Article

Current Biology

Neural evidence for modality-independent storage in working memory

Highlights

- EEG measures of working memory load generalize across auditory and visual items
- Multi-modal storage occurs in an item-based fashion
- This storage signal is independent of spatial attention and cognitive effort
- Abstract indexing operation may underlie a spatiotemporal "pointer" process

Authors

Darius Suplica, Henry M. Jones, Gisella K. Diaz, John P. Veillette, Howard C. Nusbaum, Edward Awh

Correspondence

dsuplica@uchicago.edu

In brief

Suplica et al. provide human EEG evidence for a modality-independent, item-based signature of working memory storage. The study provides further evidence for an abstract indexing operation underlying WM, potentially related to the binding of representations to their event context.

Suplica et al., 2025, Current Biology *35*, 1–11
October 6, 2025
https://doi.org/10.1016/j.cub.2025.08.007

CellPress

Current Biology

CellPress

Article

Neural evidence for modality-independent storage in working memory

Darius Suplica,[1,3,*] Henry M. Jones,[1,2] Gisella K. Diaz,[1,2] John P. Veillette,[1] Howard C. Nusbaum,[1] and Edward Awh[1,2]
[1]Department of Psychology, The University of Chicago, Chicago, IL 60637, USA
[2]Institute for Mind and Biology, The University of Chicago, Chicago, IL 60637, USA
[3]Lead contact
*Correspondence: dsuplica@uchicago.edu
https://doi.org/10.1016/j.cub.2025.08.007

SUMMARY

Working memory (WM) is a core component of intellectual ability. Traditional behavioral accounts have argued that there remain distinct memory systems based on the type and sensory modality of information being stored. However, more recent work has provided evidence for a class of neural activity that indexes the number of visual items stored in a content-independent fashion. Here, across 2 electroencephalogram (EEG) experiments, we demonstrate an item-based signature of WM storage that generalizes across visual and auditory sensory modalities. Using multivariate techniques, we observed parallel but separate neural patterns that independently track stimulus modality, the number of items stored in WM (regardless of modality), and the number of spatially attended positions. We propose that these load signals reflect a modality-independent process for binding item representations to context, reinforcing accounts arguing for a distinction between the maintenance of the content of one's thoughts and the manipulation and gating of those thoughts.

INTRODUCTION

Neural studies of working memory (WM) have made major progress by examining stimulus-specific activity that tracks the content of stored items.[1–5] These neural signals track the voluntarily stored aspects of relevant items,[5,6] predict mnemonic fidelity,[7–9] and provide insight into the dynamics of selection and access to stored content.[10–12] Nevertheless, recent work has highlighted evidence for a distinct class of neural activity that indexes the *number* of items encoded into WM, independent of the specific content associated with those items.[13–18] For instance, multivariate analyses of electroencephalogram (EEG) activity reveal a common signature of the number of items stored in WM, despite variations in both the *type* (e.g., color, orientation, and motion) and the *number* of feature values stored for each item.[16,17] Moreover, when perceptual grouping encourages multiple elements to be perceived as a unit, this neural load signal tracks the number of *perceived items*, not the number of attended elements or positions.[16,19] Thus, these findings reveal an *item-based* neural signature of WM storage that generalizes across highly distinct visual features.

Our working hypothesis is that these content-independent signatures of WM load are generated by a spatiotemporal "pointer" operation that binds item representations to the surrounding event context.[20] This kind of contextual binding has been highlighted both in major models of WM[21–25] as well as in longstanding theories of dynamic visual cognition.[26,27] Moreover, the latter theories embrace a clear separation between the indexing operation supported by pointers, on the one hand, and parallel operations that maintain the featural content of the selected items, on the other. Here, there is a useful analogy between pointers and demonstrative words such as "this" and "that": demonstratives hold no meaning by themselves, but their function is to refer to other meaningful words. Likewise, pointers may support the contextual binding of items, while parallel operations maintain the contents of the indexed representations. Thus, pointers provide an attractive explanation for the presence of content-independent signatures of WM storage.

Extant demonstrations of content-independent load signals, however, have been restricted to the visual modality, leaving open the possibility that distinct sensory modalities recruit distinct pointer systems. Indeed, prominent WM models have argued that there are independent resource pools for the storage of visual and auditory/phonological information,[28–30] and multiple studies have found minimal interference between auditory and visual loads competing for WM storage.[31,32] Moreover, past EEG studies have highlighted distinct electrode sites where activity tracks auditory and visual WM loads, with anterior electrode sites tracking auditory loads[33] while posterior parietal electrodes track the number of visual items stored.[13,34] Thus, both behavioral and neural studies have pointed toward dissociable WM systems for visual and auditory information. In the present work, we replicate the observation that ongoing EEG activity is shaped by the sensory modality that is stored, but we also present clear evidence for a modality-general component of WM storage.

Pilot work revealed that we could decode the number of discrete sounds being held in WM by applying multivariate classifiers to ongoing EEG activity, similar to recent work with visual stimuli.[16–18,35] Here, we manipulated the number and sensory

Please cite this article in press as: Suplica et al., Neural evidence for modality-independent storage in working memory, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.08.007

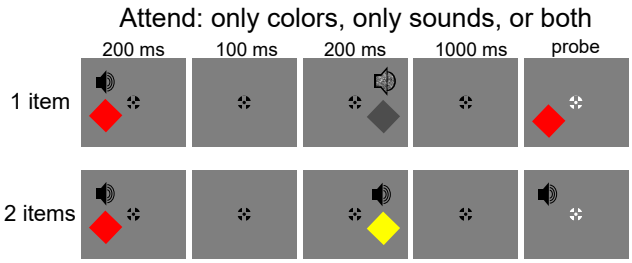

Figure 1. Experiment 1 task

Participants attended to colors, sounds, or both features. Stimuli were color-sound pairs. In set size 1 trials, one pair was replaced with white noise and a gray diamond.

modality of items stored in WM, using audiovisual displays that controlled sensory energy across experimental conditions. Experiment 1 revealed robust generalization of EEG load signatures across auditory and visual features, such that training on one sensory modality enabled precise decoding of load in the other sensory modality. Moreover, the load signature for single-feature objects (colors or sounds alone) generalized to dual-feature audiovisual objects, showing that this analysis provides an *item-based* rather than a feature-based measure of WM storage. Representational similarity analysis provided clear evidence for a sustained neural signal that tracked the number of items stored in WM, regardless of the sensory modality or informational complexity of those memories. Critically, this signal explained distinct variance in ongoing EEG activity from another robust signal that tracked the stored sensory modality. Experiment 2 replicated these key findings while independently manipulating the number of attended positions and the number of items stored in those positions. We saw clear evidence for a neural signal tracking the number of attended positions, but this explained distinct variance in EEG activity from the modality-general signal that tracked the number of stored items. Thus, our findings reinforce recent work arguing for a functional dissociation between the voluntary control of spatial attention and the encoding of items into WM,[18,36] and they provide direct neural evidence for a modality-general component of this online memory system.

RESULTS

28 healthy human participants performed a WM task in which both the number of items to be stored and the sensory modality of the stored information were manipulated (Figure 1). Our task employed sequentially presented audiovisual stimuli that consisted of a colored diamond and a real-world sound that onset simultaneously in one of three positions (left, middle, or right). To match sensory energy across set sizes, all sample displays included two audiovisual presentations. For the set size 2 trials, all audiovisual stimuli were comprised of a saturated color and a real-world sound. For set size 1 trials, 1 of those audiovisual stimuli was randomly selected and replaced with energy-matched "placeholders." Participants were instructed to store the targets and ignore the placeholders in preparation for a change detection probe. Critically, in different blocks, participants were instructed to store either the color, the sound, or the conjunction of color and sound for each audiovisual target. Following a 1,000 ms delay period, a single probe stimulus was presented, and subjects reported whether it was the same or different from the stimulus presented at that location. Test probes were in the modality that matched the storage instructions. When both visual and auditory features were stored, visual and auditory probes were randomly intermixed and equally likely.

We had behavioral data for 22 out of 24 subjects (behavioral data for subjects 1 and 2 were missing due to a programming error). Mean accuracy across subjects was high for all conditions (range 92.3% for auditory set size 2%–97.7% for visual set size 1). A repeated-measures ANOVA revealed a significant effect of attended modality ($F_{(2,42)} = 8.95$, $p = 5.8 \times 10^{-4}$), WM load ($F_{(1,21)} = 18.43$, $p = 3.2 \times 10^{-4}$), and a non-significant interaction term ($F_{(2,42)} = 2.52$, $p = 0.092$).

Neural signatures of WM load generalize across sensory modalities

Consistent with past work, multivariate analysis of EEG activity enabled robust and sustained decoding of the number of items stored in WM for all 3 conditions (visual, auditory, and conjunction). To test whether load decoding generalized across these conditions, we measured classifier performance when training and testing across different modalities (Figures 2A and 2B). Thus, we trained classifiers to discriminate between set sizes 1 and 2 within each modality, and we tested those classifiers on independent data in which either the same or a different modality was stored. Instead of using raw classification accuracy to measure decodability, we employed a metric called "hyperplane contrast."[17] Hyperplane contrast is conceptually similar to decoding accuracy, with higher numbers indicating better decoding. Hyperplane contrast has the advantage of being more robust to non-orthogonal but potentially confounding neural signals. Moreover, this signal grows proportionally to the signal-to-noise ratio, improving interpretability (see STAR Methods for additional rationale).

Overall, we observed robust decoding of WM load for nearly all time points within modalities (51/58 auditory, 52/58 visual) and for most, but not all, time points across modalities (45/58 auditory to visual, 40/58 visual to auditory). When averaging over the delay period, hyperplane contrast (a measure of classifier performance) over the delay period was significantly greater than zero both when training and testing within a modality (auditory, $t(23) = 3.91$, $p = 3.48 \times 10^{-4}$, Bayes factor [BF] = 96.4; visual, $t(23) = 3.74$, $p = 5.33 \times 10^{-4}$, BF = 66.4) and when training and testing across modalities (auditory to visual, $t(23) = 3.18$, $p = 2.10 \times 10^{-3}$, BF = 20.2; visual to auditory, $t(23) = 3.01$, $p = 3.16 \times 10^{-3}$, BF = 14.288). Hyperplane contrast when testing across modalities (crossmodal) was not significantly lower than within modalities (intramodal) at any time point (one-tailed t test, false discovery rate [FDR] corrected). In the aggregate, BFs for the difference between intramodal and crossmodal contrast were 0.77 (train auditory, $t(23) = 1.143$, $p = 0.132$) and 1.0 (train visual, $t(23) = 1.388$, $p = 0.089$), indicating ambiguous evidence for either the null or alternative hypotheses. While the ambiguous bayes factor does not permit us to rule out some domain-specificity (in fact, we show later that attended modality substantially impacts the load signal), the presence of robust

Please cite this article in press as: Suplica et al., Neural evidence for modality-independent storage in working memory, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.08.007

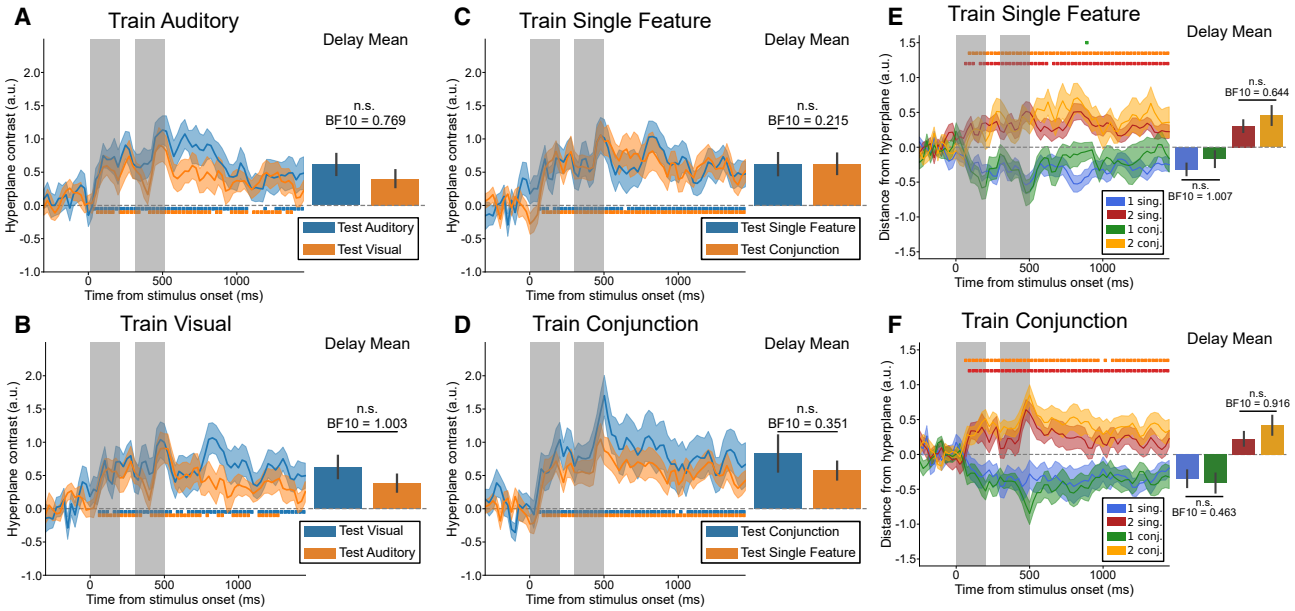

Figure 2. Decoding results for experiment 1 (n = 24)

Stimuli were presented from 0 to 200 ms and 300 to 500 ms (shown in gray). Bar graphs are the average value (hyperplane contrast for A–D, hyperplane score for E and F) over the delay period (500–1,500 ms). Lines denote mean value, and the shaded region = SEM. Squares indicate hyperplane contrast values (A)–(D) greater than zero (FDR-corrected one-tailed t test). BFs for each delay period comparison (A and B: H_A = intramodal > crossmodal; C and D: H_A = single feature ≠ conjunction; E and F: conjunction > single feature) are also given.

(A and B) Hyperplane contrast within and across sensory modalities. Blue: same modality as training; orange: opposite modality.

(C and D) Hyperplane contrast for single-feature vs. conjunction (dual-feature) trials. Blue: same number of features as training, orange: different number of features.

(E and F) Hyperplane distances for single-feature vs. conjunction trials. More negative values are higher confidence set size 1, and more positive values are higher confidence set size 2. Yellow squares indicate 2 conjunctions > 1 conjunction. Red indicates 2 singles > 1 single. Green indicates 1 conjunction > 1 single (only present at one time point).

See also Figure S4 for an analysis of the effects of trial bin size on generalization.

cross-decoding across modalities provides strong evidence for a common neural signature of load across visual and auditory modalities.

Neural signatures of WM load are item-based, not feature-based

To test whether these load classifiers were operating at the level of items or features, we examined whether load classifiers based on single-feature stimuli generalized to conjunction conditions in which both visual and auditory features were retained. Cross-decoding between single-feature and conjunction conditions was robust and nearly identical (Figures 2C and 2D), consistent with prior findings with conjunctions of color and orientation.[16] Hyperplane contrasts for conjunction and single-feature trials were not significantly different at any point (two-tailed t test, FDR corrected; delay average, train single-feature: $t(23) = -0.029$, $p = 0.976$, BF = 0.215; train conjunction: $t(23) = 1.049$, $p = 0.305$, BF = 0.351). While successful cross-decoding provides evidence for a common load signature, we also used a more specific analysis to distinguish whether load decoding was item-based or feature based. We trained a model to distinguish between set sizes 1 and 2 using only single-feature stimuli and then tested this model with a single conjunction stimulus. If load decoding is based on the number of feature values stored, then models trained on single-feature stimuli should be biased toward higher load classification when the number of features stored per item is doubled. This prediction was not supported. Even though twice as many features were stored in the conjunction condition, a single conjunction stimulus did not show an upward shift from the single-feature set size 1 condition in a hyperplane analysis (Figure 2E), with the exception of a single time point. No significant increase in load was observed in the aggregate (set size 1: $t(23) = 1.391$, $p = 0.089$, BF = 1.007; set size 2: $t(23) = 0.949$, $p = 0.176$, BF = 0.644). The same pattern when training the model on conjunction stimuli (Figure 2F, set size 1: $t(23) = -0.407$, $p = 0.656$, BF = 0.463; set size 2: $t(23) = 1.309$, $p = 0.102$, BF = 0.916). Thus, this modality-general load signal tracks the number of *items* stored in WM, not the number of feature values.

Generalization across modalities is not due to storage of irrelevant features

An alternative explanation for the generalization of load models across the auditory and visual conditions is that observers tended to store both auditory and visual features, regardless of instructions. In this case, perfect cross training would be expected even if independent neural signals tracked each modality. Two pieces of evidence contradict this hypothesis. First,

Please cite this article in press as: Suplica et al., Neural evidence for modality-independent storage in working memory, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.08.007

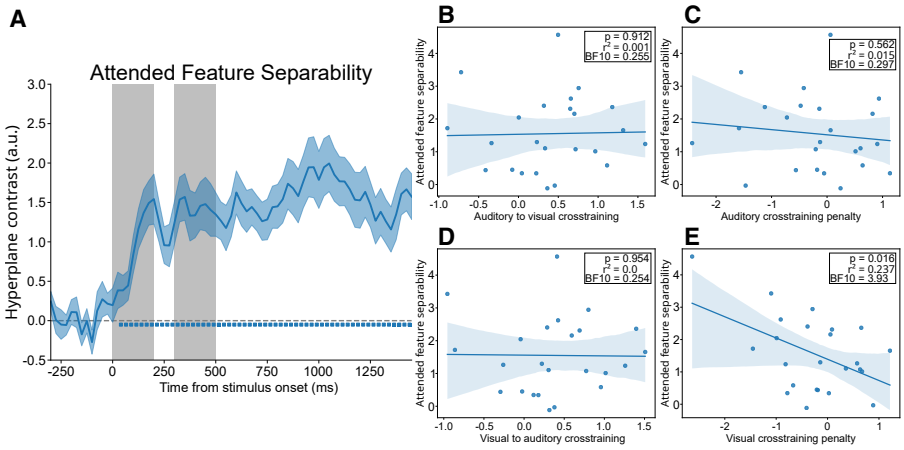

Figure 3. Leakage does not explain generalization (n = 24)

(A) Hyperplane contrast between attend auditory and attend visual trials (set sizes combined). Significant time points (FDR corrected $p < 0.05$) are denoted by blue squares.

(B and D) Correlation between attended feature discriminability and cross-decoding (B: auditory to visual, D: visual to auditory).

(C and E) Correlation between attended feature discriminability and crosstraining penalty equivalent to within modality—cross modality hyperplane contrast, (C) auditory—auditory to visual, (E) visual—visual to auditory. All values are averaged across the delay period. Correlations were evaluated by linear regression (Wald test) across subjects (line = line of best fit, shaded region = 95% CI).

attended modality (auditory/visual) was robustly discriminated by our classifier (Figure 3A) to an even greater degree than load, and this feature-specific activity was sustained across the delay period. This provides direct neural evidence that observers selectively stored the relevant features. Second, we reasoned that if cross training was driven by storage of the irrelevant feature in WM, then individuals with higher decodability of the attended feature should show worse generalization between auditory and visual load models. We evaluated this in two ways. First, we correlated attended feature discriminability with cross-modal decoding performance (Figures 3B and 3D) to test the prediction that subjects worse at filtering would show better generalization across modalities. Second, we replaced cross-modal discriminability with a "penalty" metric (Figures 3C and 3E), equal to the difference between the mean intramodal and crossmodal contrast. This hypothesis predicts that poor filterers (low attended feature discriminability) should have a smaller penalty because they were in more similar attentional states during the auditory and visual conditions. However, three of these correlations were not significant (Figure 3B p = 0.912, BF = 0.255; Figure 3C p = 0.562, BF = 0.297; Figure 3D p = 0.954, BF = 0.254), while one was statistically significant in the opposite direction of the prediction (Figure 3E, p = 0.016, BF = 3.93). However, after excluding outlier values, this association was no longer significant (p = 0.595, BF = 0.302). Thus, storage of irrelevant features cannot explain our results.

Using Representational similarity analysis for a concurrent analysis of modality-general pointers and feature-specific activity

One limitation of our decoding approach so far is that each analysis targeted one construct of interest (i.e., modality-general load vs. feature-specific activity) while ignoring the presence of the others. Moreover, even good classifier generalization does not conclusively establish overlap in neural codes.[37] Thus, to obtain converging evidence regarding these distinct classes of neural activity, we used representational similarity analysis (RSA) to simultaneously model the effects of load and stimulus modality. The logic of RSA is to examine whether specific theoretical models predict the pairwise similarity across all conditions of the experiment. We tested four separate regressors that might predict this similarity structure (Figure 4C): (1) pointer load model—predicts similarity based on the number of *items* stored, regardless of the selected sensory modality. (2) Feature load model—predicts similarity based on the total number of *feature* values stored (regardless of modality). (3) Graded feature model—predicts similarity based on the degree of overlap in attended features, such that the conjunction condition falls in between the single-feature conditions. (4) Task-set model—predicts similarity based on which task set the observer has adopted (color, sound, or conjunction), with each task set being equally distinct from the others. Thus, regressors 1 and 2 contrasted item-based vs. feature-based models of WM load activity, while regressors 3 and 4 examined whether feature-selective signals tracked the specific constellation of features stored or the discrete task sets that corresponded to each experimental condition.

Item-based pointers and feature-specific activity explain unique variance in EEG activity

By calculating the semipartial correlations between each regressor and ongoing EEG activity, we measured the unique variance explained by each regressor at each time point (Figures 4D and 4E). Two regressors reliably explained variance in neural activity. First, the pointer load model explained robust variance throughout the entire delay period (Wilcoxon signed-rank test, W = 273, $p = 7.48 \times 10^{-5}$), while the feature load model explained no unique variance (W = 183, p = 0.180). This reinforces the prior conclusion

Please cite this article in press as: Suplica et al., Neural evidence for modality-independent storage in working memory, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.08.007

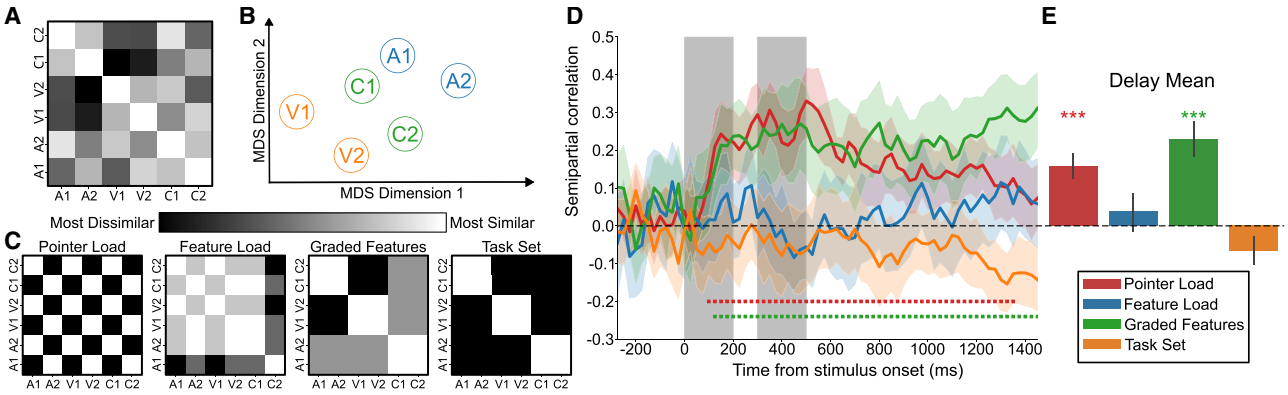

Figure 4. RSA results for experiment 1 (n = 24)

Condition labels are denoted by a letter to indicate modality condition (A: Auditory, V: Visual, C: Conjunction), as well as a number to denote set size (1 or 2).

(A) Empirical representational dissimilarity matrix (RDM), averaged across subjects over delay period (t > 500 ms).

(B) MDS projection of RDM, color coded by modality, averaged over the delay period.

(C) Graphical description of tested theoretical models. Darker = more dissimilar, lighter = more similar, arbitrary scale.

(D) Semipartial correlations of each factor to the empirical RDM over time. Stimulus period denoted by the gray-shaded region, significant time points (Wilcoxon signed-rank test, FDR corrected) denoted by colored boxes under the graph.

(E) Semipartial correlations averaged across the delay period (t > 500 ms).

***p < 0.001, **p < 0.01, *p < 0.05 (Wilcoxon signed-rank test, uncorrected).

See also Figure S2 for a replication with a pupil size regressor.

that these neural load signals are item-based, not feature based. Second, the graded feature model explained robust variance throughout the delay period (W = 278, $p = 3.19 \times 10^{-5}$), while the task-set model failed to explain unique variance (W = 92, p = 0.952). Thus, feature-selective activity in this study was better explained by activity within modality-specific sensory regions than by activity that tracked the three discrete task sets required by our instructions.[38] Critically, the pointer load and graded feature models explained *distinct variance* in EEG activity. Thus, RSA reinforced our initial evidence for modality-general pointers while simultaneously controlling for feature-specific neural activity. Finally, multidimensional scaling (MDS) allowed us to visualize the similarity relationships between all experimental conditions within a 2D space (Figure 4B). Here, two axes of separation are apparent. First, a "sensory modality" axis divides the conditions from left to right. Visual and auditory conditions are on the left and right, respectively, while the conjunction condition falls in between them, in line with the graded feature model. Second, another axis separates load 1 conditions (top) from load 2 conditions (bottom), regardless of modality. Thus, this separability aligns with our hypothesis and the semipartial correlations described above.

Experiment 2

Although the findings so far reveal a clear dissociation between load-sensitive neural signals and feature-specific neural activity, the number of items stored in WM was confounded with the number of relevant positions in the display. Could changes in the breadth of spatial attention explain the common load signature between visual and auditory modalities? Indeed, recent work has shown that RSA is sensitive to EEG signatures of spatial attention,[18] enabling a direct test of this question. Thus, we independently manipulated WM load, attended modality, and the number of spatially attended positions (Figure 5). Subjects attended to either auditory or visual features and stored one or three targets that were sequentially presented in either a single location or across three unique locations.

We had behavioral data for 15 out of 16 subjects. Mean accuracy across subjects ranged from a minimum of 89.3% (auditory set size 3, different locations) to a maximum of 98.4% (visual set size 1, same locations). A repeated-measures ANOVA revealed a non-significant effect of attended modality ($F_{(1,14)}$ = 2.85, p = 0.11), a significant effect of WM load ($F_{(1,14)}$ = 35.0, $p = 3.8 \times 10^{-5}$), and a (although very close to threshold) significant effect of the number of attended locations ($F_{(1,14)}$ = 4.68, p = 0.048). There was a non-significant interaction between modality and load ($F_{(1,14)}$ = 0.83, p = 0.38), but significant interactions between modality and number of locations ($F_{(1,14)}$ = 19.07, $p = 6.44 \times 10^{-4}$), load and number of locations ($F_{(1,14)}$ = 6.03, p = 0.028), and three-way interaction ($F_{(1,14)}$ = 14.78, p = 0.0018).

We used RSA to examine the predictive power of three regressors: (1) pointer load model—predicts similarity based on the number of *items* stored, regardless of sensory modality. (2) Spatial attention model—predicts similarity based on the total number of *locations attended* regardless of sensory modality and pointer load. (3) Sensory modality model—predicts similarity based on the stored sensory modality.

Modality-independent pointers and spatial attention reflect distinct selection processes

All models exhibited statistically significant correlations (spatial attention: W = 136, $p = 1.5 \times 10^{-5}$; attended modality: W = 128, $p = 3.8 \times 10^{-4}$; pointer load: W = 134, $p = 4.6 \times 10^{-5}$) that were sustained throughout both stimulus presentation and delay-period phases of the trial (Figure 6B). The spatial attention model explained robust variance, starting immediately after the second stimulus presentation when the number of attended positions could be disambiguated. Critically, the pointer load model explained unique variance throughout both stimulus

Please cite this article in press as: Suplica et al., Neural evidence for modality-independent storage in working memory, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.08.007

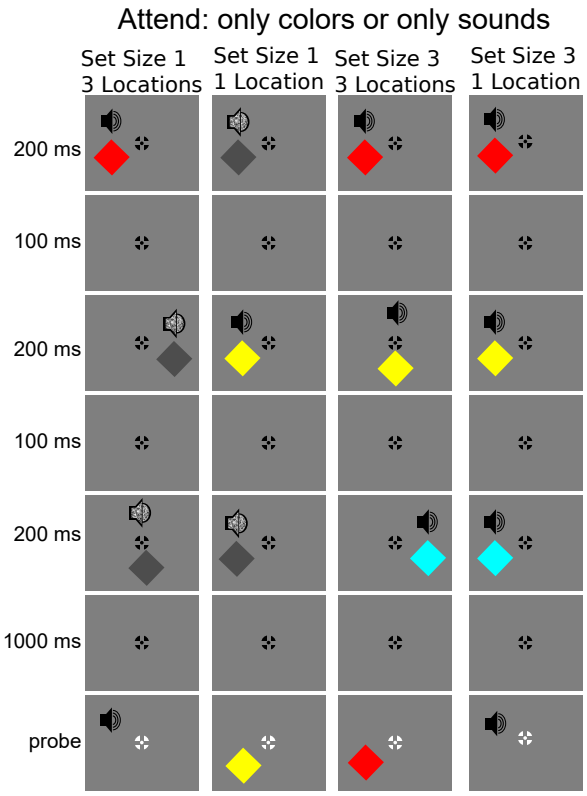

Figure 5. Experiment 2 task

Subjects attended to either only colors or only sounds. Each trial could have either one or three relevant stimuli (set size 1 trials had the same placeholders as experiment 1). Stimuli could appear in the same location for all presentations or in 3 unique locations.

presentation and the delay, showing that this load signal cannot be explained by variations in the spatial extent of attention or by modality-specific activity. We also replicated these RSA results while swapping the sensory modality model for two models: (1) visual load model—predicted similarity based on number of colors stored and (2) auditory load model—predicted similarity based on the number of sounds stored. This change did not affect the outcomes for the pointer load and spatial attention models (Figure 6C, pointer load: W = 122, $p = 1.68 \times 10^{-3}$; spatial attention: W = 136, $p = 1.5 \times 10^{-5}$). Both the visual (W = 119, $p = 3.14 \times 10^{-3}$) and auditory (W = 106, $p = 0.025$) load models explained reliable variance, although this was transient and early for the auditory load model. We note that when ranking values prior to computing correlations, as done in prior work,[39] the auditory load model no longer had a significant semi-partial correlation when averaging across the delay period. However, no other result for this analysis (crucially pointer load) changed substantially. These findings bolster the conclusion from experiment 1 that modality-specific neural activity was tracking sensory activity related to color and sound storage rather than a shift between discrete task sets. Furthermore, this modality-specific activity is independent from the pointer-related activity.

Interestingly, our spatial attention model, which assumed that attention was allocated to both targets and distractors, explained substantial variance. However, one might predict that spatial attention would eventually be restricted to locations containing targets. To test this, we reran RSA with a spatial attention model based solely on the number of relevant target locations (Figure S1). In fact, this model began to explain some variance, particularly toward the end of the delay (Figure S1B). However, it did not explain unique variance at any time point from that explained by the spatial attention model, which allocated attention to both targets and distractors (Figure S1C), although the variance explained was significant in the delay average (W = 103, $p = 0.037$). Regardless of which model for spatial attention was used, the pointer load and attended modality models explained robust variance throughout the stimulus and delay periods. These results suggest an interesting distinction: while signals tracking WM are shaped by relevance early after stimulus onset, signals tracking spatial attention exhibit spatial selectivity in a delayed fashion.[36]

We considered the possibility that our pointer load measure may reflect a more generalized signature of cognitive effort. While it has been difficult to find precise measures of effort,[40] some have argued that it plays a substantial role in the neural substrates of WM.[41,42] However, many studies that have looked at effort and WM load simultaneously, instead of using one as a proxy for the other, have identified them as unique signals that are both present in nearly all cognitive tasks.[17,43] We note that the tasks performed were not particularly difficult (mean accuracies of 95% and 94% in experiments 1 and 2, respectively), and there were only modest declines in accuracy in the higher load conditions (2% and 7% differences in experiments 1 and 2, respectively). Nevertheless, we carried out additional analyses to address this alternative explanation. Pupil size has often been used as an objective measure of effort, particularly when the sensory energy of the display is matched across putative variations in effort.[44,45] Indeed, in prior work, we have directly tested whether voltage-based decoding could be explained by variations in pupil size,[17] and we found no evidence that variations in pupil size could explain voltage-based measures of WM load. These findings notwithstanding, we examined the impact of effort by replicating all RSA analyses with an additional RDM representing each individual's average baselined change in pupil diameter per experimental condition.

The key question was whether or not including pupil data undermined our evidence for pointers. The pupil size model never explained unique variance, and including it did not have a consistent impact across the two experiments. In experiment 1's analysis (Figure S2), inclusion of the pupil size regressor reduced the variance explained of the pointer load model in the delay period (51/58 to 18/58 time points significant), although the model remained significant in the aggregate (W = 139, $p = 0.0399$). In contrast, in experiment 2, including pupil size had no reliable effect on the variance explained by pointer and sensory modality factors (Figure S3A: 67/70 significant time points both with and without pupil model, delay aggregate W = 114, $p = 4.27 * 10^{-4}$), and it showed only a slight reduction in the analysis where modality-specific load factors were included (Figure S3B: 67/70 to 46/70 significant time points, delay aggregate W = 99, $p = 0.0128$). The inconsistent impact of the pupil regressor suggests that cognitive effort is not a robust source of the modality-general pointer signal.

Please cite this article in press as: Suplica et al., Neural evidence for modality-independent storage in working memory, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.08.007

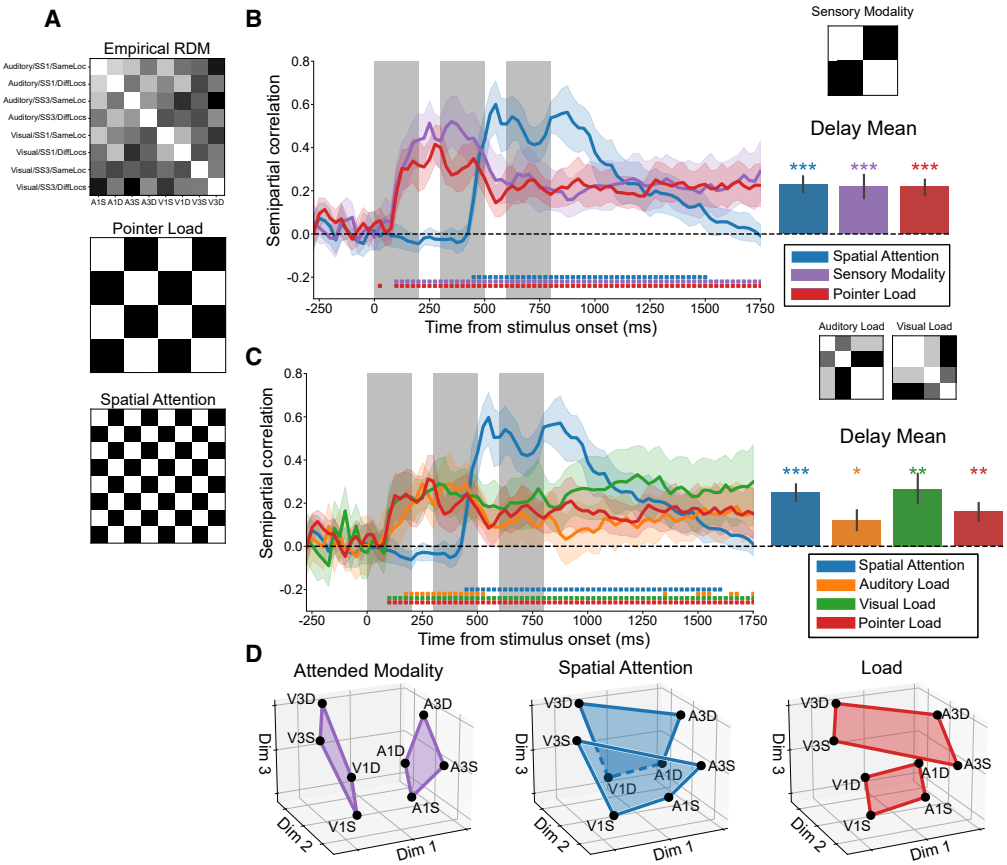

Figure 6. RSA results for experiment 2 (n = 16)

Condition labels (as shown in the empirical RDM) are such that the first letter (A/V) denotes modality, the number (1/3) denotes set size, and the final portion (Same/Diff) denotes stimuli appearing in the same or different spatial locations.

(A) RDMs used in all comparisons (representative empirical RDM, pointer load, and spatial attention).

(B) RSA results with the sensory modality model (auditory or visual, shown in top right). Left: semipartial correlations of each factor over time. The stim period is denoted by the gray-shaded region, and significant time points (Wilcoxon signed-rank test, FDR corrected) are denoted by colored boxes under the graph. Right: average correlation over delay period (t > 800 ms). ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$ (Wilcoxon signed-rank test, uncorrected).

(C) RSA results with auditory and visual modality-specific load models (in top right).

(D) 3D MDS projection of delay-period empirical RDM, points grouped by sensory modality, spatial attention demand, and WM load (points within the same colored plane are consistent within that feature). Tick marks are equal in distance.

See also Figures S1 and S3 for replications of this effect, an alternate spatial attention regressor, and a pupil size regressor.

Finally, an MDS projection of these experimental conditions into 3D space reveals a pleasingly interpretable cubic configuration (Figure 6D). Set size 1 and set size 3 define the top and bottom of the cube. Visual and auditory conditions define the left and right sides of the cube. And finally, the number of attended locations defines the front and back faces of the cube. Thus, the MDS plot implies separability between modality-independent pointers, feature-selective neural signals, and spatial attention. To summarize, experiment 2 replicated the initial demonstration of a dissociation between load-sensitive and feature-specific neural activity while also showing that the number of attended positions explained unique variance in EEG activity. These findings corroborate recent claims that the deployment of spatial attention and the gating of items into WM represent distinct aspects of attentional control.[18,20,36] Moreover, these results show that spatial attention cannot explain the modality-independent neural activity that tracked the number of items stored in WM.

DISCUSSION

WM is a cornerstone for broader intellectual function,[46–48] motivating the search for a clear taxonomy of its neural components. Our findings reveal a modality-general signature of WM storage that tracks the number of individuated items stored, independent of the sensory modality of those items. Multivariate decoding models trained on one sensory modality enabled precise decoding of WM load in the other sensory modality. Critically, RSA analyses corroborated the presence of this modality-general load signal while also showing that *distinct variance* in ongoing EEG activity was explained by the stored sensory modality and the number of attended positions in the display. Thus, parallel but

Please cite this article in press as: Suplica et al., Neural evidence for modality-independent storage in working memory, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.08.007

separate neural signals track the number of individuated items stored and the featural content of those items.[14,49] Moreover, we reinforce recent work that has argued for a dissociation between deployments of spatial attention and the selective encoding of items into WM.[18,36,50]

What is the computational role of this modality-general load activity? Our working hypothesis is that it reflects the deployment of spatiotemporal "pointers" that bind the selected items to the surrounding event context,[16,20] an operation that has been highlighted in major models of WM[21,23,51,52] and in prominent theories of dynamic visual cognition.[26,27,53] This literature elucidates how perception and action in dynamic environments require the observer to track attended items through space and time, despite changes in appearance or position.[26,27] Critically, the insight that perceptual inputs typically evolve across an unfolding event has motivated the idea that spatiotemporal tracking—sometimes referred to as *tokenization*[53]—is distinct from the maintenance of the featural details of tracked items. This separation affords continuous spatiotemporal tracking even when the featural details of the item are unstable. Thus, this contextual binding process provides an attractive explanation for a modality-independent, item-based signature of the number of relevant items stored in WM. Moreover, our findings show that these modality-general pointers operate in parallel with modality-specific signals that index the specific constellation of features held in WM (i.e., visual, auditory, or both). Thus, pointers may support item-based contextual binding while parallel processes support the maintenance of the attended featural details. This said, we are not presently able to rule out other processes that could generate this signal. For example, subitizing and judging numerosity is one process that seems to be limited by the number of individuated objects.[54] Further work is therefore necessary to directly link spatiotemporal tracking to this process of content-independent storage.

Our results are in line with past arguments for supramodal aspects of attentional networks mediating WM storage.[55] For example, top-down modulation by the dorsal attention network (DAN) has been heavily implicated in WM storage.[56–59] Majerus et al.[60] observed strong generalizability between fMRI patterns of WM for visual and verbal stimuli, particularly in posterior intraparietal sulcus, as well as across the DAN more broadly, during encoding and maintenance periods. Similarly, Rizza et al.[61] found strong correspondence between activation patterns for visual or auditory presentations of a spatial mapping task in several brain regions associated with the DAN, including the frontal eye fields and superior parietal lobule. While we see a clear connection with these ideas, our findings also show that WM encoding generates load signals that are dissociable from those tracking the deployment of spatial attention and cognition more broadly.[18,19,36,62]

Importantly, there are extant behavioral findings that, at first glance, appear to conflict with our proposal that WM storage depends on a modality-general pointer system. Various studies have found that WM performance is enhanced when the memory set includes mixed stimulus types (e.g., visual and verbal or visual and auditory stimuli).[32,63–65] In some cases, little or no interference is observed when information from separate modalities is concurrently stored, implying that each modality has its own storage capacity. Results like this have motivated models that explain behavioral response accuracy based entirely on the similarity between (noisy) competing memories[66] or due to interference during retrieval instead of limits on storage capacity per se.[52] Thus, while past work has provided both behavioral and neural evidence of item limits on WM storage,[67,68] it is nonetheless possible that these models are addressing distinct stages of processing with distinct limiting factors. Although past work suggests that item-based load signals are capacity-limited and predictive of individual WM ability,[16,34,35] interference that scales up with inter-item similarity may have a strong effect on decision stages of processing. Thus, interference during decision stages could explain the impact of inter-item similarity, even if a common pointer operation supports distinct sensory modalities.

In summary, we present electrophysiological evidence for a neural signature of WM storage that generalizes across sensory modalities, supporting a broad class of models that distinguish between abstract control processes that enable the gating and manipulation of specific thoughts and the stimulus-specific representations that determine the content of those thoughts.[20,69,70] We propose that these modality-general load signals reflect a content-independent operation for the contextual binding of items to the surrounding event, an essential component of cognition in virtually all perceptually guided tasks. Indeed, given the prominence of event codes in theories of long-term memory encoding and access, we hypothesize that the assignment of spatiotemporal pointers may serve as the initial step in the formation of durable episodic memories.[20,25,71]

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Darius Suplica (dsuplica@uchicago.edu).

Materials availability

This study did not generate unique new materials or reagents.

Data and code availability

- De-identified EEG data have been deposited in the Open Science Framework (OSF) at https://doi.org/10.17605/OSF.IO/J84R5 and are publicly available.
- All original code has been deposited at OSF and is publicly available at https://doi.org/10.17605/OSF.IO/J84R5.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This research was supported by theNational Institute of Mental Health grant no. ROIMH087214 and Office of Naval Research grant no. N00014-12-1-0972 to E.A.

AUTHOR CONTRIBUTIONS

Conceptualization, E.A., G.K.D., D.S., and H.M.J.; methodology, E.A., G.K.D., D.S., H.M.J., J.P.V., and H.C.N.; investigation, D.S. and G.K.D.; formal analysis, D.S. and H.M.J.; writing – first draft, D.S., E.A., and H.M.J.; writing – review and editing, all authors; and funding acquisition, E.A.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Experimental procedures
  - Apparati and Data Acquisition
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Overview
  - Artifact Rejection and Subject Exclusions
  - Decoding analyses
  - Statistical Tests
  - Representational Similarity Analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2025.08.007.

Received: March 20, 2025
Revised: July 7, 2025
Accepted: August 6, 2025

REFERENCES

1. Fuster, J.M., and Alexander, G.E. (1971). Neuron activity related to short-term memory. Science *173*, 652–654. https://doi.org/10.1126/science.173.3997.652.
2. Fuster, J.M., and Jervey, J.P. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. Science *212*, 952–955. https://doi.org/10.1126/science.7233192.
3. Goldman-Rakic, P.S. (1995). Cellular basis of working memory. Neuron *14*, 477–485. https://doi.org/10.1016/0896-6273(95)90304-6.
4. Harrison, S.A., and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. Nature *458*, 632–635. https://doi.org/10.1038/nature07832.
5. Serences, J.T., Ester, E.F., Vogel, E.K., and Awh, E. (2009). Stimulus-Specific Delay Activity in Human Primary Visual Cortex. Psychol. Sci. *20*, 207–214. https://doi.org/10.1111/j.1467-9280.2009.02276.x.
6. Yu, Q., and Shim, W.M. (2017). Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working memory. NeuroImage *157*, 97–107. https://doi.org/10.1016/j.neuroimage.2017.05.055.
7. Ester, E.F., Anderson, D.E., Serences, J.T., and Awh, E. (2013). A neural measure of precision in visual working memory. J. Cogn. Neurosci. *25*, 754–761. https://doi.org/10.1162/jocn_a_00357.
8. Vo, V.A., Sutterer, D.W., Foster, J.J., Sprague, T.C., Awh, E., and Serences, J.T. (2022). Shared Representational Formats for Information Maintained in Working Memory and Information Retrieved from Long-Term Memory. Cereb. Cortex *32*, 1077–1092. https://doi.org/10.1093/cercor/bhab267.
9. Zhao, Y., Kuai, S., Zanto, T.P., and Ku, Y. (2020). Neural Correlates Underlying the Precision of Visual Working Memory. Neuroscience *425*, 301–311. https://doi.org/10.1016/j.neuroscience.2019.11.037.
10. Panichello, M.F., and Buschman, T.J. (2021). Shared mechanisms underlie the control of working memory and attention. Nature *592*, 601–605. https://doi.org/10.1038/s41586-021-03390-w.
11. Libby, A., and Buschman, T.J. (2021). Rotational dynamics reduce interference between sensory and memory representations. Nat. Neurosci. *24*, 715–726. https://doi.org/10.1038/s41593-021-00821-9.
12. Lewis-Peacock, J.A., Drysdale, A.T., Oberauer, K., and Postle, B.R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. J. Cogn. Neurosci. *24*, 61–79. https://doi.org/10.1162/jocn_a_00140.
13. Vogel, E.K., and Machizawa, M.G. (2004). Neural activity predicts individual differences in visual working memory capacity. Nature *428*, 748–751. https://doi.org/10.1038/nature02447.
14. Xu, Y., and Chun, M.M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. Nature *440*, 91–95. https://doi.org/10.1038/nature04262.
15. Rajsic, J., Burton, J.A., and 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.
16. Thyer, W., Adam, K.C.S., Diaz, G.K., Velázquez Sánchez, I.N., Vogel, E.K., and Awh, E. (2022). Storage in Visual Working Memory Recruits a Content-Independent Pointer System. Psychol. Sci. *33*, 1680–1694. https://doi.org/10.1177/09567976221090923.
17. Jones, H.M., Thyer, W.S., Suplica, D., and Awh, E. (2024). Cortically Disparate Visual Features Evoke Content-Independent Load Signals during Storage in Working Memory. J. Neurosci. *44*, e0448242024. https://doi.org/10.1523/JNEUROSCI.0448-24.2024.
18. Jones, H.M., Diaz, G.K., Ngiam, W.X.Q., and Awh, E. (2024). Electroencephalogram Decoding Reveals Distinct Processes for Directing Spatial Attention and Encoding Into Working Memory. Psychol. Sci. *35*, 1108–1138. https://doi.org/10.1177/09567976241263002.
19. Diaz, G.K., Vogel, E.K., and Awh, E. (2021). Perceptual Grouping Reveals Distinct Roles for Sustained Slow Wave Activity and Alpha Oscillations in Working Memory. J. Cogn. Neurosci. *33*, 1354–1364. https://doi.org/10.1162/jocn_a_01719.
20. Awh, E., and Vogel, E.K. (2025). Working memory needs pointers. Trends Cogn. Sci. *29*, 230–241. https://doi.org/10.1016/j.tics.2024.12.006.
21. Baddeley, A. (2000). The episodic buffer: a new component of working memory? Trends Cogn. Sci. *4*, 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2.
22. Swan, G., and Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. Atten. Percept. Psychophys. *76*, 2136–2157. https://doi.org/10.3758/s13414-014-0633-3.
23. Hedayati, S., O'Donnell, R.E., and Wyble, B. (2022). A model of working memory for latent representations. Nat. Hum. Behav. *6*, 709–719. https://doi.org/10.1038/s41562-021-01264-9.
24. Oberauer, K., and Lin, H.-Y. (2024). An interference model for visual and verbal working memory. J. Exp. Psychol. Learn. Mem. Cogn. *50*, 858–888. https://doi.org/10.1037/xlm0001303.
25. Yonelinas, A.P. (2024). The role of recollection and familiarity in visual working memory: A mixture of threshold and signal detection processes. Psychol. Rev. *131*, 321–348. https://doi.org/10.1037/rev0000432.
26. Pylyshyn, Z.W. (2009). Perception, Representation, and the World: the FINST That Binds. In Computation, Cognition, and Pylyshyn, Don Dedrick, and Lana Trick, eds. (Boston Review), pp. 3–48.
27. Kahneman, D., Treisman, A., and Gibbs, B.J. (1992). The reviewing of object files: Object-specific integration of information. Cogn. Psychol. *24*, 175–219. https://doi.org/10.1016/0010-0285(92)90007-O.
28. Baddeley, A.D. (1986). Working Memory Oxford (Oxford University).
29. Saults, J.S., and Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. J. Exp. Psychol. Gen. *136*, 663–684. https://doi.org/10.1037/0096-3445.136.4.663.
30. Cowan, N., Saults, J.S., and Blume, C.L. (2014). Central and peripheral components of working memory storage. J. Exp. Psychol. Gen. *143*, 1806–1836. https://doi.org/10.1037/a0036814.
31. Fougnie, D., Zughni, S., Godwin, D., and Marois, R. (2015). Working memory storage is intrinsically domain specific. J. Exp. Psychol. Gen. *144*, 30–47. https://doi.org/10.1037/a0038211.

32. Cocchini, G., Logie, R.H., Della Sala, S.D., MacPherson, S.E., and Baddeley, A.D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. Mem. Cognit. *30*, 1086–1095. https://doi.org/10.3758/BF03194326.

33. Guimond, S., Vachon, F., Nolden, S., Lefebvre, C., Grimault, S., and Jolicoeur, P. (2011). Electrophysiological correlates of the maintenance of the representation of pitch objects in acoustic short-term memory. Psychophysiology *48*, 1500–1509. https://doi.org/10.1111/j.1469-8986.2011.01234.x.

34. Luria, R., Balaban, H., Awh, E., and Vogel, E.K. (2016). The contralateral delay activity as a neural measure of visual working memory. Neurosci. Biobehav. Rev. *62*, 100–108. https://doi.org/10.1016/j.neubiorev.2016.01.003.

35. Adam, K.C.S., Vogel, E.K., and 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.

36. Hakim, N., Adam, K.C.S., Gunseli, E., Awh, E., and Vogel, E.K. (2019). Dissecting the Neural Focus of Attention Reveals Distinct Processes for Spatial Attention and Object-Based Storage in Visual Working Memory. Psychol. Sci. *30*, 526–540. https://doi.org/10.1177/0956797619830384.

37. Sandhaeger, F., and Siegel, M. (2023). Testing the generalization of neural representations. NeuroImage *278*, 120258. https://doi.org/10.1016/j.neuroimage.2023.120258.

38. Kikumoto, A., and Mayr, U. (2020). Conjunctive representations that integrate stimuli, responses, and rules are critical for action selection. Proc. Natl. Acad. Sci. USA *117*, 10603–10608. https://doi.org/10.1073/pnas.1922166117.

39. Kiat, J.E., Hayes, T.R., Henderson, J.M., and Luck, S.J. (2022). Rapid Extraction of the Spatial Distribution of Physical Saliency and Semantic Informativeness from Natural Scenes in the Human Brain. J. Neurosci. *42*, 97–108. https://doi.org/10.1523/JNEUROSCI.0602-21.2021.

40. Westbrook, A., and Braver, T.S. (2015). Cognitive effort: A neuroeconomic approach. Cogn. Affect. Behav. Neurosci. *15*, 395–415. https://doi.org/10.3758/s13415-015-0334-y.

41. Master, S.L., Li, S., and Curtis, C.E. (2024). Trying Harder: How Cognitive Effort Sculpts Neural Representations during Working Memory. J. Neurosci. *44*, e0060242024. https://doi.org/10.1523/JNEUROSCI.0060-24.2024.

42. Kitzbichler, M.G., Henson, R.N.A., Smith, M.L., Nathan, P.J., and Bullmore, E.T. (2011). Cognitive Effort Drives Workspace Configuration of Human Brain Functional Networks. J. Neurosci. *31*, 8259–8270. https://doi.org/10.1523/JNEUROSCI.0440-11.2011.

43. Kardan, O., Adam, K.C.S., Mance, I., Churchill, N.W., Vogel, E.K., and Berman, M.G. (2020). Distinguishing cognitive effort and working memory load using scale-invariance and alpha suppression in EEG. Neuroimage *211*, 116622. https://doi.org/10.1016/j.neuroimage.2020.116622.

44. van der Wel, P., and van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. Psychon. Bull. Rev. *25*, 2005–2015. https://doi.org/10.3758/s13423-018-1432-y.

45. Keene, P.A., deBettencourt, M.T., Awh, E., and Vogel, E.K. (2022). Pupillometry signatures of sustained attention and working memory. Atten. Percept. Psychophys. *84*, 2472–2482. https://doi.org/10.3758/s13414-022-02557-5.

46. Cowan, N., Elliott, E.M., Scott Saults, J., Morey, C.C., Mattox, S., Hismjatullina, A., and Conway, A.R.A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. Cogn. Psychol. *51*, 42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001.

47. Unsworth, N., Fukuda, K., Awh, E., and Vogel, E.K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. Cogn. Psychol. *71*, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003.

48. Fukuda, K., Vogel, E., Mayr, U., and Awh, E. (2010). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. Psychon. Bull. Rev. *17*, 673–679. https://doi.org/10.3758/17.5.673.

49. Xu, Y., and Chun, M.M. (2009). Selecting and perceiving multiple visual objects. Trends Cogn. Sci. *13*, 167–174. https://doi.org/10.1016/j.tics.2009.01.008.

50. Günseli, E., Fahrenfort, J.J., van Moorselaar, D., Daoultzis, K.C., Meeter, M., and Olivers, C.N.L. (2019). EEG dynamics reveal a dissociation between storage and selective attention within working memory. Sci. Rep. *9*, 13499. https://doi.org/10.1038/s41598-019-49577-0.

51. Oberauer, K. (2019). Working Memory Capacity Limits Memory for Bindings. J. Cogn. *2*, 40. https://doi.org/10.5334/joc.86.

52. Oberauer, K., and Lin, H.-Y. (2017). An interference model of visual working memory. Psychol. Rev. *124*, 21–59. https://doi.org/10.1037/rev0000044.

53. Kanwisher, N.G. (1987). Repetition blindness: Type recognition without token individuation. Cognition *27*, 117–143. https://doi.org/10.1016/0010-0277(87)90016-3.

54. Halberda, J., Sires, S.F., and Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. Psychol. Sci. *17*, 572–576. https://doi.org/10.1111/j.1467-9280.2006.01746.x.

55. Rajan, A., Meyyappan, S., Liu, Y., Samuel, I.B.H., Nandi, B., Mangun, G.R., and Ding, M. (2021). The Microstructure of Attentional Control in the Dorsal Attention Network. J. Cogn. Neurosci. *33*, 965–983. https://doi.org/10.1162/jocn_a_01710.

56. Gazzaley, A., and Nobre, A.C. (2012). Top-down modulation: bridging selective attention and working memory. Trends Cogn. Sci. *16*, 129–135. https://doi.org/10.1016/j.tics.2011.11.014.

57. Majerus, S., Attout, L., D'Argembeau, A., Degueldre, C., Fias, W., Maquet, P., Martinez Perez, T., Stawarczyk, D., Salmon, E., Van der Linden, M., et al. (2012). Attention Supports Verbal Short-Term Memory via Competition between Dorsal and Ventral Attention Networks. Cereb. Cortex *22*, 1086–1097. https://doi.org/10.1093/cercor/bhr174.

58. Naghavi, H.R., and Nyberg, L. (2005). Common fronto-parietal activity in attention, memory, and consciousness: Shared demands on integration? Conscious. Cogn. *14*, 390–425. https://doi.org/10.1016/j.concog.2004.10.003.

59. Todd, J.J., and 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.

60. Majerus, S., Cowan, N., Péters, F., Van Calster, L., Phillips, C., and Schrouff, J. (2016). Cross-Modal Decoding of Neural Patterns Associated with Working Memory: Evidence for Attention-Based Accounts of Working Memory. Cereb. Cortex *26*, 166–179. https://doi.org/10.1093/cercor/bhu189.

61. Rizza, A., Pedale, T., Mastroberardino, S., Olivetti Belardinelli, M., Van der Lubbe, R.H.J., Spence, C., and Santangelo, V. (2024). Working Memory Maintenance of Visual and Auditory Spatial Information Relies on Supramodal Neural Codes in the Dorsal Frontoparietal Cortex. Brain Sci. *14*, 123. https://doi.org/10.3390/brainsci14020123.

62. Bae, G.-Y., and Luck, S.J. (2018). Dissociable Decoding of Spatial Attention and Working Memory from EEG Oscillations and Sustained Potentials. J. Neurosci. *38*, 409–422. https://doi.org/10.1523/JNEUROSCI.2860-17.2017.

63. Baddeley, A.D., and Hitch, G. (1974). Working Memory. In Psychology of Learning and Motivation (Elsevier), pp. 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1.

64. Fougnie, D., and Marois, R. (2011). What limits working memory capacity? Evidence for modality-specific sources to the simultaneous storage of visual and auditory arrays. J. Exp. Psychol. Learn. Mem. Cogn. *37*, 1329–1341. https://doi.org/10.1037/a0024834.

65. Henderson, L. (1972). Visual and verbal codes: Spatial information survives the icon. Q. J. Exp. Psychol. *24*, 439–447. https://doi.org/10.1080/14640747208400303.

66. Schurgin, M.W., Wixted, J.T., and Brady, T.F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. Nat. Hum. Behav. *4*, 1156–1172. https://doi.org/10.1038/s41562-020-00938-0.

67. Adam, K.C.S., Vogel, E.K., and Awh, E. (2017). Clear evidence for item limits in visual working memory. Cogn. Psychol. *97*, 79–97. https://doi.org/10.1016/j.cogpsych.2017.07.001.

68. Ngiam, W.X.Q., Foster, J.J., Adam, K.C.S., and Awh, E. (2023). Distinguishing guesses from fuzzy memories: Further evidence for item limits in visual working memory. Atten. Percept. Psychophys. *85*, 1695–1709. https://doi.org/10.3758/s13414-022-02631-y.

69. Lundqvist, M., Brincat, S.L., Rose, J., Warden, M.R., Buschman, T., Miller, E.K., and Herman, P. (2022). Spatial computing for the control of working memory. Preprint at bioRxiv. https://doi.org/10.1101/2020.12.30.424833.

70. O'Reilly, R.C., Ranganath, C., and Russin, J.L. (2022). The Structure of Systematicity in the Brain. Curr. Dir. Psychol. Sci. *31*, 124–130. https://doi.org/10.1177/09637214211049233.

71. Fukuda, K., and Vogel, E.K. (2019). Visual short-term memory capacity predicts the "bandwidth" of visual long-term memory encoding. Mem. Cognit. *47*, 1481–1497. https://doi.org/10.3758/s13421-019-00954-0.

72. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

73. Vallat, R. (2018). Pingouin: statistics in Python. J. Open Source Software *3*, 1026. https://doi.org/10.21105/joss.01026.

74. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

75. Uddin, S., Heald, S.L.M., Van Hedger, S.C., Klos, S., and Nusbaum, H.C. (2018). Understanding environmental sounds in sentence context. Cognition *172*, 134–143. https://doi.org/10.1016/j.cognition.2017.12.009.

76. Uddin, S., Heald, S.L.M., Van Hedger, S.C., and Nusbaum, H.C. (2018). Hearing sounds as words: Neural responses to environmental sounds in the context of fluent speech. Brain Lang. *179*, 51–61. https://doi.org/10.1016/j.bandl.2018.02.004.

77. Martin, R.A. (2021). PyPortfolioOpt: portfolio optimization in Python. J. Open Source Software *6*, 3066. https://doi.org/10.21105/joss.03066.

78. Gardner, W.G., and Martin, K.D. (1995). HRTF measurements of a KEMAR. J. Acoust. Soc. Am. *97*, 3907–3908. https://doi.org/10.1121/1.412407.

79. Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J.K. (2019). PsychoPy2: Experiments in behavior made easy. Behav. Res. Methods *51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y.

80. Thaler, L., Schütz, A.C., Goodale, M.A., and Gegenfurtner, K.R. (2013). What is the best fixation target? The effect of target shape on stability of fixational eye movements. Vision Res. *76*, 31–42. https://doi.org/10.1016/j.visres.2012.10.012.

81. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. B Methodol. *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

82. Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. *2*, 4. https://doi.org/10.3389/neuro.06.004.2008.

83. Diedrichsen, J., Provost, S., and Zareamoghaddam, H. (2016). On the distribution of cross-validated Mahalanobis distances. Preprint at arXiv. https://doi.org/10.48550/arXiv.1607.01371.

84. Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., and Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. PLoS Comp. Biol. *10*, e1003553. https://doi.org/10.1371/journal.pcbi.1003553.

85. Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., and 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.

86. Ledoit, O., and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. J. Multivariate Anal. *88*, 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4.

Please cite this article in press as: Suplica et al., Neural evidence for modality-independent storage in working memory, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.08.007

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
EEG data	This paper	https://doi.org/10.17605/OSF.IO/J84R5
Software and algorithms		
Analysis code	This paper	https://doi.org/10.17605/OSF.IO/J84R5
Scikit-learn	Pedregosa et al.[72]	https://github.com/scikit-learn/scikit-learn
Pingouin	Vallat[73]	https://github.com/raphaelvallat/pingouin
Scipy	Virtanen et al.[74]	https://github.com/scipy/scipy

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants were recruited from the greater University of Chicago and Hyde Park community in exchange for payment ($20 / hour). Twenty-eight participated in experiment 1 (4 excluded), and twenty in experiment 2 (4 excluded). Subjects were excluded from the final sample if fewer than 150 trials per stimulus condition were present after rejection of trials containing artifacts. Participants were between the age of 18-35, reported normal or corrected-to-normal color vision and hearing, and no history of stroke or neurological disorder. All participants gave informed consent according to procedures approved by the University of Chicago Institutional Review Board.

Our target sample size for experiment 1 was 24 participants, a conservative estimate based on prior studies in the lab.[16,17] Twenty-eight subjects (mean age 24.7, 13 male, 13 female, 1 declined to state, 1 data lost) participated. Four subjects were excluded from the final sample due to excessive eye movements and noise in EEG recordings (see artifact rejection). All pupil size analyses were run on 19 subjects with usable pupil data.

Our target sample size for experiment 2 was 16 participants. We assumed that increasing the difference between set sizes would improve overall SNR, and prior pilot studies still found reliable load effects at 16 subjects. A total of 20 participants completed the experiment (12 female, 6 male, 1 declined to state, 1 data not saved, mean age, 24.6, SD 4.3), with 16 remaining after artifact rejection. All pupil size analyses were run on 14 subjects with usable pupil data.

We recorded self-identified sex for all participants, with an option to decline to specify. We did not record ancestry, race, ethnicity, or socioeconomic status. However, we do not believe that these factors should substantially affect the results. All analyses were performed within subjects.

METHOD DETAILS

Experimental procedures

Both experiments were single-probe sequential change detection tasks. We used bimodal stimuli, in which a colored diamond was presented with each auditory stimulus, in a spatially colocalized manner. We used a set of 5 recordings of environmental sounds used in prior studies,[75,76] clips each of which lasted 200ms. The possible sounds were of a camera shutter, crow, car horn, bell, or zipper, in addition to a distractor white noise sound. Stimuli were matched on intensity (70 dB SPL), sampling rate (44100 Hz), and duration. All stimuli had a bias towards the right (90°), left (-90°) or center (30°), and were synthesized by filtering recordings through a head-related transfer function[77,78] based on volume balance. Visual stimuli were colored red #FF0000, yellow #FFFF00, green #00FF00, cyan #00FFFF, blue #0000FF, or gray (distractor) #555555 diamonds with a 1 degree visual angle radius, lateralized to the left, right (both by 6 degrees), or center of fixation. Luminances of all stimuli were, respectively, 25.2, 106.9, 83.84, 92.61, 9.50, or 17.90 cd/m^3. Stimuli were jittered by up to 2 degrees from their center point. Stimuli presented at the center location could not be less than 1.5 degrees from fixation. We predicted that spatial and temporal colocalization would act as grouping cues which would cause the stimuli would be viewed as bound audiovisual objects; this aligned with our results and with subjects' subjective experience (upon being asked after data collection).

In experiment 1, two pairs of visual and auditory stimuli were presented for 200ms, with a 100ms ISI. In set size 2 trials, 2 target stimulus pair were presented. In set size 1 trials, placeholder pairs (gray squares and white noise) were presented at one timepoint in order to match stimulus energy across set sizes. The order of target pairs and placeholder pairs was random across trials. We manipulated the information subjects stored in WM by directing subjects to only remember auditory items, visual items, or all items (conjunction) for a given block. Therefore, experiment 1 had a 3 (attended-modality conditions) x 2 (set sizes) design. Subjects completed 24 blocks of 56 trials, cycling between sensory modality conditions, for a total of 1,344 trials and 224 trials per cell of

Please cite this article in press as: Suplica et al., Neural evidence for modality-independent storage in working memory, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.08.007

the design. After a 1s delay following the second stimulus pair presentation, a single probe was presented according to the attended modality. Subjects answered whether this was one of the previously presented objects.

The goal of experiment 2 was to determine whether the modality general load signal seen in experiment 1 was driven by spatial attention, or if it persisted even when spatial attention was directly manipulated. We replicated experiment 1's design with the following modifications. First, we varied the number of attended spatial locations. In "same location" conditions, all stimuli appeared in the same location after each other, randomly selected from the options of left, right, or center, ensuring that all trials had equal demands to spatial attention regardless of working memory load. "Different location" conditions were identical to those in experiment 1, with stimulus pairs appearing at distinct locations within a presentation sequence. Second, we removed the conjunction condition to reduce the experiment run time. Finally, we changed the set sizes to 1 and 3 to increase the magnitude of the possible spatial attention confound, as well as the magnitude of the putative WM load signal. This design allows us to test the generalization of load signals while controlling the spatial information between conditions. Therefore, experiment 3 had a 2 (attended-modalities) x 2 (set sizes) x 2 (spatial conditions) design. In total, subjects completed 26 blocks of 56 trials, for a total of 1456 trials and 182 trials per cell of the design. All other details, including stimuli and placeholders, were identical to experiment 1.

Apparati and Data Acquisition

Subjects were tested in a dimly lit, electrically shielded chamber. Stimuli were presented on a neutral gray background (RGB: 127,127,127). Subjects performed the experiment on a gamma-corrected 24-in. LCD monitor, with a 1920x1080 resolution and 120Hz refresh rate, at a distance of 75 cm. Subjects gave their response by a keyboard press (z = same, / = different), and rested their heads on a foam chin rest for the duration of trial blocks. Auditory stimuli were presented via in-ear earphones (Neurospec ER3C), with disposable foam tips. Experiments were designed using PsychoPy3.[79] During all trials, subjects were expected to fixate on a cross in the center; we used the shape identified by Thaler et al. (2013)[80] as optimal for fixation. We applied a real-time eye-tracking rejection procedure during the experiment. In this, any eye movements (determined by eye-tracking) of more than 1.5 degrees visual angle from the center of fixation would cause the trial to be immediately aborted and a trial of the same condition repeated at the end of the experiment.

We recorded EEG activity from 30 active Ag/AgCl electrodes (Brain Products actiCHamp) at International 10-20 system sites Fp1/2, Fz, F3/4, F7/8, FC1/2, FC5/6, Cz, C3/4, CP1/2, CP5/6, Pz, P3/4, P7/8, PO3/4, PO7/8, Oz, and O1/2. Prior to starting recording, impedances for all electrodes were below 10 kΩ. A ground electrode was placed at position FPz, and two references were placed on the left and right mastoids. All electrodes were referenced online to the right mastoid, and rereferenced offline to the algebraic average of the left and right mastoids. Additionally, to track eye movements, we recorded EOG data using passive electrodes. We placed a ground electrode on the left cheek, two electrodes to track horizontal eye movements $\sim$ 1 cm from the horizontal canthus of each eye, and two to track vertical eye movements above and below the right eye. Data were recorded at 1000 Hz and filtered online (high cutoff: 80Hz, low cutoff: 0.01 Hz, slope: 12 dB / octave) using BrainVision Recorder on a Windows PC. Eye-tracking data was collected using a desk-mounted infrared EyeLink 1000 Plus system (SR Research) at 1000 Hz, that was calibrated between blocks and after any breaks. We used eye-tracking data preferentially for artifact rejection, and EOG when eye-tracking was not available or malfunctioning.In experiment 1, data were segmented offline to 1800 ms epochs time-locked to the onset of the first stimulus (-300 ms to +1500 ms), and baseline-corrected with the 300 ms immediately prior to stimulus onset. In experiment 2, 2100 ms epochs were used (-300 ms to +1800 ms).

QUANTIFICATION AND STATISTICAL ANALYSIS

Overview

For each analysis, sample size (number of participants) and the statistical tests used are present in the figure legends. Sample size was uniform for all analyses in the same figure. Where asterisk notation was used to denote significance, we used the following thresholds: *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$, n.s = $p > 0.05$. P values and Bayes Factors for statistical tests, where applicable, are presented in the results text. When evaluating performance over time, significant time windows are marked with a colored box, corresponding to a FDR corrected p-value (see Statistical Tests subsection) below 0.05. Bar heights and plot lines represent means, and error bars (included shaded regions) denote SEM.

Artifact Rejection and Subject Exclusions

We applied an automated artifact rejection procedure to identify trials contaminated by ocular or muscular artifacts, and then visually inspected all trials to manually reject significantly noisy trials or remove false positives from rejection (these primarily occurred due to drift in the EOG or brief lapses in eye-tracking recording. Trials were rejected if they met any of the following criteria:

- Eyetracking: Eye movements of over 1.5 degrees visual angle away from fixation.
- EOG: Any trial where the absolute EOG value was greater than 75 μV from pre-trial baseline. This was only used when eye-tracking was not available for a subject.
- Blinks: Trials where the eye-tracker lost focus on the eye for any portion. Several of these were false positives, and were re-included in the final sample if vertical EOG and frontal EEG electrodes did not show a deflection.
- Saturated Electrodes: Trials where any channel has flatlined or exhibits a step function

Please cite this article in press as: Suplica et al., Neural evidence for modality-independent storage in working memory, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.08.007

- Additional EEG artifacts: skin potentials, muscle artifacts, and excessive noise. In experiment 1, we excluded 1 electrode from analysis for 1 subject due to high noise. However, we still had sufficient data to successfully run all analyses.

In experiment 1, we rejected between 5.6% (conjunction ss1) to 6.5% (auditory ss1) of trials. In experiment 2, we rejected between 3.5% (auditory ss3 different locations) to 4.3% (auditory ss1 1 location) of trials.

Subjects were excluded from the final sample if they had fewer than 150 trials per load condition remaining after artifact rejection.

Decoding analyses

We applied a multivariate binned-trial classification procedure (mvLoad) to classify working memory load within subjects.[35] When attempting to use cross-decoding to evaluate the generalizability of neural patterns across feature dimensions, we used a parallel but higher-powered alternative metric to decoding accuracy which we called "hyperplane contrast." The goal of this approach is to draw a hyperplane in n-dimensional space at the midpoint of the logit function as determined by the training set. We then measure the average signed distance of each condition to this hyperplane (scikit-learn: decision_function) and take the difference between the two distances as the contrast. This is done both for held-out samples from the training set, as well as a test group that the classifier is entirely naïve to when appropriate. This is analogous to estimating the magnitude of the multivariate difference vector separating each pair of conditions, and is insensitive to the types of additive shifts that changes in other signals may cause.[17] These distances are additionally easier to interpret than classification accuracy, as they are not affected by ceiling effects and apply in a continuous space as opposed to the average of several binary classifications.

We used the following pipeline for decoding analyses. On a given iteration, we first formed "bins" by randomly sampling trials from the same condition into groups of 20, and averaging across the trials in each group. This was done without replacement, and any trials that were not assigned to a bin (due to the trial counts not cleanly dividing by 20) were dropped for that iteration. We then split these binned trials into a training (70%) and testing (30%) set using the train_test_split function from scikit-learn.[72] Training sets were always balanced via random dropping to ensure that an equal number of trial bins from each training condition were used. We averaged these time series using a sliding window (50ms, 25ms step). Training data were standardized for each time window using the StandardScaler function, and test data were standardized to the mean and standard deviation of the training data. These data were then fed into a regularized logistic regression classifier (scikit-learn LogisticRegression), and used to predict appropriate labels and a confidence value for the test set. This procedure was repeated over 1,000 iterations, including the initial random binning, for each subject and each time window.

We chose a bin size of 20 a priori based on prior work.[16,18] However, to ensure our results were invariant to the specific bin size used we performed a downsampling analysis while varying the varying the bin size, using trial bins of size 1, 5, 10, 20, and 30 (Figure S4). SNR generally monotonically increased with greater bin size, particularly when crosstraining. We additionally observed sustained crosstraining performance throughout the delay period when using a bin size of 10 or greater; the 1 trial and 5 trial bins were more sporadic.

Statistical Tests

Hyperplane contrasts were evaluated using a one-tailed paired t-test, as we would not expect these to be meaningly lower than zero. We compared the mean contrast of each subject at each timepoint. Because we conducted independent significance tests at multiple timepoints, all p-values were corrected according to the Benjamini-Hochberg procedure,[81] with a false discovery rate set to 0.05. When we were comparing the values of two contrasts against each other, we used a combination of one-tailed and two-tailed tests based on our alternative hypotheses.

For all parametric tests, we calculated the bayes factor (BF) using the pingouin package,[73] which compares the relative degree of evidence for the null and alternative hypotheses. Calculation of bayes factors addresses one of the primary weaknesses of null-hypothesis significance testing, which is its difficulty in interpreting nonsignificant results. Instead of providing a likelihood statistic for the null hypothesis, the bayes factor represents the likelihood ratio of the null and alternative hypotheses. Generally, a BF of 1 represents equal evidence for both (absence of evidence), while BFs of 1/3 and 1/10 represent weak and strong, respectively, evidence for the null hypothesis, and BFs of 3 and 10 represent weak and strong evidence for the alternative. All other statistical results were calculated using Scipy.[74]

To ensure robustness of our correlations in Figure 4, we removed subjects with outlier values from regressions. For each regression, we calculated Cook's distance for each datapoint to identify high-leverage outliers. We defined influential subjects as those with a Cook's distance greater than 0.5. Two high-leverage subjects, with Cook's distances of 0.54 and 0.95 were excluded. We replicated all other analyses for experiment 1 with these two subjects excluded, and did not observe any qualitative differences.

Representational Similarity Analysis

We additionally used representational similarity analysis,[82] a form of multivariate pattern analysis which analyzes the similarity of activation patterns in the brain across conditions in order to more effectively model the effects of load and stimulus modality simultaneously. Unlike evaluating classifier discriminability, it allows us to examine the effects of one factor (say, set size) when the effects of others are controlled for.

We performed RSA analyses by computing a distance metric between each possible pair of conditions, and then combining them into a single matrix of dissimilarities (representational dissimilarity matrix: RDM). The distance metric used here was a crossvalidated

Please cite this article in press as: Suplica et al., Neural evidence for modality-independent storage in working memory, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.08.007

estimate of the Mahalanobis distance, also known as the linear discriminant contrast (LDC), which produces a distance metric that is largely unbiased by noise.[83–85] We note that the LDC is conceptually analogous to the hyperplane contrast presented above, just based specifically on a Linear Discriminant Analysis (LDA), rather than a Logistic Regression model. This distance metric is particularly ideal for inference tests, as unlike non-cross-validated distance measures, which have a lower bound of 0, the LDC can take on negative values, such that if two samples vary solely due to chance, the mean of the distribution of their LDC values across iterations will be zero. This procedure used a similar pipeline as the decoding analyses, although we did not bin trials prior to classification. Training and testing trials were obtained using the StratifiedShuffleSplit function from scikit-learn using a 50/50 split. Distances were calculated over subjects and timepoints using a 50 ms sliding window (25 ms step), over 10,000 iterations, and subsequently averaged over each iteration. We computed LDC between two condition pairs as

$$d_M^2(\vec{x}, \vec{y}) = \left(\vec{x}_{train} - \vec{y}_{train}\right) * \Sigma_{train}^{-1} * \left(\vec{x}_{test} - \vec{y}_{test}\right)$$

Where $\vec{x}$ and $\vec{y}$ are the average EEG voltage values for a given condition at the given timepoint (i.e. a vector of the same length as the number of electrodes), computed for both the training and testing half, respectively, and Σ_{train} is the covariance matrix of the training set. We calculated this by first demeaning trials by subtracting the condition average voltage, computing the covariance matrix across conditions, and then regularizing by the Ledoit-Wolf procedure.[85,86]

In order to evaluate models, we designed RDMs which were intended to describe the predictions of different theoretical factors which may be present in the data. Prior work[17,18,39] has rank transformed RDMs prior to calculating correlations. We did not do so to improve interpretability, but note that the results of our models are nearly qualitatively identical with or without the rank transformation; the only noticeable effect was that the auditory load model in Figure 6C no longer became significant after ranking. We then applied a linear regression model with the empirical RDM as the independent variable, and each theoretical factor model as a predictor. Then, we calculated the semipartial correlation (a measure of the unique variance explained by each regressor) of each theoretical model to the empirical RDM.

In Experiment 1, we tested the following theoretical factors. First, we included a pointer load model, assumed a distance of 0 between all conditions of the same set size regardless of attended modality (e.g. visual-1 and auditory-1) and a distance of 1 between all conditions of the other set size (e.g. visual-1 and visual-2). Second, we included an alternative, feature based model, which coded each condition by the number of feature values being maintained. In this model, for example, conjunction set size 2 would have 4 features (2 colors and 2 sounds), while visual set size 1 would have 1 feature, meaning the difference between them is 3. We also included two models designed to code for how the different modality conditions may be represented. In the Graded model, the conjunction condition fell between the visual and auditory conditions (e.g. auditory=-1, conjunction=0, visual=1), reflecting a single attended-feature axis. In the Discrete model, each condition (auditory, conjunction, visual) was equidistant from the other 2 (a distance of 1 to each other), reflecting an equilateral triangle.

Experiment 2 included the following theoretical factors. First, the pointer load model was identical to that of experiment 1, given there were only 2 set sizes to compare against. Next, a spatial attention factor was coded based on the number of presented locations (3 vs 1), with a distance of 0 between the conditions with the same number of locations and a distance of 2 between conditions with a different number of locations. We also examined a version of this spatial attention factor that was coded based on the number of relevant target locations (i.e., all set size 1 conditions, as well as all set size 3 with repeated locations, only had 1 target location), but that model did not explain as much unique variance as the pure spatial attention factor, nor did it change the results relating to the other factors. We also included 2 different modeling approaches to capture modality-specific signals. In one model, we coded only for the sensory modality, such that all conditions of a given modality had a distance of 0 between them, and all condition pairs that spanned the 2 modalities had a distance of 1. This model was analogous to the attended feature models used in experiment 1. In contrast, we also tested a model which examined modality-specific load signals. This consisted of 2 RDMs, on per modality, using the following logic. Set size 1 trials for a given modality (e.g. visual) would have a modality-specific load of 1, set size 2 trials of that modality would have a modality-specific load of 2, and any set size of the other modality (auditory set size 1 or 2 in this example) would have a modality-specific load of 0.

Prior to running regression analyses, we calculated the variance inflation factor (VIF) for each empirical model to ensure that estimates of each theoretical model's individual contribution were reliable and not potentially deflated; in all cases, each theoretical RDM's VIF was <3, indicating low multicollinearity between variables. For each factor, the semipartial correlation was calculated as the square root of the difference between the full model's R^2 and the R^2 value of a sub-model with this factor removed, multiplied by the sign of the factor's β in the full model. This was done for each subject. We used wilcoxon tests to evaluate whether semipartial correlations were significant at the group level, while applying the Benjamini-Hochberg procedure when testing over multiple timepoints. The presence of a statistically significant semipartial correlation can provide evidence for the unique contribution of a given theoretical factor to the neural signal.

To visualize the representational structure of the conditions, we applied metric multidimensional scaling (MDS) to the distance matrices calculated for RSA. We used both 2-D and 3-D projections based on experimental conditions (2-D for experiment 2's 2-level design and 3-D for experiment 3's 3-level design). We first averaged the RDM over all subjects and delay period timepoints, and then used the metric MDS class implemented by scikit-learn to calculate the appropriate projection.