Critically, feature decoding was robust at higher set sizes, ruling out dual encoding as an explanation for content-independent load activity.

Finally, to evaluate whether this common signature of WM load across colors and words was a reflection of common changes in cognitive effort (i.e., higher set sizes require more effort), we measured pupil size. This well-established physiological marker of cognitive effort did support above-chance decoding of WM load, but the temporal dynamics of pupil-based decoding significantly deviated from the EEG-based load signals. Specifically, pupil-based decoding did not onset until 500 msec after the onset of the content-independent load signal. Although this large difference in time course casts doubt on whether effort alone can explain the generalized neural signatures of WM load, we acknowledge that pupil-based signals may be sluggish compared with voltage-based signals. We will return to this issue in Experiment 2, where a more robust generalization of voltage-based load signals provides a stronger argument against the effort-based interpretation.

EXPERIMENT 2: COLORS, WORDS, AND COLORED WORDS

The results from Experiment 1 demonstrated the presence of content-independent load signals, even when there were no perceptual differences between conditions. However, this design included a confound between the number of individuated items stored in WM and the number of attended positions in the visual display. Prior work has highlighted the tight coupling between spatial attention and visual WM, making it difficult to fully disentangle the neural mechanisms supporting each process (Foster, Sutterer, Serences, Vogel, & Awh, 2017; Chun, 2011; Awh & Jonides, 2001; but see Hakim, Feldmann-Wüstefeld, Awh, & Vogel, 2021; Fukuda, Mance, & Vogel, 2015). Nevertheless, there is growing evidence for a functional dissociation between WM storage, on the one hand, and covert spatial orienting, on the other. Jones, Diaz, and colleagues (2024) directly addressed this issue by independently manipulating the number of memory items and the spatial extent of attention by using "dot cloud" stimuli that afforded strong manipulations of stimulus breadth. They found that distinct aspects of EEG activity tracked the number of items stored in WM and the spatial extent of covert spatial attention. Thus, to extend our initial findings, we disentangled WM load and the breadth of spatial attention by sequentially presenting the memory items within a single central position, eliminating the confound between the number of stored items and the spatial extent of attention.

In addition, results from Experiment 1 revealed only transient decodability of the attended feature signals at a lower set size, suggesting that participants may have incidentally encoded both color and word features when memory demands were low. This may reflect automatic processing of verbal stimuli, such as obligatory word reading (Stroop, 1935). To strongly encourage selective encoding of the relevant feature, we modified the 2AFC test displays such that only task-relevant feature information was shown (e.g., colored rectangles when testing color, gray characters when testing word). We also introduced two additional "single-feature" conditions: color-only condition, where words were replaced by colored placeholders (e.g., "XXX"), and word-only condition, in which all texts appeared in gray. These single-feature conditions allowed us to benchmark the neural signals associated with remembering a single feature in isolation and to compare these directly against conditions where the same feature had to be selectively remembered from conjunction stimuli (i.e., colored words).

Taken together, Experiment 2 aimed to (1) eliminate the confound between WM load and spatial attention via sequential, central presentations and (2) evaluate the neural similarity and distinction between single-feature and conjunction-based memory representations for the same feature type as well as across feature types.

Methods

Participants

Participants between 18 and 35 years old with normal or corrected-to-normal visual acuity were recruited from the University of Chicago and the surrounding community. Experimental procedures were approved by the institutional review board at the University of Chicago. All participants provided informed consent and were compensated for their participation at a rate of $20 per hour.

The target sample size was 18 participants, informed by the outcomes of Experiment 1. From 27 participants (15 female; mean age = 25.0 years, SD = 3.93 years), we excluded data from eight participants because of the following reasons:

system crashed during the experiment (n = 1), participant quit in the middle of the experiment (n = 1), and participant's data had excessive EEG or ocular artifacts (n = 6). A total of 19 participants were ultimately included.

Task Procedures

Participants performed a sequential WM task using a similar task structure as Experiment 1, but with centrally presented memory items. Each trial began with a fixation display shown for a randomized duration between 1000 and 1500 msec. This was followed by a sequential memory sample display in which three items (including placeholders as needed for set size) were presented one at a time at fixation. For better readability, colors were adjusted accordingly (RGB values: red = 255, 0, 0; orange = 255, 128, 0; yellow = 255, 255, 0; green = 0, 255, 0; cyan = 0, 255, 255; blue = 0, 0, 255; magenta = 255, 0, 255; black = 0, 0, 0), as well as a gray placeholder color (RGB value = 155, 155, 155). Each item was shown for 150 msec, separated by a 350-msec ISI. A 1000-msec fixation display followed the final memory item display. The memory test display consisted of a target and foil item shown bilaterally. Depending on the attended feature condition, participants saw either two colored rectangles (width = 2, height = 0.8) or two gray words (width = 2, height = 0.8). Participants responded using the left or right arrow key with their right hand. Display remained visible until a response was made.

Experiment 2 employed a 4 × 2 within-participant design: Attended Feature (color only, word only, attend color, attend word) × Set Size (1 or 3). In the color-only condition, all words were replaced with "XXX" and varied in color, and in the word-only condition, words were presented in gray; the attend-color and attend-word conditions were identical to those in Experiment 1. Each participant completed 1600 trials (200 trials per condition), divided into 20 blocks of 80 trials. Attended feature was blocked, and set size varied within blocks. Participants were allowed to take breaks between blocks. The total session lasted approximately 3.5 hr.

Data Collection, Processing, and Analysis

Data collection, processing, and analysis procedures were identical to those in Experiment 1.

Results

Behavioral

Across all conditions, participants performed the task with above-chance accuracy (Figure 5B). By conducting a repeated-measures ANOVA, we found a significant main effect of Set Size, $F(1, 18) = 21.33$, $p < .001$, Condition Type, $F(1, 18) = 5.40$, $p = .002$, and an interaction between the two, $F(1, 18) = 12.39$, $p < .001$. To understand the source of this interaction, we conducted post hoc comparisons focusing on accuracy at the Set Size 3

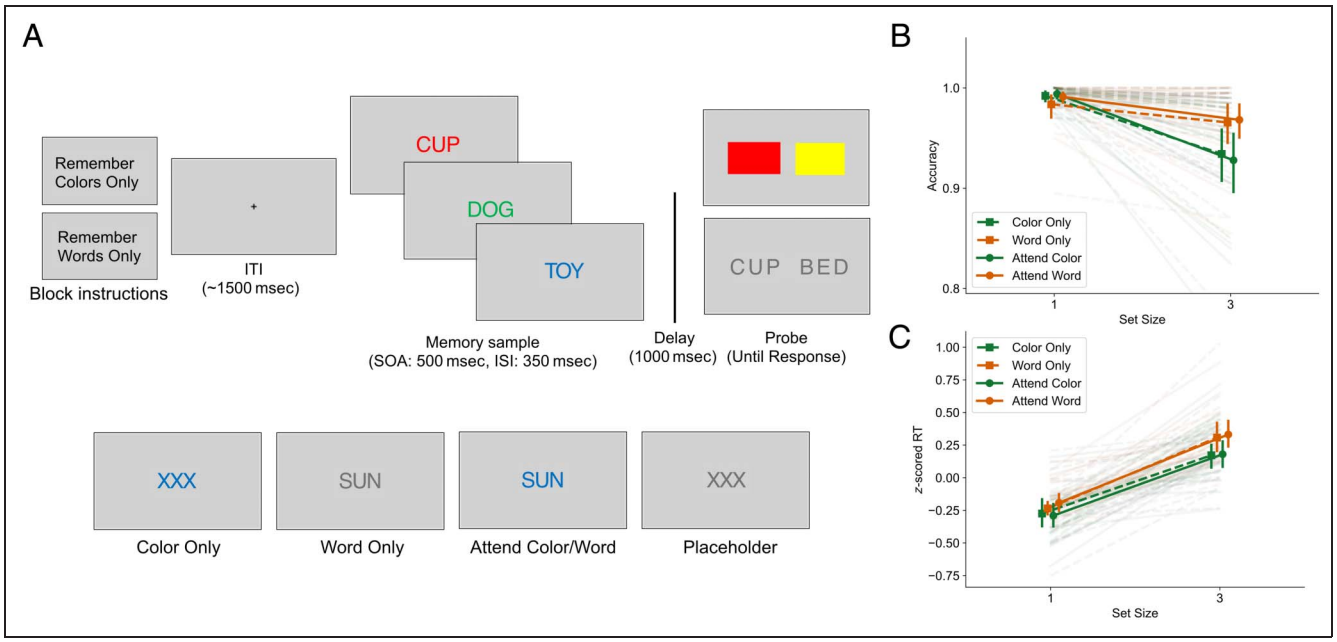

Figure 5. Experiment 2 task procedure and behavioral results. (A) A single-trial design for Experiment 2. Participants were cued at the beginning of each block to remember either the color or word of the memory items while ignoring the relevant placeholder items. Depending on which feature is cued in the beginning, participants see either two colored rectangles or two gray words bilaterally. Bottom four panels show sample displays for single-feature conditions (e.g., color only, word only), conjunction conditions (e.g., attend color/word), and placeholder. (B) 2AFC accuracy result for each condition. (C) 2AFC RT result for each condition, z scored within participant. Dim colored lines indicate individual participant data. Error bars indicate 95% confidence intervals, calculated using 1000 iterations of bootstrapped resampling of the data.

condition. Participants were significantly more accurate in the attend-word condition than in the attend-color condition, $t(18) = 3.19$, $p < .001$, and a similar pattern was observed for the word-only versus color-only conditions, $t(18) = 2.94$, $p < .001$. In addition, accuracy significantly declined from Set Size 1 to Set Size 3 in both the color-only, $t(18) = 4.73$, $p < .001$, and attend-color, $t(18) = 4.61$, $p < .001$, conditions, whereas this decline was less pronounced in the word-based conditions (word only: $t(18) = 2.37$, $p = .03$; attend word: $t(18) = 3.32$, $p = .004$). For z-scored RT (Figure 5C), we found a significant main effect of Set Size, $F(1, 18) = 84.39$, $p < .001$, but not for Condition Type, $F(1, 18) = 2.47$, $p = .07$, nor an interaction between the two, $F(1, 18) = 1.49$, $p = .23$.

Multivariate Decoding

Conjunction stimuli. Contrary to the attended feature decoding results from Experiment 1, we were able to reliably decode the attended feature information throughout most of the delay period (Figure 6A), including at Set Size 1 (Supplemental Figure S1). For Set Size 1 conditions, decodability was significant starting moments after the

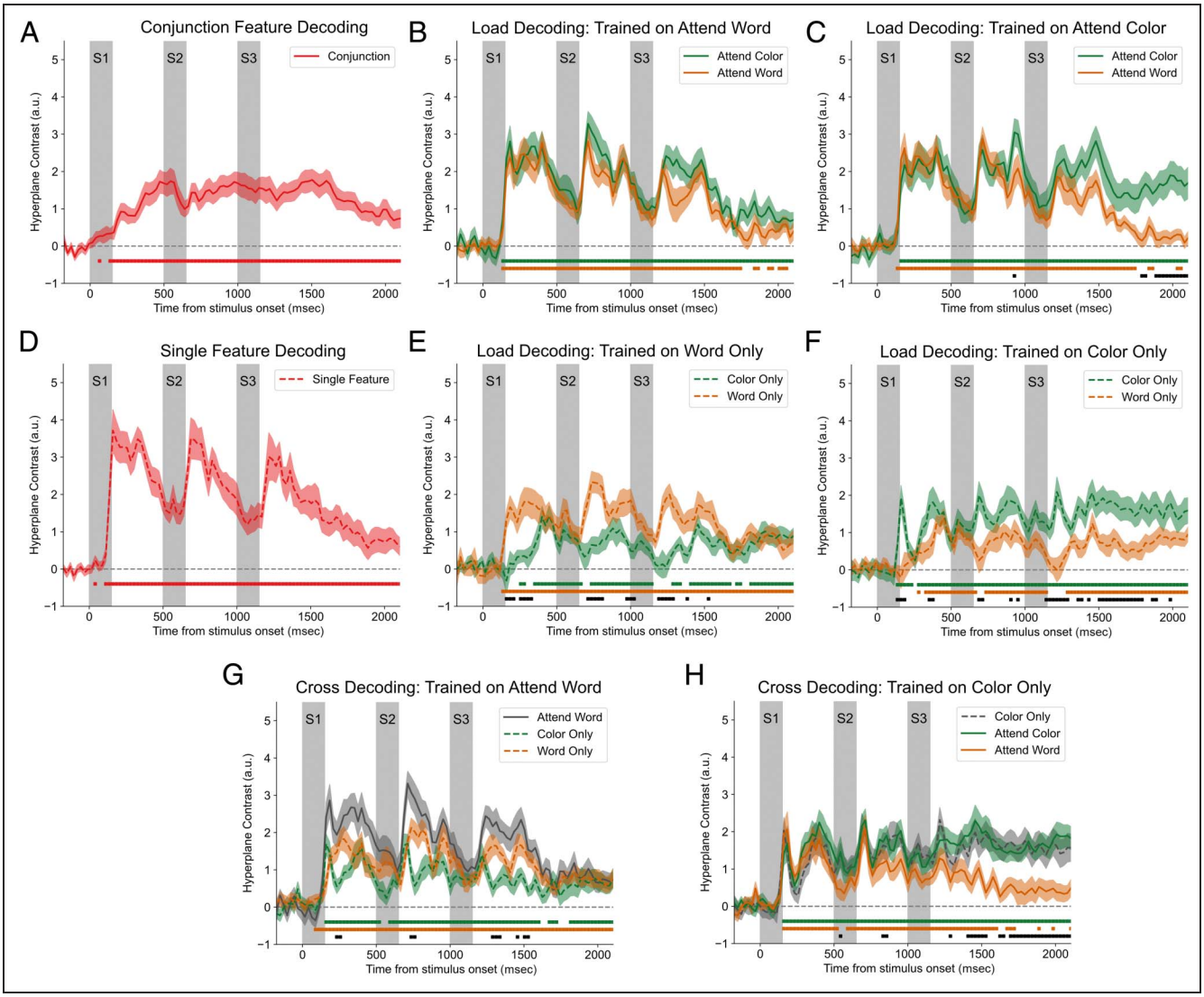

Figure 6. Experiment 2 multivariate decoding results. The first row represents the results from conjunction stimuli conditions. The second row represents the results from single-feature stimuli conditions. The last row represents the cross-decoding results across conditions. (A, D) Aggregated attended feature decodability. (B, E) Within-feature and across-feature decodability when trained on attend-word/word-only conditions. (C, F) Within-feature and across-feature decodability when trained on attend-color/color-only conditions. (G) Cross-decoding generalization when trained on attend-word conditions and tested on single-feature stimuli conditions. (H) Cross-decoding generalization when trained on color-only conditions and tested on conjunction stimuli conditions. The gray vertical rectangle indicates the time when the memory array was displayed. The gray lines represent the decodability of the trained condition. The colored lines represent the decodability of the tested conditions. The colored squares represent time points where the hyperplane contrast values are significantly greater than zero (corrected $p < .05$, FDR corrected using the Benjamini–Hochberg procedure). The black squares indicate when the within-feature decodability was significantly different from across-feature decodability.

first stimulus display and lasted until the first half of the delay period. For Set Size 3, decodability was significant starting moments after the first stimulus display and lasted until the end of the delay period.

We then assessed whether we could still decode the load for conjunction conditions (i.e., colored words) when sequentially presented. We found a strong within-feature load decodability, replicating the previous findings in Experiment 1. For across-feature decoding (Figure 6B), we again observed sustained generalization of WM load signals between conditions when trained on attend-word conditions, indicating the presence of a shared content-independent load signal. However, we found imperfect generalization during the late delay period (start: window centered on 1816 msec) when trained on attend-color conditions (Figure 6C). The color-based contrast was significantly higher than the word-based contrast, suggesting that the model was able to leverage on color-specific load signals that are not equivalently available in the attend-word condition.

Once again, to evaluate whether these effects could be explained by effort-related processes, we also decoded WM load from pupil size (Supplemental Figure S2). Although we observed sustained generalization of load decoding, the temporal profile differed from voltage-based load dynamics, with an onset that was 350 msec later for pupil than for voltage-based decoding. Notably, when trained on attend-color conditions and tested on attend-word conditions, pupil-based load decoding exhibited perfect generalization across the entire delay period unlike voltage-based decoding results.

Single-feature stimuli. We were able to reliably decode the attended feature information for single-feature conditions (Figure 6D) as well as for each set size condition (Supplemental Figure S1). Notably, we saw similar decoding patterns to Rajsic and colleagues' (2019) decoding results (Figure 2B), where we see an initial decodability peak after the stimulus onset. We both found strong within-feature and across-feature load decoding, with imperfect generalization for several time points. In general, we found more evidence of imperfect generalization across time when trained on color-only conditions compared with word-only conditions, but nonetheless, the results suggested the presence of generalizable load signals for both color and word features. Finally, pupil-based load decoding showed weak performance (Supplemental Figure S2). Training on color-only conditions yielded above-chance performance that did not generalize to word-only conditions, whereas training on word-only conditions produced no significant decoding at any time point and no evidence of across-feature generalization. Thus, this well-accepted objective measure of cognitive effort did not appear to provide robust tracking of load in this experiment. In turn, this reduces the concern that varying effort was the driver of the observed content-independent load signal.

Cross-decoding across stimulus types. Prior analyses focused on decoding performance within and across feature types under the same stimulus type (i.e., single-feature or conjunction stimuli). Here, we asked whether we see generalizable neural representations across single-featured and multifeatured stimuli. Specifically, we tested whether encoding a color (or word) in isolation yields the same neural pattern as when the same feature was selectively encoded from a dual-feature stimulus, and vice versa. To address this question, we performed a cross-decoding analysis in which classifiers trained on single-feature trials (e.g., color only or word only) were tested on conjunction trials (e.g., attend color or attend word), and vice versa.

First, we found robust and sustained generalization across single-featured and multifeatured stimuli within the same feature dimension (e.g., trained on color only, tested on attend color), and this was true across all training–testing combinations (see Supplemental Figure S3 for full results). We also found strong cross-feature generalization, specifically when the model was trained on word feature conditions (either word only or attend word) and tested on the corresponding color feature condition of the opposite stimulus type (Figure 6G). In contrast, when trained on color feature conditions, we observed imperfect generalization to word feature conditions during the latter half of the delay period (Figure 6C and H). These results suggest that the model captured color-specific load signal during training, which may have constrained its ability to generalize to word feature conditions.

RSA

Multiple signals were present in Experiment 2 (Figure 7). We found strong evidence for content-independent load signal (Figure 7B), which emerged shortly after the onset of the first stimulus presentation (window centered on 136 msec) and persisted throughout the delay period (window centered on 2104 msec). We found evidence for color-specific and word-specific load signals that tracked the load within their respective modalities across time points. However, these feature-specific load models were highly correlated with the attended feature model, as reflected by a variance inflation factor of 2.84 (roughly two thirds of the variance explained by other RDMs). As a result, the unique contribution of the attended feature signal was not reliably detected in the full model. When the color- and word-specific load models were removed (Figure 7C), we observed a reemergence of the attended feature signal.

Discussion

Experiment 2 provided a more comprehensive demonstration of content-independent load signals across both stimulus and feature types. By presenting memory items

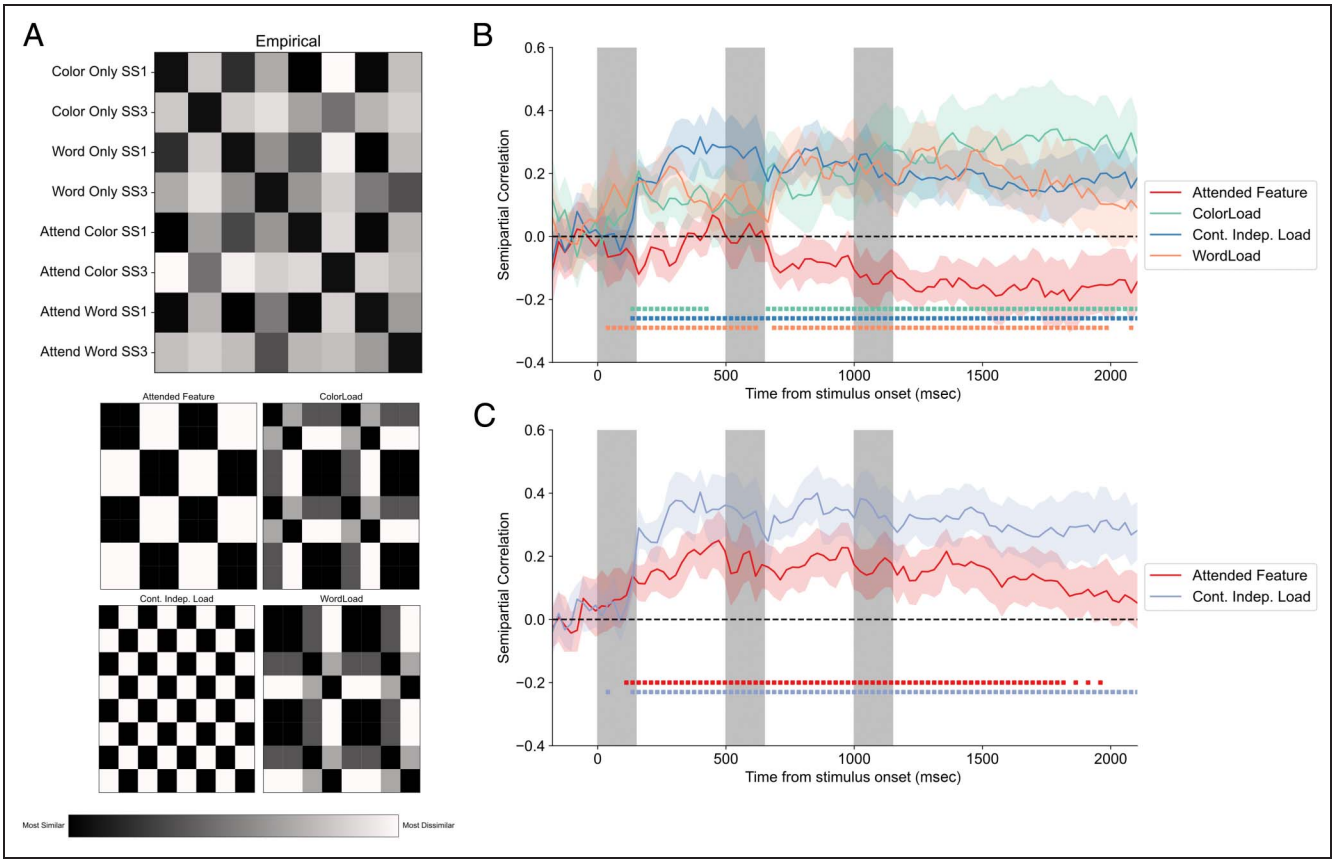

Figure 7. Experiment 2 RSA results. (A) Empirical and four theoretical RDMs, organized by attended feature (color, word), set size (1, 3), and stimulus type (single-feature, multifeature). (B) Semipartial correlations between empirical and theoretical RDMs estimated across time points, averaged across participants. (C) Semipartial correlations between empirical and theoretical RDMs without content-specific load models. The gray vertical rectangle indicates the time when the memory array was displayed. The colored lines represent time points where semipartial correlations are significantly greater than zero (one-tailed Wilcoxon signed-rank test with FDR correction using the Benjamini–Hochberg procedure). Cont. Indep. = content independent.

centrally in a sequential manner, we ruled out spatial attention as a source of the content-independent load signal. Multivariate decoding and RSA again revealed a robust, sustained load signal that generalized across color and word features, in line with a domain-general component of WM storage.

We also observed reliable decoding of the attended feature for both single-feature and conjunction conditions, a notable contrast to Experiment 1, where only higher set size conditions showed sustained attended feature decodability. Moreover, cross-decoding analyses revealed near-perfect generalization when models were trained and tested on the same feature across different stimulus types, with modest asymmetries when trained on color and tested on word features. This suggests that although the load signal is broadly generalizable, feature-specific signals, especially for words, may contribute to representational differences across stimulus types, perhaps indicating a transformation in WM maintenance format (e.g., concurrent storage to verbal coding of relevant features). For pupil-based decoding, we replicated the prior findings in Experiment 1, again observing a dissociation between the temporal dynamics of pupil- and voltage-based WM load decoding. Although pupil size carried above-chance load information in some conditions, these effects showed different generalization patterns compared with voltage-based results. This consistent mismatch across experiments reinforces the interpretation that the content-independent load signal observed in EEG data reflects neural processes that are distinct from effort-related signals indexed by pupil size.

RSA results further corroborated these findings by revealing both content-independent and feature-specific load signals for color and word features. In the full RSA model, we did not observe a reliable unique contribution of the attended feature model, which may reflect collinearity with the feature-specific load models—as indicated by a variance inflation factor of 2.84. When these collinear models were removed, the attended feature model accounted for unique variance in EEG signals. Although this result demonstrates that variance associated with the attended feature model is present in the data, it remains ambiguous whether this reflects a distinct task set like signal or overlap with feature-specific load patterns. Overall, these findings reinforce the existence of a content-independent load signal while also highlighting

the contribution of feature-specific representations that may differ in strength and temporal dynamics across modalities.

GENERAL DISCUSSION

Across three data sets, we found consistent evidence for a modality-general neural signal that tracks the number of items stored in WM, regardless of whether those items were simple colors or meaningful words. This content-independent load signal was robust to differences in their perceptual format (Rajsic et al., 2019) and remained reliable when controlling for stimulus-evoked activity (Experiment 1) and spatial attention (Experiment 2). These results build on prior findings that have implicated a common indexing operation for highly distinct visual features (Jones, Diaz, et al., 2024; Thyer et al., 2022) and even distinct sensory modalities (Suplica et al., 2025). Despite clear evidence for nonoverlapping aspects of visual and verbal storage in WM (e.g., Baddeley, 2000; Baddeley & Hitch, 1974), our findings reveal a common component of WM storage across visual and verbal items. At the same time, distinct variance in neural activity was explained by the specific features that were stored, underscoring that this generalizable load signal operates in parallel with content-specific representations. Taken together, these findings argue against a strictly modular view of WM and instead support a model in which visual and verbal information are maintained via both shared and separable mechanisms.

A potential alternative interpretation is that the decodable differences between WM load conditions reflect task-general variations in cognitive effort, which has been argued to covary with set size (Shenhav et al., 2017; D'Esposito & Postle, 2015). Although effort-related factors may explain some portion of the variance in the EEG activity, several aspects of our results suggest that effort does not provide a complete account of the observed evidence for content-independent load signals. For example, we measured pupil dilation, a well-accepted measure of cognitive effort when stimulus displays are matched (van der Wel & van Steenbergen, 2018). Although pupil-based decoding enabled above-chance load classification in some conditions, pupil decoding revealed markedly different patterns of generalization than did EEG-based markers of WM load. Whereas EEG measures of WM load generalized across stimulus types in all cases, pupil signals did not exhibit such consistent generalization, undermining the claim that effort alone explains the near-perfect generalization of EEG-based load signatures across visual and verbal stimuli. Especially in Experiment 2, pupil-based decoding failed when trained on word-only conditions and failed to generalize to color-only conditions, whereas EEG load signals remained robust (Supplemental Figure S2D). Furthermore, conjunction conditions showed strong generalization with single-feature conditions despite greater selective attention demands, which would require more cognitive effort. Finally, whereas behavioral accuracy varied across stimulus types in Experiment 2, indicating differences in task difficulty, the neural load patterns remained consistent across stimuli.

These distinctions between neural and pupil-based measures of WM load dovetail with past work that has also shown that sustained content-independent load signals can be dissociated from multiple effort-related measures (Suplica et al., 2025; Jones, Thyer, et al., 2024). For example, prior work has demonstrated that these multivariate load signals are insensitive to the total amount of feature information associated with each item, such that load signals track the total number of items or chunks maintained in WM, rather than the total number of features that observers must store (Suplica et al., 2025; Thyer et al., 2022). Consistent with this dissociation, ongoing work in our laboratory suggests that participants subjectively experience multifeature memoranda as more effortful, although we observe near-perfect generalization between single-feature and multifeature load signatures. To sum up, although we acknowledge that further work is needed to examine the possible links between cognitive effort and neural signatures of WM load, the present work combined with past studies suggests that a content-independent signature of WM storage provides a more compelling explanation of our findings than cognitive effort.

Although domain-specific subsystems, such as the visuospatial sketchpad and phonological loop, have traditionally been emphasized in the literature, our findings are also consistent with a less frequently highlighted component of this framework: the episodic buffer (Baddeley, 2000). This component was proposed to integrate information across modalities by binding content into coherent, capacity-limited chunks that can interface with long-term memory. Rather than indicating that visual and verbal WM operate in isolation, our results support a view in which modality-specific representational codes are coordinated by a shared, capacity-limited architecture that flexibly supports storage across domains. This architecture converges with contemporary accounts proposing content-independent "pointer" mechanisms that support contextual binding in WM (Awh & Vogel, 2025; Jones, Thyer, et al., 2024). Importantly, this interpretation also aligns closely with Cowan's embedded process model (Cowan, 2001), which posits a central focus of attention capable of maintaining a limited number of chunks regardless of modality. Crucially, our RSAs show that even after accounting for this shared, load-related variance, reliable modality-specific representational structure remains. Thus, the present findings suggest that the perspectives advanced by Baddeley and Cowan are not competing explanations but instead capture complementary levels of WM organization: a domain-general capacity limit operating over modality-specific representational systems.

Multivariate decoding results from Experiment 2 also revealed instances of imperfect generalization between color and word features, particularly during the latter half of the delay period. These results suggest that, whereas a

domain-general architecture supports core WM operations such as item individuation and indexing (Jones, Thyer, et al., 2024; Thyer et al., 2022), feature-specific codes might be evolving over time. One possibility is that the neural representation of verbal content undergoes a transformation during maintenance, increasingly relying on phonological or semantic codes (Camos, Lagner, & Loaiza, 2017). This divergence could account for the reduced generalization, especially at later time points when these representational shifts are most pronounced, as well as higher accuracies for Set Size 3 word conditions. However, these asymmetries emerged against a backdrop of sustained, generalizable load signals across conditions. That is, even when feature-specific codes emerged, the shared neural signature reflecting the number of items in WM remained robust, as indicated by the decoding results when trained on word and tested on color features. This suggests that the domain-general indexing of stored items remained stable, even as the nature of content-specific representations evolved over time.

Notably, RSA also revealed the presence of both color- and word-specific load signals, consistent with the multivariate decoding results above. However, even after controlling for these content-specific components, a robust and sustained signal reflecting content-independent WM load persisted throughout the delay period. Unlike multivariate decoding approaches, which optimize classification performance by leveraging all available discriminative patterns (Hebart & Baker, 2018), RSA provides a theory-driven framework for isolating distinct representational components (Kriegeskorte et al., 2008). By modeling the structure of neural dissimilarities relative to explicit hypotheses, RSA enables us to quantify the unique variance explained by each theoretical model. In doing so, it disentangles overlapping sources of information and allows both content-specific and content-independent signals to be examined in parallel. Thus, while controlling for a range of factors that have typically been confounded with WM load in past work (e.g., sensory energy, spatial attention, feature-specific activity), we have provided compelling evidence for a content-independent load signal that operates across visual and verbal modalities. These results build on a growing body of work, emphasizing the distinction between content-specific representations and operations that bind those representations to the surrounding context (Awh & Vogel, 2025).

Corresponding author: Woohyeuk Chang, Department of Psychology, University of Chicago, 5848 S. University Avenue, Chicago, IL 60637, e-mail: woohyeukchang@uchicago.edu.

Data Availability Statement

All data, analysis scripts, and task scripts are publicly available via OSF and can be accessed at https://osf.io/dvtqr/. Supplemental Material can be accessed on this article's homepage: https://doi.org/10.1162/JOCN.a.2522.

Author Contributions

Woohyeuk Chang: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Writing—Original draft; Writing—Review & editing. Will Epstein: Conceptualization; Data curation; Investigation; Writing—Review & editing. Henry Jones: Methodology; Resources; Writing—Review & editing. William X. Q. Ngiam: Conceptualization; Writing—Review & editing. Edward Awh: Conceptualization; Funding acquisition; Writing—Original draft; Writing—Review & editing.

Funding Information

This research was supported by National Institute of Mental Health (https://dx.doi.org/10.13039/100000025), grant no. ROIMH087214 and Office of Naval Research (https://dx.doi.org/10.13039/100000006), grant no. N00014-12-1-0972 to E. A.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

Adam, K. C. S., Vogel, E. K., & Awh, E. (2020). Multivariate analysis reveals a generalizable human electrophysiological signature of working memory load. *Psychophysiology*, *57*, e13691. https://doi.org/10.1111/psyp.13691, PubMed: 33040349

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*, 119–126. https://doi.org/10.1016/s1364-6613(00)01593-x, PubMed: 11239812

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science*, *7*, 25–31. https://doi.org/10.1111/j.1467-9280.1996.tb00662.x

Awh, E., & Vogel, E. K. (2025). Working memory needs pointers. *Trends in Cognitive Sciences*, *29*, 230–241. https://doi.org/10.1016/j.tics.2024.12.006, PubMed: 39779443

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2, PubMed: 11058819

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Elsevier. https://doi.org/10.1016/S0079-7421(08)60452-1

Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2023). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. *Psychonomic Bulletin & Review*, *30*, 421–449. https://doi.org/10.3758/s13423-022-02179-w, PubMed: 36260270

Camos, V., Lagner, P., & Loaiza, V. M. (2017). Maintenance of item and order information in verbal working memory. *Memory*, *25*, 953–968. https://doi.org/10.1080/09658211.2016.1237654, PubMed: 27745428

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia*, *49*, 1407–1409. https://doi.org/10.1016/j.neuropsychologia.2011.01.029, PubMed: 21295047

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114. https://doi.org/10.1017/s0140525x01003922, PubMed: 11515286

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142. https://doi.org/10.1146/annurev-psych-010814-015031, PubMed: 25251486

Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., & Awh, E. (2017). Alpha-band oscillations enable spatially and temporally resolved tracking of covert spatial attention. *Psychological Science*, *28*, 929–941. https://doi.org/10.1177/0956797617699167, PubMed: 28537480

Fougnie, D., Zughni, S., Godwin, D., & Marois, R. (2015). Working memory storage is intrinsically domain specific. *Journal of Experimental Psychology: General*, *144*, 30–47. https://doi.org/10.1037/a0038211, PubMed: 25384164

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*, 14009–14016. https://doi.org/10.1523/JNEUROSCI.5003-14.2015, PubMed: 26468201

Hakim, N., Feldmann-Wüstefeld, T., Awh, E., & Vogel, E. K. (2021). Controlling the flow of distracting information in working memory. *Cerebral Cortex*, *31*, 3323–3337. https://doi.org/10.1093/cercor/bhab013, PubMed: 33675357

Hebart, M. N., & Baker, C. I. (2018). Deconstructing multivariate decoding for the study of brain function. *Neuroimage*, *180*, 4–18. https://doi.org/10.1016/j.neuroimage.2017.08.005, PubMed: 28782682

Ikkai, A., McCollough, A. W., & Vogel, E. K. (2010). Contralateral delay activity provides a neural measure of the number of representations in visual working memory. *Journal of Neurophysiology*, *103*, 1963–1968. https://doi.org/10.1152/jn.00978.2009, PubMed: 20147415

Jones, H. M., Diaz, G. K., Ngiam, W. X. Q., & Awh, E. (2024). Electroencephalogram decoding reveals distinct processes for directing spatial attention and encoding into working memory. *Psychological Science*, *35*, 1108–1138. https://doi.org/10.1177/09567976241263002, PubMed: 39159181

Jones, H. M., Thyer, W. S., Suplica, D., & Awh, E. (2024). Cortically disparate visual features evoke content-independent load signals during storage in working memory. *Journal of Neuroscience*, *44*, e0448242024. https://doi.org/10.1523/JNEUROSCI.0448-24.2024, PubMed: 39327004

Kikumoto, A., & Mayr, U. (2020). Conjunctive representations that integrate stimuli, responses, and rules are critical for action selection. *Proceedings of the National Academy of Sciences, U.S.A.*, *117*, 10603–10608. https://doi.org/10.1073/pnas.1922166117, PubMed: 32341161

Kikumoto, A., Sameshima, T., & Mayr, U. (2022). The role of conjunctive representations in stopping actions. *Psychological Science*, *33*, 325–338. https://doi.org/10.1177/09567976211034505, PubMed: 35108148

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008, PubMed: 19104670

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica*, *75*, 55–74. https://doi.org/10.1016/0001-6918(90)90066-o, PubMed: 2260493

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. https://doi.org/10.1016/j.neubiorev.2016.01.003, PubMed: 26802451

McCollough, A. W., Machizawa, M. G., & Vogel, E. K. (2007). Electrophysiological measures of maintaining representations in visual working memory. *Cortex*, *43*, 77–94. https://doi.org/10.1016/s0010-9452(08)70447-7, PubMed: 17334209

Ngiam, W. X. Q., Adam, K. C. S., Quirk, C., Vogel, E. K., & Awh, E. (2021). Estimating the statistical power to detect set-size effects in contralateral delay activity. *Psychophysiology*, *58*, e13791. https://doi.org/10.1111/psyp.13791, PubMed: 33569785

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, *362*, 342–345. https://doi.org/10.1038/362342a0, PubMed: 8455719

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Rajsic, J., Burton, J. A., & Woodman, G. F. (2019). Contralateral delay activity tracks the storage of visually presented letters and words. *Psychophysiology*, *56*, e13282. https://doi.org/10.1111/psyp.13282

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., et al. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *Neuroimage*, *60*, 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050, PubMed: 22178808

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., et al. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*, 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526, PubMed: 28375769

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662. https://doi.org/10.1037/h0054651

Suplica, D., Jones, H. M., Diaz, G. K., Veillette, J. P., Nusbaum, H. C., & Awh, E. (2025). Neural evidence for modality-independent storage in working memory. *Current Biology*, *35*, 4620–4630. https://doi.org/10.1016/j.cub.2025.08.007, PubMed: 40912253

Thyer, W., Adam, K. C. S., Diaz, G. K., Velázquez Sánchez, I. N., Vogel, E. K., & Awh, E. (2022). Storage in visual working memory recruits a content-independent pointer system. *Psychological Science*, *33*, 1680–1694. https://doi.org/10.1177/09567976221090923, PubMed: 36006809

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*, 751–754. https://doi.org/10.1038/nature02466, PubMed: 15085133

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review.

Psychonomic Bulletin & Review, *25*, 2005–2015. https://doi.org/10.3758/s13423-018-1432-y, PubMed: 29435963

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751. https://doi.org/10.1038/nature02447, PubMed: 15085132

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage*, *137*, 188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012, PubMed: 26707889

Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review*, *15*, 223–229. https://doi.org/10.3758/pbr.15.1.223, PubMed: 18605507

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91–95. https://doi.org/10.1038/nature04262, PubMed: 16382240

Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences*, *13*, 167–174. https://doi.org/10.1016/j.tics.2009.01.008, PubMed: 19269882