

Multivariate analysis reveals a generalizable human electrophysiological signature of working memory load

Kirsten C. S. Adam^{1,2}  | Edward K. Vogel^{3,4,5} | Edward Awh^{3,4,5}

¹Department of Psychology, University of California San Diego, La Jolla, CA, USA

²Institute for Neural Computation, University of California San Diego, La Jolla, CA, USA

³Grossman Institute for Neuroscience, Quantitative Biology, and Human Behavior, University of Chicago, Chicago, IL, USA

⁴Department of Psychology, University of Chicago, Chicago, IL, USA

⁵Institute for Mind and Biology, University of Chicago, Chicago, IL, USA

Correspondence

Kirsten C. S. Adam, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109, USA.
Email: kadam@ucsd.edu

Funding information

Research was supported by National Institute of Mental Health grant 5R01-MH087214 (to E.A. & E.V.) and Office of Naval Research grant N00014-12-1-0972 (to E.V.). K.A. was supported by National Institute of Mental Health grant 5T32-MH020002

Abstract

Working memory (WM) is an online memory system that is critical for holding information in a rapidly accessible state during ongoing cognitive processing. Thus, there is strong value in methods that provide a temporally resolved index of WM load. While univariate EEG signals have been identified that vary with WM load, recent advances in multivariate analytic approaches suggest that there may be rich sources of information that do not generate reliable univariate signatures. Here, using data from four published studies ($n = 286$ and $>250,000$ trials), we demonstrate that multivariate analysis of EEG voltage topography provides a sensitive index of the number of items stored in WM that generalizes to novel human observers. Moreover, multivariate load detection (“mvLoad”) can provide robust information at the single-trial level, exceeding the sensitivity of extant univariate approaches. We show that this method tracks WM load in a manner that is (1) independent of the spatial position of the memoranda, (2) precise enough to differentiate item-by-item increments in the number of stored items, (3) generalizable across distinct tasks and stimulus displays, and (4) correlated with individual differences in WM behavior. Thus, this approach provides a powerful complement to univariate analytic approaches, enabling temporally resolved tracking of online memory storage in humans.

KEY WORDS

classification, EEG, multivariate pattern analysis, working memory

1 | INTRODUCTION

Working memory serves as a critical interface between perception, memory, and action. Given the critical role of working memory in complex cognition, much prior work has been dedicated to identifying measures of the human electroencephalogram (EEG) signal that track working memory load in near real-time. The most widely used of these measures is the contralateral delay activity (CDA) (Vogel & Machizawa, 2004; Vogel, McColough, & Machizawa, 2005), though other univariate measures such as suppressed alpha power and a sustained negative slow-wave have also been identified (Fukuda, Kang, & Woodman, 2016; Fukuda, Mance, & Vogel, 2015).

Tracking online working memory storage is critical for testing working memory's role as an interface between varied cognitive demands. Studies taking advantage of univariate measures like the contralateral delay activity (CDA) have demonstrated filtering within working memory (Vogel et al., 2005), the role of working memory in guiding visual search (Carlisle, Arita, Pardo, & Woodman, 2011; Emrich, Al-Aidroos, Pratt, & Ferber, 2009; Olivers, Peters, Houtkamp, & Roelfsema, 2011; Woodman & Arita, 2011), the role of working memory in buffering retrieval from long-term memory (Fukuda & Woodman, 2017), and the role of existing long-term memories in shaping working memory encoding (Xie & Zhang, 2018). In cases where behavior is equivocal, neural measures are critical for disentangling

competing explanations of an observed behavioral pattern. For example, unobtrusively monitoring the CDA during a typical visual search task revealed that search templates initially held working memory are moved to long-term memory with experience (Carlisle et al., 2011). Given the clear utility of near real-time measures of working memory load, we present a novel analytic approach that provides a strong leap forward in the search for more sensitive and precise measures of storage in working memory.

Although univariate measures have been very productive for tackling many important questions about how and when working memory resources are deployed, they may miss some important aspects of the memory signal. For example, in the domain of spatial attention, it has long been known that lateralized, univariate changes to alpha power (i.e., contralateral alpha suppression) can be used to track attention to the left versus right hemifield, and that the topography of alpha power is modulated by finer-grained manipulations of spatial position (e.g., Rihs, Michel, & Thut, 2007). In this context, multivariate analysis of alpha topography has been shown to provide a spatially and temporally resolved index of covert attention that substantially improves the utility of this signal for covert tracking of spatial attention (Foster, Sutterer, Serences, Vogel, & Awh, 2017). In addition to expanding the utility of known univariate measures, multivariate tools allow us to track information that was previously opaque to univariate analysis. For example, recent work has shown that motion direction of a dot cloud (Bae & Luck, 2019a) and a single remembered orientation (Bae & Luck, 2018, 2019b; Wolff, Ding, Myers, & Stokes, 2015; Wolff, Jochim, Akyürek, & Stokes, 2017) can be decoded from the topography of event-related potentials (ERPs; time-locked to stimulus or memory array onset), despite the absence of clear univariate signals that track this information.

Here, we show that a similar multivariate approach (“Multivariate load detection,” or “mvLoad”) enables tracking of online memory load in a sensitive and temporally resolved fashion. Note, throughout the article we define working memory load as the increasing amount of information held in mind with increasing memory set size. Although there is an ongoing debate about the format of mnemonic representations (e.g., item-based vs. a flexible resource;

Bays, 2018; Hakim, Adam, Gunseli, Awh, & Vogel, 2019), we do not directly address this issue here. In three experiments, we demonstrate that we can predict working memory set size from ERPs of small groups of trials (time-locked to memory array onset) and even with single trials of EEG data. Further analyses demonstrate that this multivariate decoding signal has the expected profile of a working memory signal (e.g., modulated by working memory task demands; shows higher confusability for supra-capacity working memory loads) and carries promise for future cross-subject, cross-experiment decoding applications.

2 | METHOD

2.1 | Overview of data sets

We used four previously published data sets to examine whether and why we can decode working memory load from the EEG signal. Further methodological details about the participants, tasks, and data acquisition can be found in each of the original published articles (Fukuda, Mance, et al., 2015; Fukuda, Woodman, & Vogel, 2015; Hakim et al., 2019; Unsworth, Fukuda, Awh, & Vogel, 2014, 2015). Some high-level information about the studies is provided in Table 1 and below, and additional information about preprocessing of the EEG data are available in the Supplemental Methods. Figure 1 shows examples of the stimuli and task procedures. Data for all experiments are freely available the Open Science Framework at <https://osf.io/6jkqu/>.

2.1.1 | Experiment 1

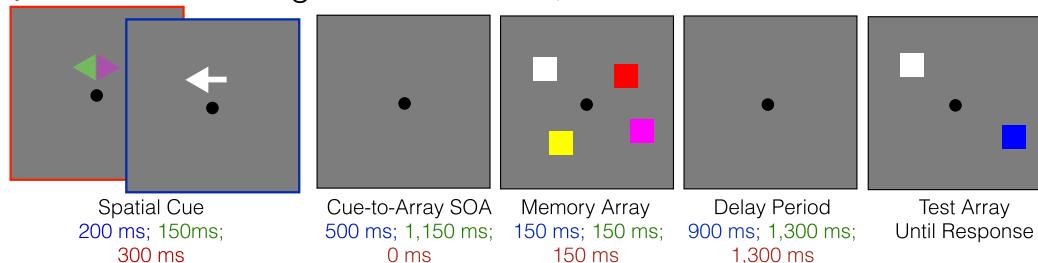
Participants performed two conditions of a lateralized change detection task (Unsworth, Fukuda, Awh, & Vogel, 2014, 2015); EEG data were collected from 20 passive electrodes. In one condition, participants remembered the colors of items (color change detection; set size 2 and 6). In the other condition, participants remembered the shapes of items (shape change detection, set size 2 and 6). Here, we collapse across the color and shape conditions (i.e., just examining set size 2

TABLE 1 Overview of data sets

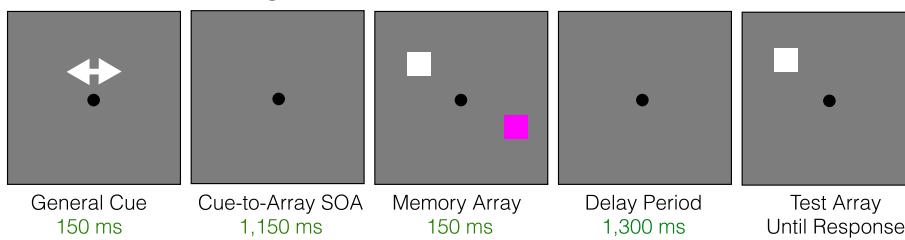
Experiment	Article	Task	Set sizes	Total trials	Subjects (n)
1	Unsworth et al. (2014, 2015)	Lateralized change detection	2, 6	116,101	152
2a	Fukuda, Mance, et al. (2015)	Lateralized and Whole-field change detection	1–4, 6, 8	23,951	30
2b	Fukuda, Woodman, et al. (2015)	Lateralized change detection	1–8	49,916	31
3	Hakim et al. (2019)	Lateralized change detection; Lateralized attention task	2, 4	102,358	73 ^a

^aSeventy-three unique subjects participated in one or more of the four sub-experiments (total of 97 sessions).

(a) Lateralized change detection task, Set Size 2



(b) Whole-field change detection task, Set Size 2



(c) Attention task, Target Present trial

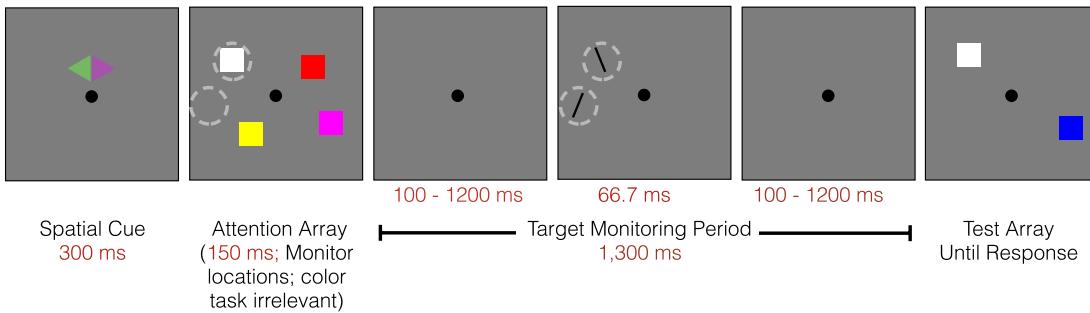


FIGURE 1 Task schematics for a typical Set Size 2 trial. (a) Lateralized change detection task used in Experiments 1, 2a, 2b, and 3. Participants are first cued to one side of space (Red cue box shows symbolic cue used in Experiment 3, for example, attend green side; Blue cue box shows spatial cue used in other Experiments). Participants remember the items on the cued side and are tested on one of the items. Blue text shows task timing for Experiments 1 and 2b; Green text shows timing for Experiment 2a; Red text shows task timing for Experiment 3. (b) Whole-field change detection task used in Experiment 2a. Participants are given a non-spatial cue that the trial is upcoming, remember all items on the screen across a delay, and are tested on one of the items. (c) Attention task used in Experiment 3. This task serves as a control for working memory task demands. Participants are cued to one side of the display. Rather than remember the colors of the squares, participants are asked to pay attention to the spatial positions occupied by the squares. They monitor these positions across a target monitoring period. A target (tilted line at the same location as one of the item positions) and/or distractor (tilted line at a foil location where no item appeared) appear on 25% of trials. For illustrative purposes only, the locations of the target and distractor lines are circled. The colored squares shown in the test array are task-irrelevant (visual control for WM condition); the array indicates when participants should make their response about the target

vs. 6 overall). See, Figure S1 for confirmation that the general results hold for separately analyzed color and shape conditions. The full data set includes 183 participants, with 200 trials per sub-condition (e.g., “shape set size 2”). Participants were analyzed if they had at minimum 160 trials per sub-condition after artifact rejection, leaving 152 subjects for analysis ($M = 382$ trials per set size after collapsing across color and shape).

2.1.2 | Experiment 2a

Participants completed two conditions of a change detection task: lateralized change detection and whole-field

change detection (Fukuda, Mance, et al., 2015); EEG data were collected from 20 passive electrodes. Participants ($n = 29$) completed 80 trials per set size (1–4, 6, and 8) in each condition ($M = 68.8$ trials per set size in each task condition).

2.1.3 | Experiment 2b

Participants ($n = 31$) completed a lateralized change detection task (Fukuda, Woodman, et al., 2015); EEG data were collected from 20 passive electrodes. There were eight set size conditions (1–8; $M = 201.3$ trials per set size).

2.1.4 | Experiment 3

The full data set includes 97 sessions from four sub-experiments. Some additional sessions were collected but excluded according to the artifact rejection criteria detailed in the original paper (Hakim et al., 2019; EEG data were collected from 32 active electrodes. In published work, we demonstrated no major differences between these four sub-experiments and we performed analyses collapsed across all four sub-experiments (Hakim et al., 2019). We likewise combined data across all four sub-experiments here. All participants completed two key task conditions: (1) a lateralized working memory task, and (2) a lateralized attention task. In some sub-experiments, the lateralized working memory task employed color memoranda and in other cases the lateralized working memory task employed spatial memoranda. In all cases, there were two set sizes (2 or 4 items) in each task ($M = 263.8$ trials per set size in each condition). Although participants could complete each sub-experiment only once, some participants completed multiple sub-experiments. This resulted in a total of 73 unique subjects. If a participant completed more than one sub-experiment, we averaged their behavioral and classification results across sessions so that each participant was equally represented in the full data set.

2.2 | Tasks

2.2.1 | Lateralized change detection

On each trial, participants were first cued to attend one hemi-field (left or right) with a brief spatial cue. After the cue, there was a blank interval (“cue-to-array SOA”), and then, the memory array appeared. The memory array consisted of brightly colored squares drawn in both hemifields. Participants were instructed to remember only the colored squares in the cued hemi-field across a blank delay period. At test, a single colored square was presented at one of the remembered locations. On 50% of trials (“same trials”), the test square was the same color as the item presented at that position. On the other 50% of trials (“change trials”), the test square was a different color from before. Participants pressed one of two keys to indicate whether the test square was the same color or had changed colors. The exact task timing varied slightly across data sets. Experiments 1 and 2b used a cue duration of 200 ms, cue-to-array SOA of 500 ms, memory array duration of 150 ms, and a delay period of 900 ms. Experiment 2a used a cue duration of 150 ms, cue-to-array SOA of 1,150 ms, memory array duration of 150 ms, and a delay period of 1,150 ms. The working memory condition from Experiment 3 used a cue duration of 300 ms, cue-to-array SOA of 0 ms, array presentation of 150 ms, delay period of 1,300 ms, and a blank inter-trial interval of 750 ms. Experiment

3 is a combination of four sub-experiments reported in Hakim et al., 2019. The task events and timing were consistent across all four sub-experiments. In two sub-experiments participants remembered color (as described); in the other two sub-experiments, participants remembered the spatial position of items, and were tested with an item that was either at the same location (“same trial”) or with an item that appeared at a foil location a minimum of 1.5 objects’-width away from any of the remembered locations (“different trial”). Prior work revealed that these stimulus-specific differences did not greatly alter the CDA, and that it was justified to collapse across these sub-experiments for further analysis (Hakim et al., 2019).

2.2.2 | Whole-field change detection

Experiment 2a used a whole-field version of the change detection task. This task was very similar to the lateralized change detection task, except there was no spatial cue. Instead, participants received a task-general cue (e.g., a double-sided arrow that did not indicate a side to attend but gave a temporal warning that the memory array was coming). Participants remembered all items from the array, and to-be-remembered items were presented in both the left and right hemifields (i.e., “whole-field”). As before, participants remembered colors across a delay and were probed on one item, and they reported whether the probed item was the same as or different from the remembered item. Experiment 2a used a cue duration of 150 ms, cue-to-array SOA of 1,150 ms, memory array duration of 150 ms, and a delay period of 1,150 ms.

2.2.3 | Lateralized attention task

In Experiment 3, the relative need for working memory task demands was manipulated. Participants viewed identical stimuli as in the lateralized change detection task described above, but they were given different task instructions which could be achieved with sustained spatial attention. This task used identical stimuli and task timing as the lateralized change detection task, but we asked participants to perform a sustained attention task rather than a working memory task. This allowed us to compare mental processes for “attention” and “working memory” tasks while holding visual stimulation and task timing constant. In prior work (Hakim et al., 2019), we showed that the CDA was present in the working lateralized change detection task but not in the lateralized attention task, indicating that the CDA is associated with working memory task demands. In the task, participants were first cued to one visual hemi-field (e.g., attend the hemi-field indicated by the green side of a double-sided arrow; colors counterbalanced across participants). Rather than remembering the colors and/or locations of the items in the

“memory array” (e.g., array of colored squares), participants were instead instructed to maintain their spatial attention to the positions occupied by the items. Participants maintained spatial attention to the positions during a blank array in order to detect and discriminate a rare target (titled line) that briefly appeared (66.67 ms) at one of the attended positions on 25% of trials. Note, there was always one target (line that occupied one of the same spatial positions that was cued). In addition, in three of four sub-experiments a distractor line was shown at a “foil” location.

During the attention task, the color and position of the item in the “test array” were task-irrelevant. Instead, the test array simply indicated the time when participants should make their response. The participants made one of three button presses: (1) target absent, (2) target present, top tilted left, and (3) target present, top tilted right. Note, we discarded the 25% of target-present trials from EEG analyses to avoid any potential physical display confounds. We analyzed only the 75% of trials where there was a fully blank delay period. As expected, in the working memory task participants stored slightly more items for set size 4 ($M = 1.73$) versus set size 2 trials ($M = 1.52$; $p = .003$). Likewise, participants had poorer performance in the attention task when they monitored four locations ($M = 78\%$ correct) versus two locations ($M = 82\%$ correct; $p = .01$), see, Hakim et al. (2019) for further discussion.

2.3 | EEG data acquisition

2.3.1 | Experiments 1–2

Experiments 1, 2a, and 2b were collected from 20 passive tin electrodes (SA Instrumentation Co., San Diego, CA) mounted in an elastic cap (ElectroCap International, Eaton, OH). Electrode positions included International 10/20 sites F3, Fz, F4, T3, C3, Cz, C4, T4, P3, Pz, P4, T5, T6, O1, and O2 and five nonstandard sites: OL midway between T5 and O1, OR midway between T6 and O2, PO3 midway between P3 and OL, PO4 midway between P4 and OR, POz midway between PO3 and PO4. Data were recorded with a left-mastoid reference and re-referenced offline to the algebraic average of the left and right mastoid. Horizontal electrooculogram (EOG) and vertical EOG were collected from three additional passive electrodes affixed to the face with stickers. Trials containing ocular artifacts, movement artifacts, or blocking were excluded from analyses.

2.3.2 | Experiment 3

Experiment 3 was collected from 30 active Ag/AgCl electrodes (actiCHamp, Brain Products, Munich Germany)

mounted in an elastic cap positioned according to the international 10–20 system (Fp1, Fp2, F7, F8, F3, F4, Fz, FC5, FC6, FC1, FC2, C3, C4, Cz, CP5, CP6, CP1, CP2, P7, P8, P3, P4, Pz, PO7, PO8, PO3, PO4, O1, O2, Oz). Two additional active electrodes were affixed with stickers to the left and right mastoids, and a ground electrode was placed at position Fpz. Data were referenced online to the right mastoid and re-referenced offline to the algebraic average of the left and right mastoids. Passive electrodes (HEOG, VEOG) and eye tracking were used to monitor eye movements and blinks. Trials containing ocular artifacts, movement artifacts, or blocking were excluded from analyses.

2.4 | Classification and significance testing

2.4.1 | Single-trial classification (within-subject)

Classification was performed within a subject, on single trials, and within a given time window on raw, baselined EEG data (Experiment 1). We divided each trial into 50-ms windows and calculated the average voltage for each electrode within this window (e.g., 20 electrodes = 20 predictors). Classification was performed separately within each time point using a linear discriminant classifier (“classify.m,” with option “diagLinear” to use the diagonal covariance matrix estimate; MATLAB, MathWorks, Natick MA). We chose to use a relatively simple linear classifier to first demonstrate our effect, but our results are expected to also generalize to other classification methods such as support vector machines (SVM; Figure S2). We performed 100 iterations of the classification analysis at each point; on each iteration, we randomly assigned 2/3 of the trials to an independent training set and 1/3 of the trials to a held-out test set. We also confirmed that the trial voltage distributions on each iteration of the analysis were approximately normally distributed (Figure S3). A schematic of the classification procedure is shown in Figure 2a. In all classification procedures, we balanced the number of trials per set size in both the training and test sets.

2.4.2 | Mini-block classification (within-subject)

Instead of performing classification on single trials, we averaged together groups of like trials (i.e., the same set size condition) into “mini-blocks,” and we shuffled the assignment of trials to different mini-blocks across 100 iterations (Experiments 2 and 3). Classification was still performed using the same linear classification routine, training on 2/3 of mini-blocks and testing on a held-out 1/3 of mini-blocks

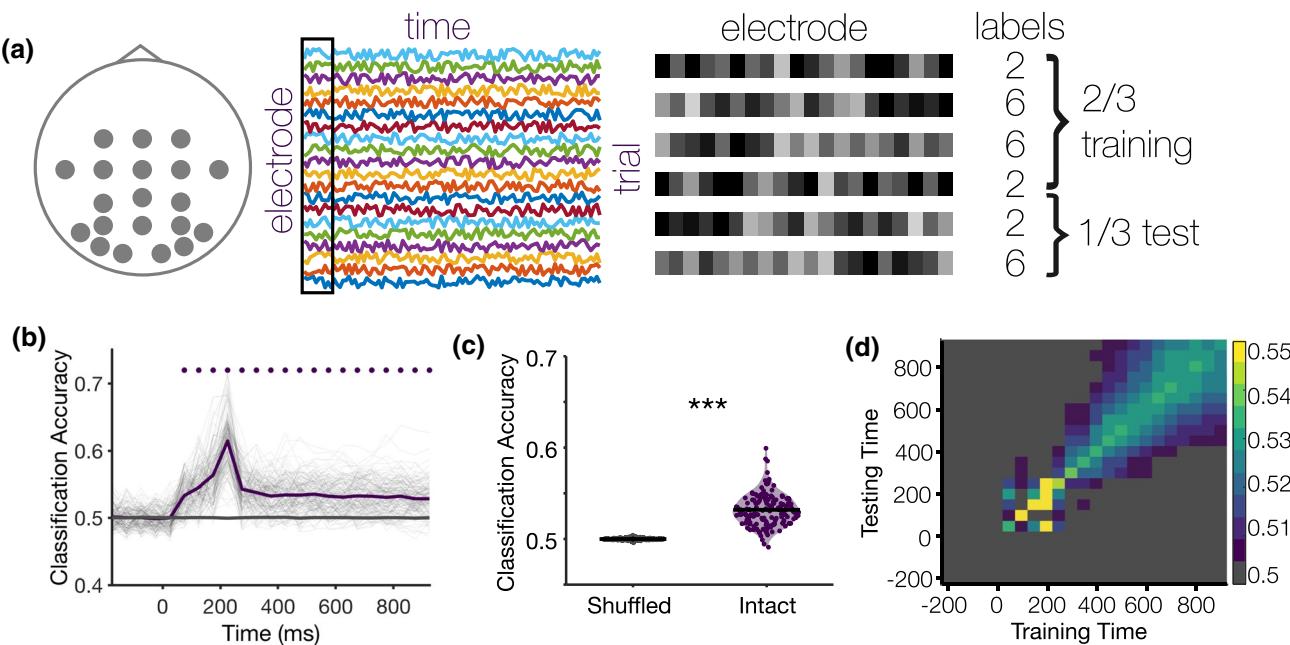


FIGURE 2 Single-trial decoding of working memory set size in Experiment 1. (a) Schematic of single-trial decoding approach. Decoding was performed within a participant, separately at each time bin (1 average value per electrode for a 50 ms time bin) using data from all electrodes (Data: t trials \times 20 electrodes; Labels: 1 \times t trials). On each iteration of the analysis, we picked a random 2/3 of trials to serve as a training data set, and the remaining 1/3 of trials were used as a testing data set. (b) Single trial decoding performance over time. Expected chance is 50%; dots indicate Bonferroni-corrected $p < .001$. (c) Average single trial decoding performance during the delay period (400–1,000 ms; *** $p < .001$). (d) Cross-temporal generalization of classification performance (training and testing across different time bins in the trial). Gray indicates that the pixel did not survive the cluster-based permutation test

on each iteration. Also note, this classification function handles both binary and multi-class classification, so the same general classification methods were used for both Experiments 1 and 3 (binary) as well as Experiment 2 (multi-class). To assess whether classification generalized across tasks (Experiments 2 and 3), we used the same method except we trained on 2/3 of mini-blocks from one task (e.g., lateralized change detection) and tested on 1/3 of mini-blocks from the other task (e.g., whole-field change detection).

2.4.3 | Mini-block classification (across-subject)

To test the generalizability of the classification signal across subjects, we performed a leave-1-subject-out analysis (Experiment 2). To do so, we blocked trials in the same way as in the within-subjects version of the mini-block analysis. We trained the classifier on all data on a random 2/3 of mini-blocks from $n-1$ subjects and tested data on a random 1/3 of mini-blocks from one held-out subject. For each held-out subject, we ran 100 iterations of randomly assigning individual trials to mini-blocks in both the training and test sets.

2.4.4 | Statistical tests

In the standard within-subject classification analyses, we trained and tested on data from the same time bin (e.g., train and test on data averaged from 0–50 ms). Significance of the time-course of overall classification was assessed via Bonferroni-corrected t tests (one-sided t tests when comparing to chance level, as we would not expect to find meaningfully below-chance values; two-sided tests when comparing between conditions). To assess the generalizability of the signal across time points, we also performed a cross-temporal analysis where we trained and tested on all possible combinations of time-points (e.g., train on the first time point, test on all other time points). Significance of the generalizability of decoding was assessed via a cluster-based permutation test statistic, based on comparing significant clusters of adjacent significant t -values to a permuted distribution (Maris & Oostenveld, 2007), using a subject-wise permutation function (1,000 iterations) adapted from Fahrenfort, van Driel, van Gaal, and Olivers (2018). Significance was always assessed in comparison to empirical chance values (rather than to theoretical chance, see, Combrisson & Jerbi, 2015); empirical chance was estimated by repeating the same classification analysis using randomly shuffled training labels.

3 | RESULTS

3.1 | Experiment 1

3.1.1 | Single-trial classification predicts working memory load in a large sample

Using a large sample ($n = 152$, $trials = 116,101$) and a linear classifier on raw EEG amplitudes from single trials (Figure 2a), we could predict working memory load (set size 2 vs. set size 6) in a sustained fashion throughout the delay period (Figure 2b; dots indicate $p < 1 \times 10^{-5}$, Bonferroni-corrected for 23 time-bins). We could classify set size quite early in the trial (50–100 ms time bin), although this early classification could be due to a physical display difference between the two- and six-item arrays. Importantly, classification was sustained throughout the delay in the absence of any physical display differences. Mean decoding accuracy during the delay period (400–950 ms) was 53.2% ($SD = 1.7\%$; Figure 2c), significantly above the chance level produced by giving the same classifier shuffled labels, $t(151) = 22.5$, $p < 1 \times 10^{-49}$, 95% CI [2.92%, 3.48%], Cohen's $d = 1.85$. Finally, decoding generalized to other time-points beginning at the 400–450 ms time bin and lasting throughout the delay (Figure 2d; gray boxes indicate that the pixel did not survive a cluster-based permutation test). Notably, the decoding signal observed during encoding (100–300 ms) did not generalize throughout the delay period, indicating that it is unlikely that a sensory imbalance signal early in the trial drove the sustained delay period decoding.

3.1.2 | Global versus lateralized contributions to decoding

Although our analysis shows that the topography of voltage values across electrodes predicted working memory load on a single trial basis, this simple decoding approach was blind to *lateralized* EEG signals that track working memory storage such as the CDA, a well-documented electrophysiological marker of storage in visual working memory. Thus, the robust performance of our decoder could not be explained by contributions from CDA activity. Nevertheless, this leaves open the interesting question of whether load detection could be further improved by taking lateralized storage signals into account. To test this, we added eight new predictors corresponding to single-trial paired difference waves (contralateral minus ipsilateral, e.g., PO8 minus PO7 for a “remember left” trial, PO7 minus PO8 for a “remember right” trial). Surprisingly, we found that adding lateralized predictors did not predict substantial additional variance beyond the single electrodes (Figure 3a). On its own, activity from just the eight lateralized predictors likewise tracked working memory load in a

sustained fashion throughout the delay period ($M = 52.9\%$, $SD = 1.7\%$, $p < 1 \times 10^{-16}$, $d = 1.67$), in line with many past demonstrations in the CDA literature. Including both single electrodes and the eight lateralized predictors somewhat improved classification performance beyond the lateralized predictors alone (53.1% combined vs. 52.9% lateralized alone, $p < .001$). Critically, however, a classifier combining both the eight lateralized predictors and the 20 single electrode predictors did slightly *worse* than the 20 single electrode predictors alone (53.1% combined vs. 53.2% single electrodes alone, $p < .001$). Further control analyses revealed that training and testing within a cued side (e.g., train and test on “remember right” trials) offered only marginal benefits over training and testing across cued sides (e.g., train on “remember right” test on “remember left”), see, Figure S4.

Although it is somewhat surprising that adding lateralized predictors did not improve decoding, topographic plots for each set size condition suggest that this failure to explain additional variance may be due to the coarse spatial distribution of the signal (Figure 3b) and/or to noisiness of using difference waves as predictors on a single trial basis. As adding the lateralized predictors failed to improve classification performance, we will continue to use the single electrode classifier (i.e., giving the classifier the raw voltage value at each electrode). Arguably, insensitivity to stimulus laterality makes the analysis more flexible and powerful, providing the potential to train and test classifiers across lateralized and nonlateralized working memory tasks; we provide an example of such cross-training in Experiment 2A/B. In supplemental analyses, we examined decoding accuracy separately for individual electrodes and groups of electrodes (e.g., occipital alone, Figure S5), and we confirmed that decoding was not driven by a global signal alone (Figure S6). Together, the topography of voltage changes and the supplementary analyses suggest that broadly distributed changes to voltage values (i.e., changes to the degree/extent of frontal positivity and changes to the degree/extent of posterior negativity) are likely driving overall decoding performance.

3.1.3 | Improving decoding with mini-block classification

Although the prior analyses demonstrated that single-trial classification is extremely robust during the delay period ($d = 1.85$), single trials are noisy and thus classification accuracy is numerically low (~53% given a chance level of 50%). To accommodate the lower numbers of trials and participants in some experiments, we tested whether averaging across small sub-sets of trials (“mini blocks”) would improve overall classification accuracy (Figure 4). Here, we performed the same basic analysis (training on 2/3 of data and testing on 1/3 of data, for 100 iterations of randomly assigning trials to

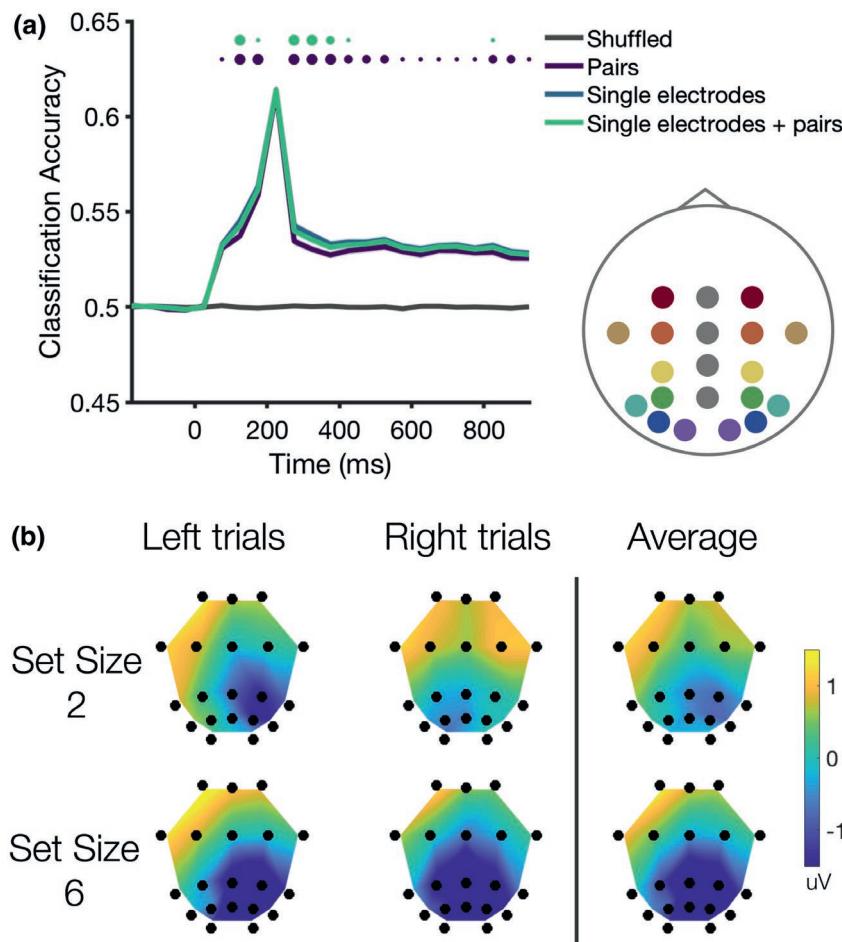


FIGURE 3 Adding lateralized predictors did not improve classification performance. (a) We created eight additional predictors by taking the difference (contra–ipsi) of all matched pairs of lateralized pairs (“pairs,” color-coded in the inset diagram). Lateralized predictors did a good job of predicting working memory load but did not improve classification above the level of the 20 single electrodes. In this figure, dots represent Bonferroni-corrected significance (small $p < .05$, medium $p < .01$, large $p < .001$). Purple dots represent the difference between single electrodes and pairs alone. Green dots represent the comparison between “single electrodes + pairs” and pairs alone. (b) Topographical plots of delay period activity for each set size condition, separated by side or averaged across both sides. Color scale is in microvolts (μ V)

training or testing). However, rather than training on single trials, we averaged small groups of trials together to reduce noise of each instance given to the classifier. This “mini-block” procedure was quite effective at improving overall classification accuracy, both early and late in the trial (Figure 2a). During the encoding period (100–300 ms), classification accuracy improved monotonically with the number of trials per mini-block, $F(1.1, 155.7)^1 = 2,615, p < 1 \times 10^{-106}$, $\eta_p^2 = .95$. In the peak sensory time bin (200–250 ms), decoding accuracy reached as high as 89.4% ($SD = 7.7\%$; 25 trial mini-blocks). During the delay period, classification accuracy likewise improved monotonically with the number of trials per mini-block, $F(1.03, 156.14) = 571.6, p < 1 \times 10^{-54}$, $\eta_p^2 = .79$, topping out at 63.9% ($SD = 7.1\%$) with 25 trial mini-blocks (Figure 2b). Note, however, the number of trials that may be used in mini-blocks is limited by the number of available trials per condition. For consistency of comparisons across experiments, we will use 10-trial mini-blocks for further analyses. This will allow us to improve classification accuracy while still accommodating the varied numbers of trials per experiment (~80–400 trials per set size).

¹Greenhouse–Geisser correction applied when the assumption of sphericity is violated.

3.2 | Experiment 2

3.2.1 | Decoding differentiates item-by-item increments in load

In Experiment 1, we demonstrated that we can discriminate between two memory load conditions (set size 2 vs. set size 6) using the multivariate EEG signal across electrodes (training and testing the classifier within a subject). In Experiment 2A and 2B, we tested whether this within-subject classification signal is sensitive to finer-grained set size manipulations. In Experiment 2A, participants performed two working memory tasks (lateralized and whole-field) with six set sizes (1–4, 6, and 8); In Experiment 2B, participants performed a lateralized working memory task with eight set sizes (1–8). As shown in Figure 5a, we could robustly predict set size in a sustained fashion throughout the delay period (all p 's $< 1 \times 10^{-8}$, Cohen's $d = 1.55, 1.49$, and 2.27 for panels left to right in Figure 5a). As in Experiment 1, this classification signal sustained through the end of the delay and generalized to all later time bins beginning mid-way through the delay (650–700 ms, 600–650 ms, and 600–550 ms, for the panels shown left to right in Figure 5b).

In addition to examining overall decoding, the inclusion of more set sizes in Experiments 2A and 2B provided the chance

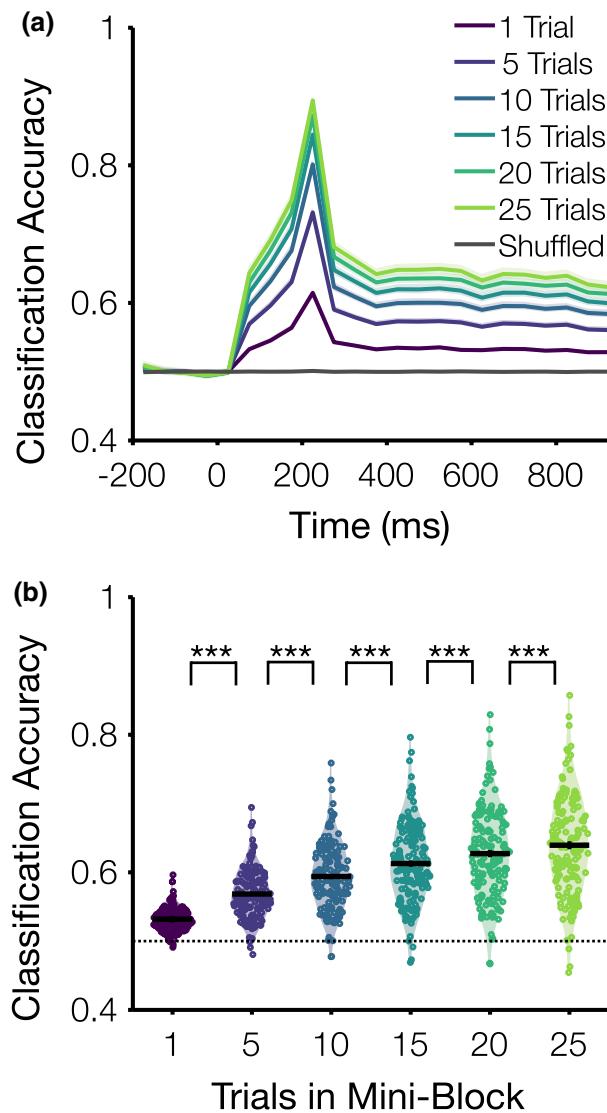


FIGURE 4 Decoding accuracy by number of trials in classification “mini-blocks.” We examined changes to decoding accuracy for training and testing on single trials versus training and testing on averaged values from small groups of trials (e.g., a “mini block” of five Set Size 2 trials). (a) Classifier accuracy over time for single trial decoding and mini-block decoding of various block sizes. Shaded error bars indicate ± 1 SEM. (b) Mean classification accuracy during the delay period (400 ms—end of delay). *** $p < .001$ (Bonferroni corrected, five comparisons)

to look at classifier errors and discriminability of set sizes. Figure 6a,b shows confusion matrices for the encoding period (100–300 ms) and delay period (400 ms to end of delay) in Experiments 2A and 2B. Much prior work on univariate neural signatures of working memory maintenance (e.g., Todd & Marois, 2004; Vogel & Machizawa, 2004; Xu & Chun, 2006) has found that, consistent with a capacity limit of 3–4 items, univariate measures increase from set sizes 1 to 3, but reach an asymptote around 3–4 items (but see Bays, 2018). Given this key signature of univariate working memory measures, we predicted that delay period decoding, but not sensory period

decoding, should show particularly poor or no discriminability among larger set sizes. To test this, we compared confusion matrix discriminability among the lower set sizes (1–3 in Experiment 2A, 1–4 in Experiment 2B) versus discriminability among higher set sizes (4,6,8 in Experiment 2A, 5–8 in Experiment 2B) in both the encoding period and delay period. Discriminability was quantified as the difference between the true category value (e.g., proportion of the times the classifier chose set size 1 when the true value was 1) and the mean of the incorrect values (e.g., how often the classifier instead chose other low set sizes 2 or 3). As we predicted for the delay period signal, we observed significantly higher decoding among low set sizes than high set sizes for all three experiments ($p < .001$) and we observed a null effect for discriminability among high set sizes during the delay in all three experiments ($p > .10$). We likewise observed higher discriminability among low set sizes than high set sizes during the sensory period (100–300 ms, all p 's < .01) but significant encoding period decoding for both high and low set sizes (all p 's < .001). For uncluttered visualization, here we have shown confusion matrices as color-scales; Please see, Figure S7 in the supplemental for numerical values in each cell of the confusion matrix. These effects was similar when we instead used many pairwise classifiers to test which set sizes were discriminable from one another during the delay period (Figure S8). As such, we found a general pattern of results that is consistent with a delay period working memory signal, which is expected to show higher confusability among supra-capacity set sizes. However, future work will be needed to examine changes to decoding while perfectly controlling for display-wise sensory differences. For example, it is unclear whether the poorer discriminability for high versus low set sizes during the encoding period (100–300 ms) was driven by bottom-up sensory differences or by attentional selection of items (e.g., a capacity limit in the number of selected items could contribute to the increased confusability for higher set sizes even during the early encoding period).

3.2.2 | Generalization of classification across subjects and tasks

Up to this point, we have always trained and tested the classifier within a given subject. Experiment 2 provided an opportunity to examine the generalizability of the classification signal across tasks, subjects, and experiments with distinct stimulus displays. In Experiment 2A, the same subjects performed two different working memory task variants (one lateralized with distractors presented in the irrelevant hemifield, one whole-field display that contained no distractors). Thus, Experiment 2A provided the opportunity to look at decoding within the same subject, but across distinct tasks (Figure 7a-b). Across-task decoding was overall significant during the delay period, indicating some degree of generalizability when training and

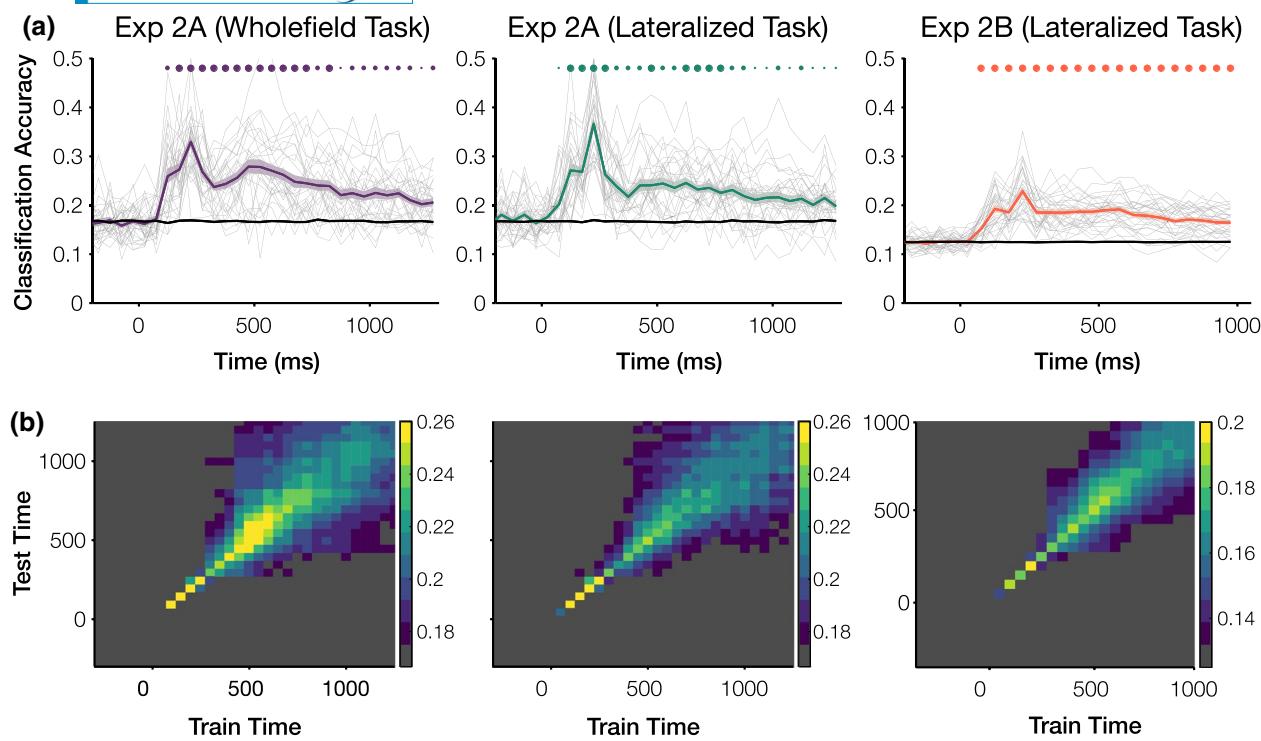


FIGURE 5 Decoding accuracy in Experiments 2A and 2B. (a) Classification accuracy over time for Experiments 2A and 2B. Chance level is 1/6 in Experiment 2A and 1/8 in Experiment 2B (black line depicts empirical chance from shuffled analysis). Shaded error bars represent ± 1 SEM. Transparent gray lines represent individual subjects. Dots indicate Bonferroni-corrected significance of each 50 ms time-bin (small dots, $p < .05$, medium, $p < .01$, large, $p < .001$). (b) Cross-temporal generalization of classification performance (training and testing across different time bins in the trial) for Experiments 2A and 2B. Gray indicates that the pixel did not survive the cluster-based permutation test

testing within- versus across tasks ($p < 1 \times 10^{-20}$). However, decoding accuracy was significantly lower when training and testing across tasks versus within a task ($p < 1 \times 10^{-9}$). We quantified the difference in classification accuracy between within- and across-task decoding across three task epochs: encoding (100–300 ms), early delay (400–900 ms), and late delay (900 ms to end of delay). We found a main effect of Training (better performance for within-vs. across tasks), $F(1,29) = 79.25$, $p < 1 \times 10^{-9}$, $\eta_p^2 = .73$, and of Epoch (better performance earlier in the trial), $F(2,58) = 31.49$, $p < 1 \times 10^{-9}$, $\eta_p^2 = .52$. We also found an interaction of Training and Epoch, $F(1.50,43.44) = 34.81$, $p < 1 \times 10^{-7}$, $\eta_p^2 = .55$, indicating that the late delay period signal was more robust to cross-generalization across task variants than the early sensory signal (i.e., during the late delay period there was no difference in classifier accuracy for within-vs. across-task training).

In Figure 7c-d, we examined the ability of the classification signal to generalize across subjects within Experiment 2A, showing that there is a generalizable multivariate signature of working memory load in humans. We trained the classifier on 2/3 of data from $n-1$ subjects, and tested the classifier on 1/3 of data from one held-out subject. For both the within- and across-subjects analyses, we performed the analysis separately for training within-a task versus across tasks (as above). Across-subject decoding was overall significant

during the delay period ($p < 1 \times 10^{-24}$). In Figure 7e-f, we have collapsed across this task dimension but it is depicted in Figure S7. Similar to generalizing across tasks, we found that found a main effect of Training (better performance for within-vs. across subjects), $F(1,29) = 45.33$, $p < 1 \times 10^{-6}$, $\eta_p^2 = .61$, and of Epoch (better performance earlier in the trial), $F(2,58) = 27.19$, $p < 1 \times 10^{-8}$, $\eta_p^2 = .48$. We again found an interaction of Training and Epoch, $F(2,58) = 44.94$, $p < 1 \times 10^{-11}$, $\eta_p^2 = .61$, indicating that the late delay period signal was more robust to cross-generalization across subjects than the early sensory signal.

Finally, in Figure 7e-f, we examined the ability of the classification signal to generalize across subjects from different experimental samples. These experiments were the same in some important ways (e.g., same electrode montage, sampling rate, and similar tasks), but differed in many minor ways (e.g., exact size of stimuli). We trained the classifier on 2/3 of data from $n-1$ subjects in one experiment (e.g., train Experiment 2A) and tested the classifier on 1/3 of data from one subject in the other experiment (e.g., test Experiment 2B). For both the within and across experiment analyses, we again performed the analysis separately for training within-versus across tasks. Across-experiment decoding accuracy was overall significant during the delay period ($p < 1 \times 10^{-29}$). In Figure 7e-f, we have collapsed across the task dimension but it is depicted in Figure S9. We found

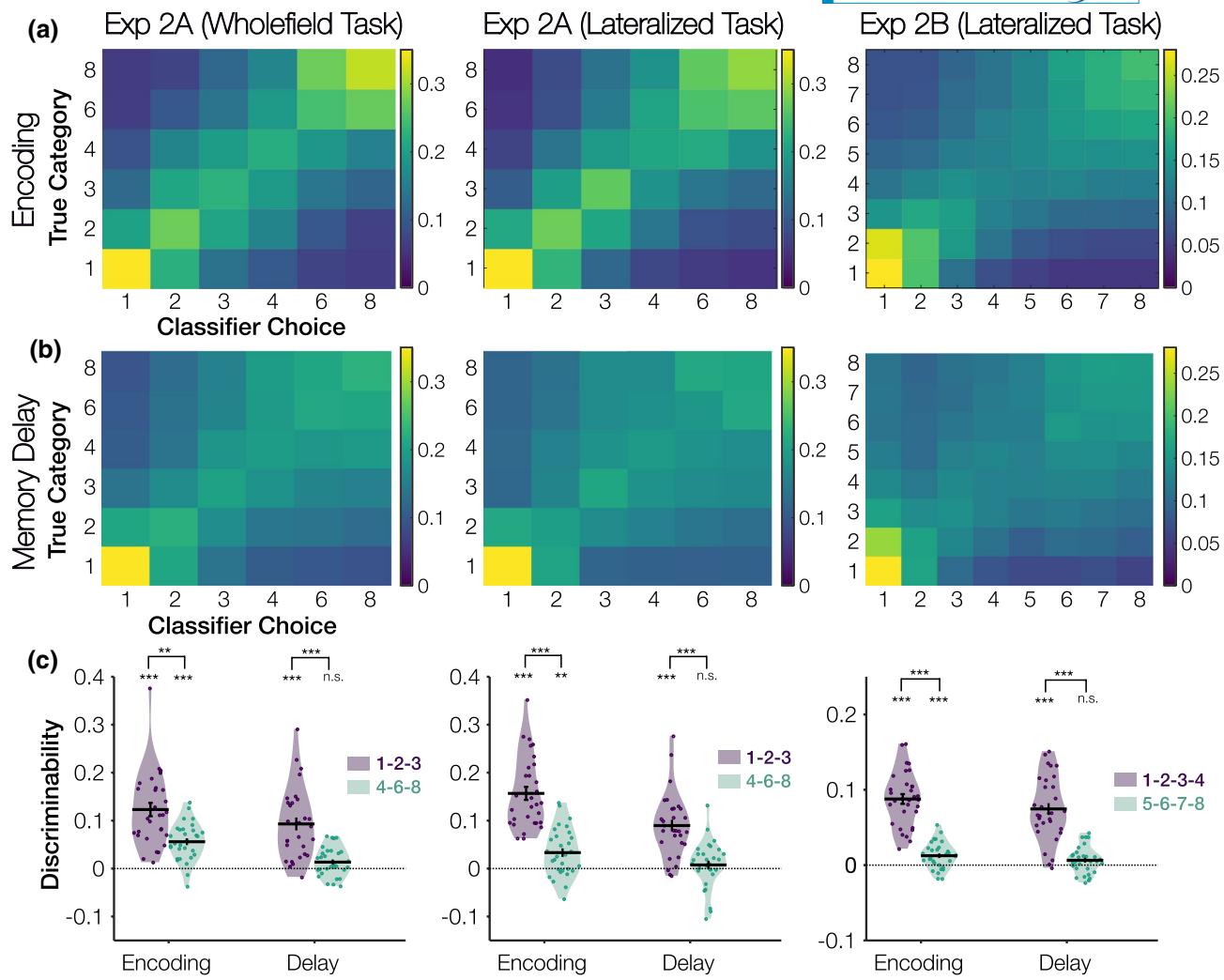


FIGURE 6 Confusion matrices and discriminability between set sizes. (a) Confusion matrix during the encoding period (100–300 ms). (b) Confusion matrix during the delay period (400 ms–end of delay). (c) Discriminability metric among lower versus higher set sizes during the encoding period and the delay period. Symbols indicate Bonferroni-corrected significance (six comparisons), n.s. $p > .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

that, relative to training across subjects but within a specific experiment, training across subjects and across experiments was slightly worse (Figure 7e-f), as indicated by a main effect of Training, $F(1,29) = 31.33$, $p < 1 \times 10^{-5}$, $\eta_p^2 = .52$. Here, we found no main effect of Epoch, $F(2,58) = 1.91$, $p = .16$, $\eta_p^2 = .06$, and no interaction of Training and Epoch, $F(2,58) = .60$, $p = .55$, $\eta_p^2 = .02$, indicating that the performance decrement for training across versus within experiments was consistent throughout the trial.

3.3 | Experiment 3

3.3.1 | Delay-period decoding is specific to working memory task demands

In all experiments discussed so far, the amount of sensory stimulation was confounded with set size (i.e., there was

more sensory stimulation on higher set size trials). Although our analyses suggest that sustained delay period decoding likely was not driven by this transient sensory confound (e.g., decoding of set sizes generalized among time points within the delay period, but decoding during the stimulus period did not generalize to the delay period), we wanted to test whether delay-period decoding is modulated by working memory task demands while holding visual stimulation constant. To do so, we examined data from Hakim et al. (2019). In this experiment, participants performed two different cognitive tasks using visually identical stimuli. In one condition (“attention”) participants performed a spatial attention task. Prior work has shown that this condition did not recruit neural signatures of working memory maintenance (i.e., CDA was absent). In the other condition (“working memory,” WM), participants performed a typical working memory task, and robust signatures of working memory maintenance were observed. This experiment thus provides a critical test of whether delay

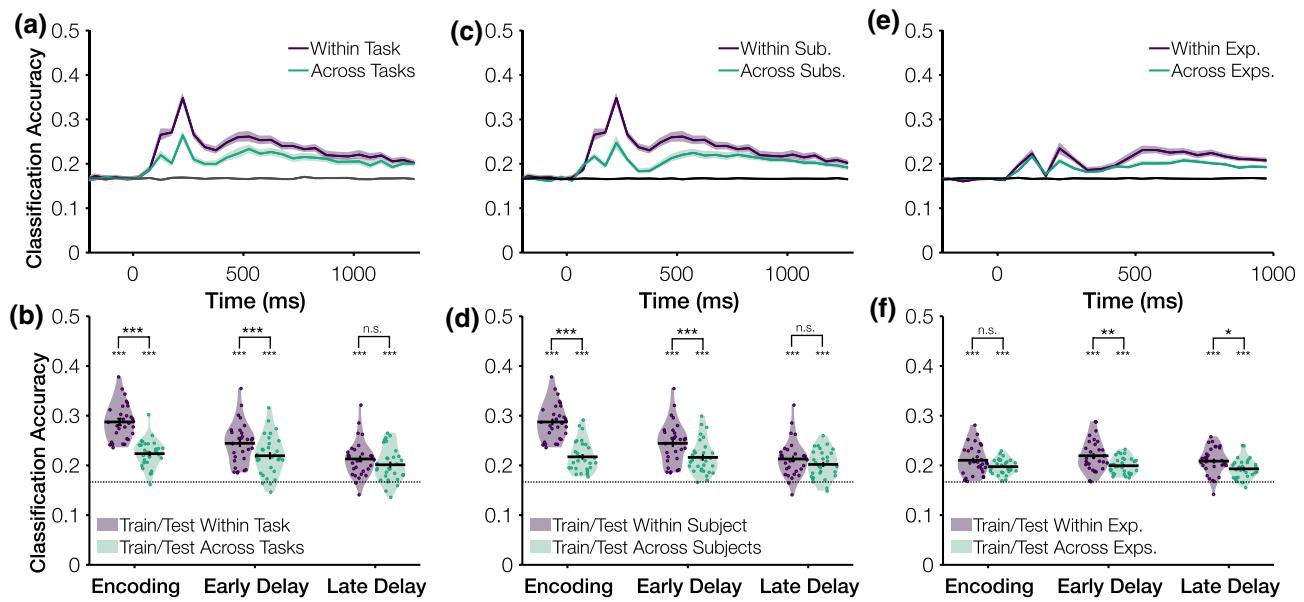


FIGURE 7 Classifier success at generalizing across task variants and across subjects. (a-b) Time-course and violin plot summaries for training/testing within a task versus across tasks (within-subjects, e.g., train on lateralized task, test on whole-field task). Shaded errors bars represent ± 1 SEM. (c-d) Time-course and violin plot summaries for training/testing within-subjects versus across subjects (averaged within-/across tasks) (e-f) Time-course and violin plot summaries for training/testing within an experiment versus across Experiments 2A and 2B (across subjects; average of within-/across tasks)

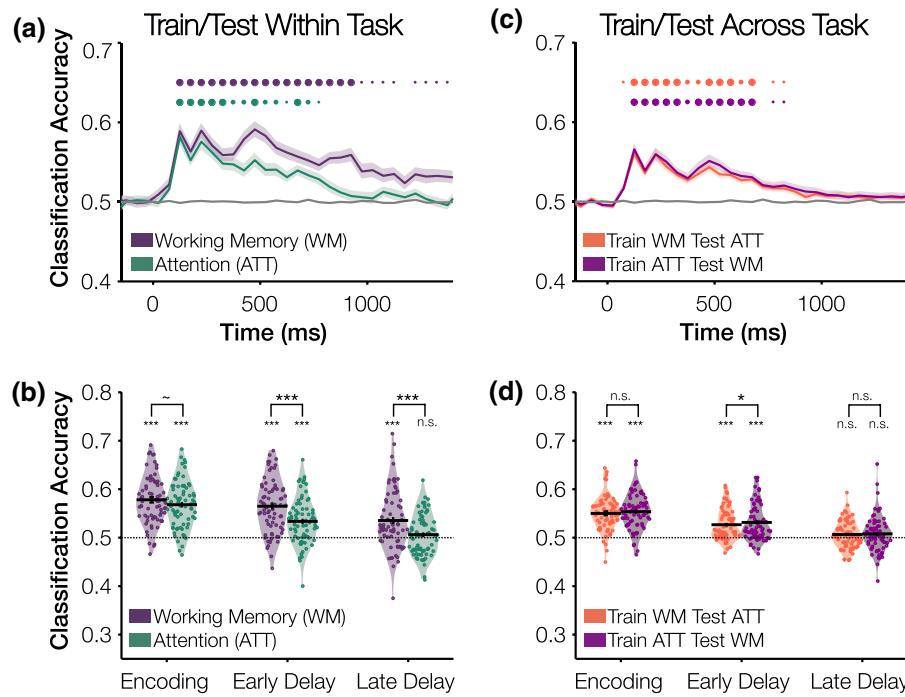
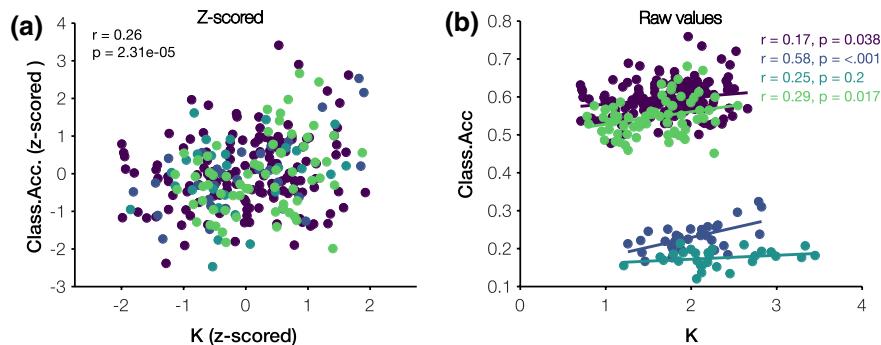


FIGURE 8 Decoding within- and across tasks in Experiment 3. (a) Classification time-course for the working memory task and the lateralized attention task. Dots indicate one-tailed t tests with Bonferroni-corrected significance (23 timepoints \times 2 conditions = 46 comparisons), small dots, $p < .05$, medium $p < .01$, large, $p < .001$. (b) Average classification within tasks during encoding, early delay, and late delay. Stars indicate Bonferroni-corrected significance (nine comparisons), n.s. $p > .10$, ~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$. (c) Classification time-course training and testing across tasks (e.g., train on attention task, test on working memory task). (d) Average classification across tasks during encoding, early delay, and late delay

FIGURE 9 Individual differences in working memory performance and decoding accuracy. (a) Correlation between capacity and classification accuracy (z-scored) for all unique subjects in all experiments. (b) Correlations between capacity and classification accuracy (raw values) for individual experiments



period decoding of set size respects the relative recruitment of working memory task demands while holding the sensory confound constant (i.e., the difference in sensory stimulation between set sizes 2 and 4 is identical for the working memory and attention task conditions).

Consistent with a signature of working memory maintenance, we observed sustained decoding of set size throughout the delay period when participants performed the working memory task, but not when they performed the spatial attention task (Figure 8a). To quantify this effect, we again divided data into the same three task epochs (encoding, early delay, and late delay). A repeated measures ANOVA with within-subjects factors epoch and task revealed a main effect of task, $F(1.80, 129.68) = 47.79, p < 1 \times 10^{-14}, \eta_p^2 = .40$, a main effect of epoch, $F(1, 72) = 25.90, p < 1 \times 10^{-5}, \eta_p^2 = .27$, and an interaction of task and epoch, $F(2, 144) = 3.46, p = .034, \eta_p^2 = .05$, Figure 8b. This demonstrates that early in the trial, when the sensory confound likely contributed to decoding, we observed no difference in decoding strength between the two task conditions. However, during the delay period, decoding was significantly weaker in the attention condition and completely disappeared by the late delay period (Figure 8b). This pattern of results is consistent with sustained delay period decoding as being driven by working memory task demands. To show that the null effect in the attention condition was not driven by a relatively weaker training set, we performed classification while training and testing across tasks. If the lack of decoding in the attention condition was just due to the attention condition serving poorly as a training set, then training on the working memory task should rescue delay period decoding for the attention task. However, we found no evidence of sustained delay period decoding when training across tasks. Although we were initially able to discriminate between set sizes, this early classification dissipated by around 700 ms (Figure 8c). We again performed an ANOVA with factors epoch and task (train attention, test WM vs. train WM, test attention). We found a main effect of epoch, $F(1.74, 125.41) = 46.55, p < 1 \times 10^{-13}, \eta_p^2 = .39$, an effect of task, $F(1, 72) = 6.38, p = .014, \eta_p^2 = .08$ (slightly better overall decoding when training on the attention task, counter to the hypothetical explanation of poor attention decoding), and no interaction of task and

epoch, $F(2, 144) = 1.19, p = .31, \eta_p^2 = .02$ (Figure 8d). Thus, this analysis suggests that multivariate load detection is determined by storage in working memory rather than by the physical characteristics of the display, or the deployment of spatial attention alone.

3.4 | Cross-experiment analyses

3.4.1 | Individual differences in decoding predict overall behavioral performance

Finally, we combined data from all experiments (unique subjects only) to test whether classification performance relates to behavioral performance. We reasoned that participants with higher working memory capacity would have more states to discriminate between (e.g., we would expect someone with a capacity estimate of 1.0 items to show similar neural signatures on all trials, whereas someone with a higher capacity estimate would have greater variability in delay period signatures across set size conditions). As such, we predicted that those with higher working memory capacity would likewise have higher delay period classification accuracy. Within each experiment, we z-scored behavior (average capacity, or “K”) and average delay period classification accuracy. This step was necessary to normalize differences in chance level (e.g., 50% vs. 16%) and behavioral performance across experiments. Also note, we used the 10-trial “mini-block” data for all four experiments. We excluded participants with capacity values lower than two SD’s below the group mean (i.e., who were not performing the task as instructed). We then performed a correlation using all unique subjects from all experiments. We found that classification accuracy and capacity were correlated, $r = .26, p < 1 \times 10^{-4}, 95\% \text{ CI} = [.14, .36]$. Correlations of raw classification accuracy and behavior values are shown in Figure 9b. Note, the overall correlation values did not noticeably change if we included subjects with poor behavioral performance or if we included duplicate subjects (Figure S10). Likewise, the correlation between classification accuracy and was similar in size to previously observed correlations between CDA amplitude and behavior (Figure S11). Interestingly, however,

classification accuracy and CDA amplitude predicted unique behavioral variance (Analysis S1).

4 | DISCUSSION

Here, we used a large data set (four published experiments, $n = 286$, $>250,000$ trials) to develop *multivariate load detection* (“*mvLoad*”), a novel approach for tracking human working memory load using the raw EEG signal. Multivariate load detection offers several key advances over existing univariate working memory signals (e.g., contralateral delay activity, “CDA”). Most importantly, multivariate load detection is generalizable across stimulus/task differences, observers, and experiments, suggesting that it taps into a common human electrophysiological signature of working memory load. In the contexts analyzed here, it was possible to train the classifier on one set of observers, and then, examine the deployment of working memory resources in a new task and set of observers. As such, multivariate load detection will allow us to study working memory in new, more flexible task contexts (e.g., without relying on lateralized displays which incur a dual task of filtering out one hemifield) and applied settings (e.g., brain-computer interfaces).

The benefits of multivariate load detection mirror similar advances made in multivariate detection of the locus of spatial attention (Foster et al., 2017; Rihs et al., 2007), attentional selection (Fahrenfort, Grubert, Olivers, & Eimer, 2017; Munneke, Fahrenfort, Sutterer, Theeuwes, & Awh, 2019), and an item’s visual features (Bae & Luck, 2018, 2019a; Wolff et al., 2015). Lateralized univariate EEG signals (e.g., lateralized alpha power, N2PC, CDA), have been fundamental for developing an understanding of human attention and working memory. By presenting identical visual stimuli in both hemifields, these lateralized signals exploit the contralateral organization of the human visual system. Conversely, however, to take advantage of these lateralized signals we *must* use specialized lateralized displays.

Lateralized displays offer some advantages, such as eliminating physical confounds and allowing for clever designs that place stimuli that are “invisible” to the analysis on the vertical mid-line (Feldmann-Wüstefeld & Vogel, 2019; Hickey, Di Lollo, & McDonald, 2009; Hillyard & Anllo-Vento, 1998; Hillyard, Hink, Schwent, & Picton, 1973; Vogel & Machizawa, 2004). However, lateralized CDA designs also introduce potential disadvantages. First, when presenting to-be-remembered items in both hemifields, there is some ambiguity as to whether differences in the CDA and behavior are confounded by the joint need to suppress irrelevant visual information. With increasing memory set size, there is both an increased need to remember more information and an increased amount of irrelevant visual information to suppress. Second, to measure lateralized components such as the

CDA, we must construct a difference score (contralateral—ipsilateral). The statistical reliability of difference scores is poor when the two underlying measures are highly correlated (Rodebaugh et al., 2016). Due to the poor spatial resolution of EEG, trial-by-trial voltage scores for contralateral and ipsilateral electrodes are highly correlated, thus the reliability of single-trial difference scores is lower than from single electrodes. Although we still obtain reliable estimates of CDA amplitude when averaging across many hundreds of trials, this traditional univariate approach potentially throws away valuable single-trial information that could be exploited to better predict working memory load.

In the current work, we used several data sets to demonstrate the utility of multivariate load detection in many contexts. We consistently found robust, sustained decoding of working memory load throughout the memory delay period, and this decoding predicted individual differences in working memory behavior. Multivariate load detection was sensitive to fine-grained variations in memory load as well as to working-memory specific (as opposed to general attentional) task demands. Furthermore, we showed that multivariate load detection generalized across stimulus differences (e.g., remembering colors vs. shapes; lateralized vs. whole-field presentation of the items) and generalized across observers (e.g., we can train the decoder on a large group of subjects, then predict memory load in a new subject whose data the classifier has never seen). The high generalizability, in particular, will be critical for future work; using the approach outlined here, we think it is possible to build a generalizable, pre-trained classifier which will be able to predict visual working memory load using relatively small amounts of data from new tasks and observers.

Future work will need to address some potential limitations of the current work. First, because we used previously published data sets, we were limited in our ability to perfectly control for potential confounds such as visual stimulation (i.e., transient luminance changes also increased with memory load). Future work using manipulations such as selective encoding (i.e., holding visual stimulation constant but varying which items are encoded) or retro-cues (Christophel, Hebart, & Haynes, 2012; Christophel, Iamshchinina, Yan, Allefeld, & Haynes, 2018; Griffin & Nobre, 2003; Harrison & Tong, 2009; Lepsién & Nobre, 2007; Sprague, Ester, & Serences, 2016), will be critical for disentangling encoding-related decoding from bottom-up, visually driven decoding of load during the early part of the trial. Second, a key contribution of this work is its demonstration of the feasibility of building a generalizable, pre-trained classifier for detecting working memory load in new tasks and observers. Here, we demonstrate that this type of generalizability is feasible across some stimulus differences and across observers within a site (i.e., controlling for high-level differences in experimental procedures such as amplifier, referencing,

and montage). Future work will be needed to further investigate the cross-site generalizability of multivariate load detection (e.g., variations in EEG systems, reference, experimental procedures, and subject pools) to build a generalizable, pre-trained working memory load detector (Dansereau et al., 2017; Scheinost et al., 2019).

Finally, we anticipate that similar approaches will also be useful outside of EEG. In the fMRI literature, for example, prior work has found univariate, load-dependent changes in parietal and prefrontal cortex (Braver et al., 1997; Cohen et al., 1997; Todd & Marois, 2004, 2005; Xu & Chun, 2006). In early visual cortex, in contrast, there are no load-dependent changes to the univariate signal with load. Despite this, the identity of a single item can be robustly decoded from visual cortex (Harrison & Tong, 2009; Kamitani & Tong, 2005; Serences, Ester, Vogel, & Awh, 2009). Furthermore, the fidelity of item-specific decoding in visual cortex is degraded with load (Emrich, Riggall, LaRocque, & Postle, 2013; Sprague, Ester, & Serences, 2014). However, no extant work has looked at multivariate neural signatures of load in visual cortex for supra-capacity set sizes, in part because of limitations of the spatial resolution of the method (with more than 2–3 items, the voxel population receptive fields would start to overlap substantially). Multivariate load detection, as performed here, could be used to probe whether working memory load, per se, can be decoded in visual cortex. Furthermore, this method could be used to test whether long-observed univariate signatures in parietal and frontal cortex are purely univariate in nature, or if more information about memory load can be gleaned by applying multivariate methods.

We argue that multivariate load detection (“mvLoad”) is a generalizable electrophysiological marker of human working memory load, and that this approach will allow for the unprecedented combination of disparate data sets to build a powerful, generalizable model of human working memory load. Because multivariate load detection is generalizable across tasks and observers, we anticipate that this method will be useful in many basic and applied research settings (i.e., unobtrusively monitoring the contribution of working memory during other cognitive contexts). All data shown here are available on the Open Science Framework (upon publication). We encourage other labs using distinct populations (e.g., developmental; clinical), research sites (e.g., outside of the U.S.), and task variants to use our published data to test the extent of the generalizability of this new method.

ACKNOWLEDGMENTS

We thank our co-authors on the published data sets used to conduct this work: Keisuke Fukuda, Eren Gunseli, Nicole Hakim, Irida Mance, Nash Unsworth, and Geoff Woodman.

CONFLICTS OF INTEREST

None.

AUTHOR CONTRIBUTIONS

Kirsten Adam: Conceptualization; Data curation; Formal analysis; Methodology; Software; Validation; Visualization; Writing-original draft; Writing-review & editing. **Edward Vogel:** Conceptualization; Funding acquisition; Resources; Supervision; Writing-review & editing. **Edward Awh:** Conceptualization; Funding acquisition; Resources; Supervision; Writing-review & editing.

DATA AVAILABILITY STATEMENT

Data for all experiments are freely available the Open Science Framework at <https://osf.io/6jkqu/>.

ORCID

Kirsten C. S. Adam  <https://orcid.org/0000-0002-4950-327X>

REFERENCES

Bae, G.-Y., & Luck, S. J. (2018). Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials. *The Journal of Neuroscience*, 38(2), 409–422. <https://doi.org/10.1523/JNEUROSCI.2860-17.2017>

Bae, G.-Y., & Luck, S. J. (2019a). Decoding motion direction using the topography of sustained ERPs and alpha oscillations. *NeuroImage*, 184, 242–255. <https://doi.org/10.1016/j.neuroimage.2018.09.029>

Bae, G.-Y., & Luck, S. J. (2019b). Reactivation of previous experiences in a working memory task. *Psychological Science*, 30(4), 587–595. <https://doi.org/10.1177/0956797619830398>

Bays, P. M. (2018). Reassessing the evidence for capacity limits in neural signals related to working memory. *Cerebral Cortex*, 28(4), 1432–1438. <https://doi.org/10.1093/cercor/bhx351>

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*, 5(1), 49–62. <https://doi.org/10.1006/nimg.1996.0247>

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *Journal of Neuroscience*, 31(25), 9315–9322. <https://doi.org/10.1523/JNEUROSCI.1097-11.2011>

Christophel, T. B., Hebart, M. N., & Haynes, J.-D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *Journal of Neuroscience*, 32(38), 12983–12989. <https://doi.org/10.1523/JNEUROSCI.0184-12.2012>

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., & Haynes, J.-D. (2018). Cortical specialization for attended versus unattended working memory. *Nature Neuroscience*, 21(4), 494–496. <https://doi.org/10.1038/s41593-018-0094-4>

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, 386(6625), 604–608. <https://doi.org/10.1038/386604a0>

Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, 250, 126–136. <https://doi.org/10.1016/j.jneumeth.2015.01.010>

Dansereau, C., Benhajali, Y., Risterucci, C., Pich, E. M., Orban, P., Arnold, D., & Bellec, P. (2017). Statistical power and prediction

accuracy in multisite resting-state fMRI connectivity. *NeuroImage*, 149, 220–232. <https://doi.org/10.1016/j.neuroimage.2017.01.072>

Emrich, S. M., Al-Aidroos, N., Pratt, J., & Ferber, S. (2009). Visual search elicits the electrophysiological marker of visual working memory. *PLoS One*, 4(11), e8042. <https://doi.org/10.1371/journal.pone.0008042>

Emrich, S. M., Riggall, A. C., LaRocque, J. J., & Postle, B. R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *Journal of Neuroscience*, 33(15), 6516–6523. <https://doi.org/10.1523/JNEUROSCI.5732-12.2013>

Fahrenfort, J. J., Grubert, A., Olivers, C. N. L., & Eimer, M. (2017). Multivariate EEG analyses support high-resolution tracking of feature-based attentional selection. *Scientific Reports*, 7(1), 1886. <https://doi.org/10.1038/s41598-017-01911-0>

Fahrenfort, J. J., van Driel, J., van Gaal, S., & Olivers, C. N. L. (2018). From ERPs to MVPA using the Amsterdam decoding and modeling toolbox (ADAM). *Frontiers in Neuroscience*, 12, 368. <https://doi.org/10.3389/fnins.2018.00368>

Feldmann-Wüstefeld, T., & Vogel, E. K. (2019). Neural evidence for the contribution of active suppression during working memory filtering. *Cerebral Cortex*, 29(2), 529–543. <https://doi.org/10.1093/cercor/rhw336>

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(7), 929–941. <https://doi.org/10.1177/0956797617699167>

Fukuda, K., Kang, M.-S., & Woodman, G. F. (2016). Distinct neural mechanisms for spatially lateralized and spatially global visual working memory representations. *Journal of Neurophysiology*, 116(4), 1715–1727. <https://doi.org/10.1152/jn.00991.2015>

Fukuda, K., Mance, I., & Vogel, E. K. (2015). Alpha power modulation and event-related slow wave provide dissociable correlates of visual working memory. *Journal of Neuroscience*, 35(41), 14009–14016. <https://doi.org/10.1523/JNEUROSCI.5003-14.2015>

Fukuda, K., & Woodman, G. F. (2017). Visual working memory buffers information retrieved from visual long-term memory. *Proceedings of the National Academy of Sciences*, 114(20), 5306–5311. <https://doi.org/10.1073/pnas.1617874114>

Fukuda, K., Woodman, G. F., & Vogel, E. K. (2015). Individual differences in visual working memory capacity: Contributions of attentional control to storage. In P. Jolicoeur, C. Lefebvre, & J. Martinez-Trujillo (Eds.), *Mechanisms of sensory working memory: Attention and performance XXV* (pp. 105–120). San Diego, CA: Elsevier.

Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience*, 15(8), 1176–1194. <https://doi.org/10.1162/089892903322598139>

Hakim, N., Adam, K. C. S., Gunseli, E., Awh, E., & Vogel, E. K. (2019). Dissecting the neural focus of attention reveals distinct processes for spatial attention and object-based storage in visual working memory. *Psychological Science*, 30(4), 526–540. <https://doi.org/10.1177/0956797619830384>

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238), 632–635. <https://doi.org/10.1038/nature07832>

Hickey, C., Di Lollo, V., & McDonald, J. J. (2009). Electrophysiological indices of target and distractor processing in visual search. *Journal of Cognitive Neuroscience*, 21(4), 760–775. <https://doi.org/10.1162/jocn.2009.21039>

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, 95(3), 781–787. <https://doi.org/10.1073/pnas.95.3.781>

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, 182(4108), 177–180. <https://doi.org/10.1126/science.182.4108.177>

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8(5), 679–685. <https://doi.org/10.1038/nn1444>

Lepsién, J., & Nobre, A. C. (2007). Attentional modulation of object representations in working memory. *Cerebral Cortex*, 17(9), 2072–2083. <https://doi.org/10.1093/cercor/bhl116>

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>

Munneke, J., Fahrenfort, J., Sutterer, D., Theeuwes, J., & Awh, E. (2019). Multivariate analysis of EEG activity indexes contingent and non-contingent attentional capture. *Neuroscience*, 734004. <https://doi.org/10.1101/734004>

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, 15(7), 327–334. <https://doi.org/10.1016/j.tics.2011.05.004>

Rihs, T. A., Michel, C. M., & Thut, G. (2007). Mechanisms of selective inhibition in visual spatial attention are indexed by alpha-band EEG synchronization. *European Journal of Neuroscience*, 25(2), 603–610. <https://doi.org/10.1111/j.1460-9568.2007.05278.x>

Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, A., ... Lenze, E. J. (2016). Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. *Journal of Abnormal Psychology*, 125(6), 840–851. <https://doi.org/10.1037/abn0000184>

Scheinost, D., Noble, S., Horien, C., Greene, A. S., Lake, E. M. R., Salehi, M., ... Constable, R. T. (2019). Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage*, 193, 35–45. <https://doi.org/10.1016/j.neuroimage.2019.02.057>

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, 20(2), 207–214. <https://doi.org/10.1111/j.1467-9280.2009.02276.x>

Sprague, T. C., Ester, E. F., & Serences, J. T. (2014). Reconstructions of information in visual spatial working memory degrade with memory load. *Current Biology*, 24(18), 2174–2180. <https://doi.org/10.1016/j.cub.2014.07.066>

Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring latent visual working memory representations in human cortex. *Neuron*, 91(3), 694–707. <https://doi.org/10.1016/j.neuron.2016.07.006>

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, 428(6984), 751–754. <https://doi.org/10.1038/nature02466>

Todd, J. J., & Marois, R. (2005). Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cognitive, Affective & Behavioral Neuroscience*, 5(2), 144–155. <https://doi.org/10.3758/CABN.5.2.144>

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and

secondary memory retrieval. *Cognitive Psychology*, 71, 1–26. <https://doi.org/10.1016/j.cogpsych.2014.01.003>

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2015). Working memory delay activity predicts individual differences in cognitive abilities. *Journal of Cognitive Neuroscience*, 27(5), 853–865. https://doi.org/10.1162/jocn_a_00765

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428(6984), 748–751. <https://doi.org/10.1038/nature02447>

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, 438(7067), 500–503. <https://doi.org/10.1038/nature04171>

Wolff, M. J., Ding, J., Myers, N. E., & Stokes, M. G. (2015). Revealing hidden states in visual working memory using electroencephalography. *Frontiers in Systems Neuroscience*, 9, 123. <https://doi.org/10.3389/fnsys.2015.00123>

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, 20(6), 864–871. <https://doi.org/10.1038/nn.4546>

Woodman, G. F., & Arita, J. T. (2011). Direct electrophysiological measurement of attentional templates in visual working memory. *Psychological Science*, 22(2), 212–215. <https://doi.org/10.1177/0956797610395395>

Xie, W., & Zhang, W. (2018). Familiarity speeds up visual short-term memory consolidation: Electrophysiological evidence from contralateral delay activities. *Journal of Cognitive Neuroscience*, 30(1), 1–13. https://doi.org/10.1162/jocn_a_01188

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, 440(7080), 91–95. <https://doi.org/10.1038/nature04262>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Adam KCS, Vogel EK, Awh E. Multivariate analysis reveals a generalizable human electrophysiological signature of working memory load. *Psychophysiology*. 2020;00:e13691. <https://doi.org/10.1111/psyp.13691>