

# Electroencephalogram Decoding Reveals Distinct Processes for Directing Spatial Attention and Encoding Into Working Memory



Henry M. Jones<sup>1,2</sup>, Gisella K. Diaz<sup>1,2</sup>, William X. Q. Ngiam<sup>1,2</sup>,  
and Edward Awh<sup>1,2</sup>

<sup>1</sup>Department of Psychology, The University of Chicago, and <sup>2</sup>Institute for Mind and Biology, The University of Chicago

Psychological Science  
1–31

© The Author(s) 2024

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09567976241263002

www.psychologicalscience.org/PS



## Abstract

Past work reveals a tight relationship between spatial attention and storage in visual working memory. But is spatially attending an item tantamount to working memory encoding? Here, we tracked electroencephalography (EEG) signatures of spatial attention and working memory encoding while independently manipulating the number of memory items and the spatial extent of attention in two studies of adults ( $N = 39$ ;  $N = 33$ ). Neural measures of spatial attention tracked the position and size of the attended area independent of the number of individuated items encoded into working memory. At the same time, multivariate decoding of the number of items stored in working memory was insensitive to variations in the breadth and position of spatial attention. Finally, representational similarity analyses provided converging evidence for a pure load signal that is insensitive to the spatial extent of the stored items. Thus, although spatial attention is a persistent partner of visual working memory, it is functionally dissociable from the selection and maintenance of individuated representations in working memory.

## Keywords

working memory, spatial attention, decoding, spatiotemporal pointers, electroencephalography, open data, open materials

Received 10/9/23; Revision accepted 4/1/24

## Introduction

There has been longstanding interest in our ability to exert voluntary control over attention to direct limited processing resources toward the most relevant aspects of the environment. But is voluntary attentional control a unitary process, or is it better understood as a constellation of distinct forms of control? We examined this question in the context of two well-known examples of goal-driven selective attention.

First, covert spatial attention can be voluntarily deployed to relevant regions of space and can modulate some of the earliest stages of sensory processing (Hillyard et al., 1998; Martinez et al., 1999). Second, people may selectively control which items enter *working memory*, an online memory system that allows us to store, manipulate, and rapidly access information (Cowan, 1999; Panichello & Buschman, 2021).

Importantly, working memory gating operates during relatively late stages of processing; items can be excluded from working memory storage even after full perceptual and semantic processing (e.g., Chun & Potter, 1995; Luck et al., 1996; Vogel et al., 1998, Vogel & Luck, 2002). Although it has been established that attention can operate during both early and late stages of processing, the possibility remains that a single control process mediates both types of selection. For example, directing covert attention toward an object may simultaneously modulate sensory processing (Martínez et al., 2006) and encode that item into working memory. In line with this possibility, spatial attention is sustained

---

### Corresponding Author:

Henry M. Jones, The University of Chicago, Department of Psychology  
Email: henryjones@uchicago.edu

at the location of items encoded into working memory even when location is irrelevant to the memory task (Foster et al., 2017), and disruption of spatial attention also disrupts working memory performance (Awh et al., 1998; Williams et al., 2013). This close partnership between spatial attention and visual working memory has motivated the perspective that storage in visual working memory is best understood as internally directed visual attention (Chun, 2011). Here, we present evidence that challenges this unitary model.

To study the relationship between spatial attention and working memory gating, we examined the neural signals that track each of these forms of selection. With scalp electroencephalogram (EEG) recordings, spatial attention can be tracked via measures of oscillatory activity in the alpha frequency band (8–12 Hz; see, e.g., Foster et al., 2017; Woodman et al., 2022), whereas the number of items encoded into working memory has been tracked via univariate and multivariate analyses of raw voltage (Adam et al., 2020; Luria et al., 2016; Thyer et al., 2022; Vogel & Machizawa, 2004). Recent work has identified plausible dissociations among these signals. They explain distinct variance in individual differences in working memory performance (Fukuda et al., 2015). They respond differently to distractors, and with different time courses (Hakim et al., 2021). Finally, they respond differently to manipulations of the number of relevant locations, compared to the number of individuated items occupying those locations (Diaz et al., 2021; Hakim et al., 2019; Thyer et al., 2022). These dissociations indicate that spatial attention may be controlled separately from encoding into working memory, but these findings are limited in two ways. First, almost every study dissociating spatial attention and working memory load signals conflates the number of relevant items with the number of relevant locations (but see Diaz et al., 2021). In addition, the alpha measures in these studies were limited to univariate measures of power or laterality within posterior electrodes, precluding a precise link between those alpha oscillations and the deployment of covert attention to specific regions of space. Thus, we employed refined EEG measures of covert spatial attention that track both the position and breadth of the spatially attended regions, and we designed a task that deconflates the number of relevant items and the breadth of the relevant locations. As discussed below, these refinements provided strong traction for examining the separability of spatial attention and WM gating.

Two experiments provided clear evidence that distinct neural signals tracked the deployment of spatial attention and the maintenance of individuated items in working memory. In Experiment 1, we independently

### Statement of Relevance

What does it mean to pay attention to something? Psychologists have long known that people can voluntarily direct attention toward relevant aspects of the environment to succeed at everyday tasks. Here, we present new evidence for a division of voluntary attention into at least two distinct types. On the one hand, people can direct their attention toward specific locations. On the other hand, people can decide which items will be encoded into visual working memory (i.e., voluntarily selected for storage). In two studies, we recorded electroencephalograms (EEGs) while participants performed a memory task in which we separately manipulated where participants should attend and how many items they needed to encode. We found that the number of items encoded into working memory is disconnected from the allocation of spatial attention to those items' positions. These findings help to refine our understanding of the voluntary control processes that allow a person to allocate limited resources to parse and understand complex visual scenes.

manipulated the number of relevant items and the number of relevant locations by sequentially presenting stimuli within overlapping or unique locations. We found that posterior alpha power was sensitive to the number of spatial locations, and not the number of items, whereas a load-decoding model tracking the number of individuated items was insensitive to the number of locations occupied by those items. In Experiment 2, we used novel dot cloud stimuli that varied strongly in spatial extent and that could overlap with minimal perceptual interference. These stimuli enabled independent manipulation of the number of individuated items and the spatial extent of the relevant regions. Using inverted encoding models to decode both the location and precision of spatial attention (Foster et al., 2017), we found that spatial attention tracked the breadth of the spatial area occupied by the memorized items but was minimally impacted by the number of separate clouds in the display. Simultaneously, multivariate decoding of the number of items in the display (mvLoad) precisely tracked the number of dot clouds held in working memory, despite strong variations in the spatial extent of those stimuli. Finally, we used representational similarity analysis (RSA) to assess the contributions of multiple independent factors that could influence our load decoding results. RSA provided

converging evidence for a pure load signal that is uniquely determined by the number of individuated objects that are stored in working memory, as well as unique variance in EEG activity that tracked the spatial extent of the stored items. Thus, spatial attention and visual working memory storage are intertwined but distinct aspects of voluntary attentional control.

## Open Practices Statement

Neither of the studies reported in this article was pre-registered. The code is available at <https://github.com/henrymj/DissociatingWMLoadFromSpatialAttention>. The data and code are available on the Open Science Framework at <https://osf.io/8svur2/>.

## Experiments 1a and 1b

### Method

**Participants.** Participants were recruited from the University of Chicago and the surrounding community. In total, 12 (9 female; mean age = 25 years,  $SD = 3.7$ ) and 27 (18 female; mean age = 25 years,  $SD = 4.0$ ) participants were used in Experiments 1a and 1b, respectively. We excluded data from 2 participants in Experiment 1a and 7 participants in Experiment 1b because of excessive EEG artifacts (< 200 trials remaining per condition in Experiment 1a and < 140 in Experiment 1b). For Experiment 1a, the intended sample size was 12 participants, because previous research had shown that this is a sufficient number of participants for observation of set-size effects on parieto-occipital alpha power (see Experiment 1 of Diaz et al., 2021). The 2 participants who were excluded from our analyses because of an insufficient number of trials were not replaced in favor of collecting data using a modified task design that balanced visual stimulation across displays (i.e., Experiment 1b). For Experiment 1b our intended sample size was 20 participants, because previous work had demonstrated that this is an appropriate number of participants for observation of more nuanced set-size effects on alpha power (see Experiment 2 of Diaz et al., 2021); we anticipated such effects when we introduced placeholders to balance visual stimulation across displays, because we were concerned that some participants might inadvertently attend the placeholders. Participants with excessive EEG artifacts were replaced until we completed our intended sample size.

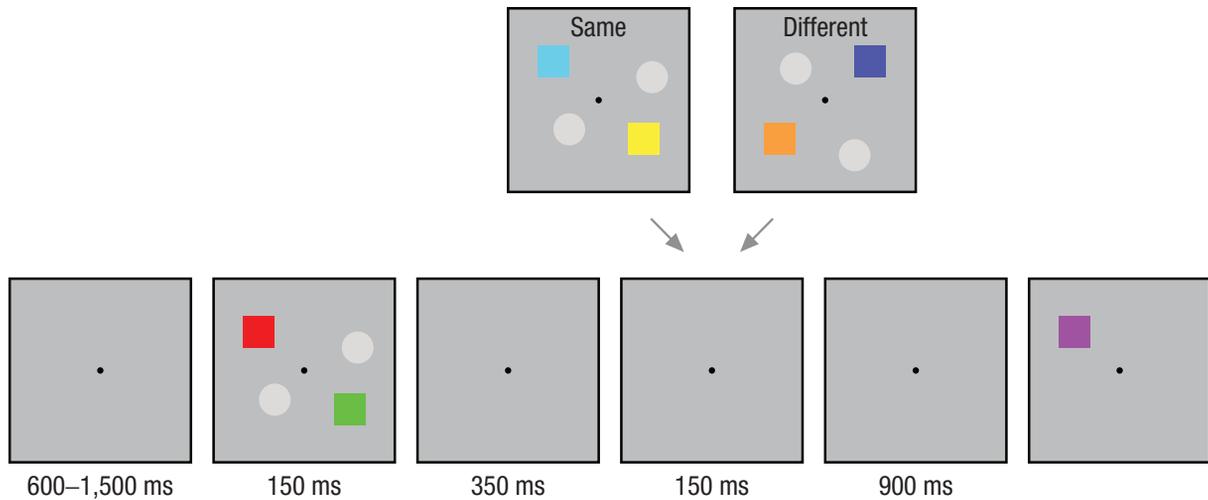
Experimental procedures were approved by the Institutional Review Board at the University of Chicago. All participants gave informed consent and were compensated for their participation at a rate of \$15 per hour. Participants reported normal color vision and normal or corrected-to-normal visual acuity.

**Apparatus.** Participants were tested in a dimly lit, electrically shielded chamber. Stimuli were generated using MATLAB (The MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Stimuli were presented on a 24-in. LCD monitor (refresh rate: 120 Hz, resolution: 1080 × 1920 pixels) at a viewing distance of approximately 75 cm and against a dark-gray background.

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

**Eye tracking.** We recorded gaze position using a desk-mounted infrared eye-tracking system (EyeLink 1000 Plus, SR Research, Ottawa, Ontario, Canada). Gaze position was sampled at 1000 Hz. Stable head position was maintained during the task using a chin rest. The eye tracker was recalibrated as needed throughout the session, including whenever participants removed their chin from the chin rest.

**Artifact rejection.** For artifact rejection, each trial was segmented into –400 ms pretrial and 1,550 ms poststimulus array onset epochs. We used an automated procedure to flag trials that were contaminated by ocular or EEG artifacts. Next, we used this procedure as a guideline during manual visual inspection where it was ultimately determined which trials were to be rejected. Experimenters were blind to condition when inspecting the data for artifacts. Trials contaminated by artifacts were excluded from EEG analyses but not from behavioral analyses. Participants were excluded from the final sample if they had fewer than 200 artifact-free trials per condition in Experiment 1a (average trials per condition = ~277 out of 320).



**Fig. 1.** Example trials for sequential change-detection task. Gray placeholders were not presented in Experiment 1a. For both Experiment 1a and Experiment 1b, two stimulus arrays were separated by a brief interstimulus interval. The first stimulus array contained either one or two colored squares, depending on the set size (two or four, respectively). The second stimulus array also contained one or two squares that were either presented in different locations or presented in the same locations occupied in the first array.

and fewer than 140 artifact-free trials per condition in Experiment 1b (average = ~198 out of 240).

An automated artifact-detection procedure was used to detect eye movements, blinks, and EEG artifacts. Trials were flagged as containing a saccade if the euclidean vector between the mean gaze positions in the first and second halves of an 80-ms sliding window (advanced in 10-ms increments) was greater than  $0.5^\circ$  of visual angle. When eye tracking data were not available, we used horizontal EOG to detect saccades. Trials were flagged as containing a saccade if the mean voltage during the first and second halves of a 150-ms sliding window (advanced in 10-ms steps) exceeded  $20 \mu\text{V}$ .

For blinks, trials were flagged as containing a blink if the eye tracker could not detect the pupil at any point during the trial. When eye-tracking data were not available, we used vertical EOG to detect blinks. Trials were flagged as containing a blink if the mean voltage during the first and second halves of a 150-ms sliding window (advanced in 10-ms steps) exceeded  $30 \mu\text{V}$ .

For EEG artifacts, we flagged trials as containing voltage drifts (e.g., skin potentials) if the absolute change in voltage from the first quarter of the trial to the last quarter of the trial exceeded  $100 \mu\text{V}$ . We flagged trials as including a sudden step in voltage (which can occur when an electrode is damaged) if the mean voltage during the first and second halves of a 250-ms sliding window (advanced in 20-ms increments) differed by more than  $100 \mu\text{V}$ . We marked trials as containing high-frequency noise (e.g., muscle artifacts) if any electrode had a peak-to-peak amplitude greater than

$150 \mu\text{V}$  within a 15-ms sliding window (advanced in 50-ms increments). Finally, we flagged trials as containing amplifier saturation if any electrode had 60 time points within a 200-ms sliding window (advanced in 50-ms increments) that were within  $1 \mu\text{V}$  of each other.

**Experiment 1a procedure.** Participants performed a change-detection task (Fig. 1). The trial began with a black fixation dot (diameter =  $0.20^\circ$ ) presented at the center of a dark-gray background for a randomly determined duration between 600 ms and 1,500 ms. The fixation dot remained visible throughout the trial. A stimuli array followed consisting of one or two colored squares (set size two or four, respectively; length =  $2^\circ$ ) presented for 150 ms. The stimuli were presented within a predetermined area (Experiment 1a:  $9.90^\circ \times 9.90^\circ$ ; Experiment 1b:  $12.5^\circ \times 12.5^\circ$ ) and at least  $2.1^\circ$  (Experiment 1a) or  $3^\circ$  (Experiment 1b) away from fixation. The predetermined area was divided into a  $4 \times 4$  grid; each of the sections could contain a single stimulus. The positions needed for the first array were randomly chosen from the 16 possible locations without replacement. In the same-location condition, the same locations were used for the second array. Otherwise, new locations were randomly selected from the set of possible locations remaining that were not already used in the first array. Jitter ( $\sim 0.25^\circ$ ) was added to the actual locations occupied by stimuli.

The stimuli in the first array were rendered in a color drawn randomly from nine possible colors without replacement (red, green, blue, yellow, magenta, cyan, white, black, and orange). Next, an interstimulus interval followed in which only the fixation dot remained

on the screen for 350 ms. Then the second stimuli array followed, consisting of either one or two squares to complete the set size to be maintained for that trial (set size two or four, respectively). The stimuli in this second array were either presented in the same locations that were previously occupied or in new locations (same or different locations, respectively).

The colors for the stimuli presented in the second array were determined randomly from the colors remaining and excluded the colors used in the first array. Participants were asked to remember the color and locations of the stimuli over a 900-ms blank delay interval during which only the fixation dot remained on the screen. After the delay, a single probe stimulus reappeared in one of the locations that was previously occupied and was either rendered in the same color as one of the original stimuli presented there or in a different color drawn randomly from the entire set of colors, excluding the color that was actually presented there. This meant that when the locations of the stimuli were the same between the first and second array (same-location condition), the set of possible change colors excluded both colors presented in the probed location. Participants used a keyboard button press to indicate whether this probe stimulus was the same or not. Participants pressed the “z” or “/” key to indicate whether the color of the probe stimulus was the same or different, respectively. There were no practice trials given before the formal experiment. Participants were given verbal and written task instructions with the aid of an example trial image similar to that of Figure 1.

Participants completed 20 blocks with each containing 64 trials. Within a block, half of the trials were no-change trials, and the remaining half were change trials in which the probe stimulus was rendered in a different color than the one (or any) presented there. Similarly, half of the trials were set size 2 (SS2); the first array contained one square and the second array contained an additional square. The remaining half were set size 4, with two squares presented in each of the first and second arrays. Within each block, there were also an equal number of trials using the same location and different locations that determined whether the stimuli in the second array were presented in the same locations as the first array or not. Finally, the probe stimulus was drawn from the first and second array with equal probability.

Participants self-initiated each block by pressing the space bar key. The experiment session was scheduled to take 3 hours, but the actual duration of the session depended on the participants' pace because they initiated each block and decided when (and if) to take breaks between blocks.

**Experiment 1b procedure.** The procedure was similar to that of Experiment 1a, with the following exceptions. First, an additional condition was included that simultaneously presented two or four squares in the first array, though the data from this condition are not analyzed further. Additionally, gray placeholder circles (diameter =  $\sim 2.26^\circ$ ) were presented in both the first and second arrays so that four items were always presented. For set size 2, this meant that each array contained three placeholders, whereas set size 4 trials contained two placeholders in each array. On each trial, four positions were randomly chosen from the set of possible locations. Depending on the set size, positions were assigned to stimuli and placeholders. In same-location-condition trials, the locations of the gray placeholders (and stimuli) were the same for both arrays. The locations were switched on the different-location-condition trials so that the stimuli were placed in the placeholder locations from the first array, and the placeholders were placed in the stimuli locations from the first array.

Participants completed 20 blocks with each containing 72 trials. The experimental session was scheduled to take 3.5 hours, but the actual duration of the session depended on the participants' pace because they initiated each block and decided when (and if) to take breaks between blocks.

**Experimental design.** Both experiments used a  $2 \times 2$  within-participants design. The factors were set size (2 or 4) and location condition (same or different locations). Behavioral data (i.e., accuracy) were analyzed using a repeated-measures analysis of variance (ANOVA), computed in JASP version 0.18.1 (Love et al., 2019). For this and all ANOVAs, we also computed a Bayesian repeated-measures ANOVA in JASP. All models were given an equal prior probability, and the change in model odds from prior to posterior ( $BF_M$ ) of the winning model is reported, along with the Bayes factor of the winning model against the null model (including only participant intercepts and random slopes).

**Parieto-occipital alpha power analysis.** EEG signal processing was performed in MATLAB. We band-pass filtered the raw EEG data using a filter from the FieldTrip toolbox (`ft_preproc_bandpassfilter.m`; Oostenveld et al., 2011), and then extracted instantaneous power values for the alpha band (8–12 Hz) by applying a Hilbert transform from the Signal Processing Toolbox in MATLAB (`hilbert.m`) to the filtered data. We calculated alpha power for the parieto-occipital electrodes: P7, P3, Pz, P4, P8, PO7, PO3, PO4, PO8, O1, Oz, O2. For illustrative purposes in the figures, we subtracted the mean baseline (–400 ms to 0 ms) at each time point in the trial for each condition and converted to percent change from baseline.

**Multivariate classification analysis.** For the mvLoad analysis, we used a logistic regression model to classify the number of items in working memory (i.e., working memory load) using baselined EEG (Thyer et al., 2022). EEG activity was calculated using a baseline from  $-500$  ms to  $-100$  ms relative to the onset of the stimulus array. The mean baseline amplitude was subtracted from EEG amplitude at each time point in the trial. To improve our signal-to-noise ratio, we randomly selected trials within each condition of interest to create groups of 20 trials and then averaged across the trials in each group. The classification procedure was performed by averaging voltage for each electrode across a 50-ms time window that advanced in 25-ms steps. At each time point the training data were standardized, and the testing data were standardized using the mean and standard deviation of the training data (*StandardScaler* in Scikit-learn; Pedregosa et al., 2011). The classifiers were trained to discriminate between our conditions of interest and then tested on a held-out set of data (*StratifiedShuffleSplit* in Scikit-learn). In this cross-validation procedure, the data for any given condition were split so that 80% of the data were used in training and the remaining 20% of data were used in testing. The training data were stratified by condition, and the number of trial groups per condition were equated by randomly down-sampling the condition with more trial groups. This procedure was repeated 1,000 times with results averaged across all iterations.

**Statistical analysis.** Behavioral data were analyzed using a repeated-measures ANOVA as well as a Bayesian repeated-measures ANOVA. To increase the power of our EEG results, we averaged alpha and mvLoad results across the delay period for testing. To avoid bleedover from the test period in alpha, we excluded the last 100 ms; we did the same for the mvLoad results. We tested alpha power using a repeated-measures ANOVA and a Bayesian repeated-measures ANOVA with a factor for set size (2 or 4) and a factor for location condition (same or different). We also planned to test whether the two conditions that differed in set size without differing in the number of locations—the set size 2 condition with different locations and the set size 4 condition with repeated locations—actually differed in terms of alpha power. We tested this using a repeated-measures  $t$  test. For this and all  $t$  tests, we also examined the Bayes factor (BF; computed with a standard Cauchy scale of .707) for evidence against the null ( $BF_{10}$ ), the reciprocal of which reflects evidence in favor of the null.

We tested decoding accuracy for each location condition separately using a repeated-measures  $t$  test of decoding accuracy for true labels against decoding accuracy for shuffled labels in each model. In the second variation of mvLoad analysis, repeated-measures  $t$

tests were used to test for differences in the classifiers' confidence between conditions of interest at each time window. Confidence scores refer to the signed distance of the test sample to the hyperplane in arbitrary units (*decision\_function* in Scikit-learn).

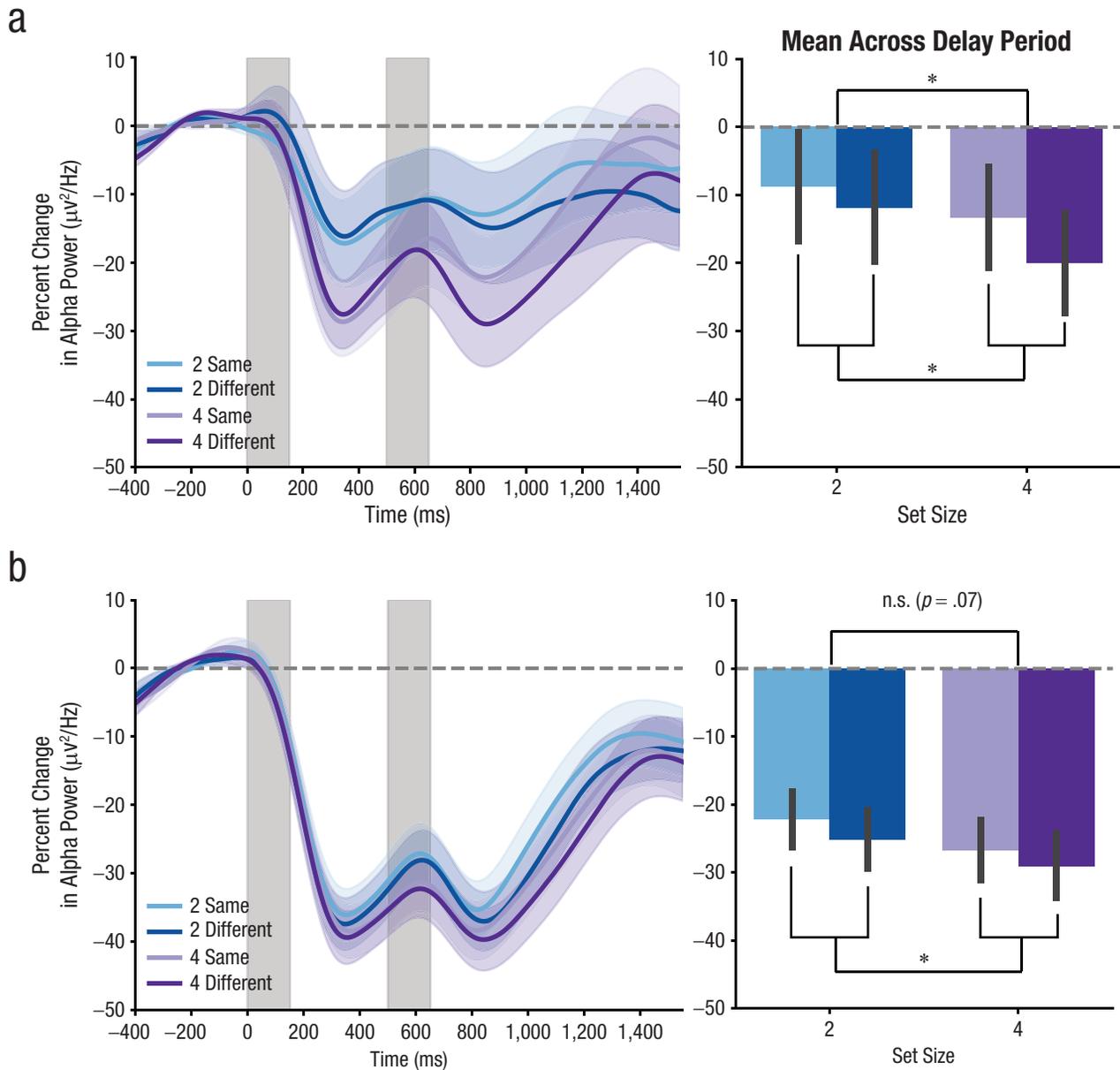
## Results

**Behavior.** For Experiment 1a, there was a main effect of set size,  $F(1, 9) = 61.90$ ,  $p < .001$ ,  $\eta_G^2 = .64$ , and location condition,  $F(1, 9) = 8.80$ ,  $p = .016$ ,  $\eta_G^2 = .048$ , on accuracy: Accuracy was higher for set size 2 ( $M = 0.98$ ,  $SD = 0.01$ ) than set size 4 ( $M = 0.87$ ,  $SD = 0.06$ ), and when different locations were occupied in the second array ( $M = 0.93$ ,  $SD = 0.07$ ) compared to the same locations ( $M = 0.92$ ,  $SD = 0.07$ ). There was a significant interaction between set size and location condition on accuracy,  $F(1, 9) = 8.17$ ,  $p = .019$ ,  $\eta_G^2 = .027$ , so that the benefit of appearing in different locations was greater for set size 4 than for set size 2, perhaps because of a ceiling effect in set size 2. A Bayesian repeated-measures ANOVA also supported a model with main effects of set size and location, along with an interaction ( $BF_M = 10.29$ , next highest = 0.91;  $BF_{10} = 8,889.37$ ).

The pattern of behavioral results was replicated in Experiment 1b. There was a main effect of set size,  $F(1, 19) = 69.47$ ,  $p < .001$ ,  $\eta_G^2 = .41$ , and a trending effect of location condition,  $F(1, 19) = 4.09$ ,  $p = .057$ ,  $\eta_G^2 = .007$ , so that accuracy was higher for set size 2 ( $M = 0.95$ ,  $SD = 0.06$ ) than for set size 4 ( $M = 0.82$ ,  $SD = 0.10$ ) and when different locations were occupied in the second array ( $M = 0.89$ ,  $SD = 0.11$ ) compared to the same locations ( $M = 0.88$ ,  $SD = 0.10$ ). Similar to Experiment 1a, there was a significant interaction between set size and location condition on accuracy,  $F(1, 19) = 9.49$ ,  $p = .006$ ,  $\eta_G^2 = .005$ , in which the benefit of appearing in different locations was present only for set size 4 compared to set size 2. A Bayesian repeated-measures ANOVA also supported a model with main effects for set size and location, along with an interaction ( $BF_M = 11.19$ , next highest = 0.75;  $BF_{10} = 868,661.57$ ).

Across both experiments, there was a reliable set size effect: Performance was higher for set size 2 than for set size 4. For set size 4, there was also a small benefit when the stimuli locations occupied in the second array were different from those used in the first array.

**Parieto-occipital alpha power.** We examined the effects of set size and location condition on alpha power at parieto-occipital electrodes (Maris & Oostenveld, 2007). On the basis of previous research, we predicted that alpha power would be sensitive to both factors (Diaz et al., 2021; Fukuda et al., 2015). Thus, we computed a repeated-measures ANOVA with set size and location



**Fig. 2.** Averaged alpha power suppression observed at parieto-occipital electrodes in Experiment 1a (a) and 1b (b). At left, the time course of alpha power suppression across the trial is shown. Light gray shaded regions indicate the duration of stimulus arrays. At right, the average suppression across the delay period is shown. Significant markers (\*) above indicate a main effect of set size, and significance markers below indicate a main effect of location ( $p < .05$ ). There was no significant interaction in either experiment.

condition as factors on the average alpha power during the delay period (Fig. 2). In Experiment 1a, there was a significant main effect of set size,  $F(1, 9) = 8.06, p = .019, \eta_G^2 = .0027$ , as well as of location condition,  $F(1, 9) = 10.60, p = .010, \eta_G^2 = .0054$ , and no significant interaction,  $F(1, 9) = .05, p = .82, \eta_G^2 = 3\text{e-}6$ . A Bayesian repeated-measures ANOVA also supported a model with main effects for set size and location, along with an interaction ( $\text{BF}_M = 5.03$ , next highest = 1.06;  $\text{BF}_{10} = 26.47$ ). A planned  $t$  test comparing set size 2 at different locations with set

size 4 at the same location found no significant difference between the conditions, but weak evidence in favor of the null,  $t(9) = 1.57, p = .15, d = 0.04, \text{BF}_{10} = 0.787$ .

Experiment 1b produced similar results to those of Experiment 1a. There was a trend toward a main effect of set size,  $F(1, 19) = 3.5, p = .076, \eta_G^2 = .0019$ , and a main effect of location condition,  $F(1, 19) = 5.09, p = .036, \eta_G^2 = .0023$ , and no significant interaction,  $F(1, 19) = 0.20, p = .66, \eta_G^2 = 8\text{e-}6$ . A Bayesian repeated-measures ANOVA also supported a model with main

effects for set size and location, but no interaction, though evidence was weak ( $BF_M = 1.66$ , next highest = 1.21;  $BF_{10} = 2.39$ ). A planned  $t$  test comparing set size 2 at different locations with set size 4 at the same location found no significant difference between the conditions and moderate evidence in favor of the null,  $t(19) = 0.30$ ,  $p = .76$ ,  $d = 0.007$ ,  $BF_{10} = 0.242$ .

As predicted, alpha power at parieto-occipital electrodes appeared to be sensitive to set size. For both experiments, there was greater alpha suppression for four items than two items during the delay. However, alpha power was also sensitive to whether items were added to the same locations or different locations. Specifically, there was greater alpha suppression when items were added to different locations for both set size 2 (Experiment 1) and set size 4 (Experiments 1 and 2). In addition, when we directly compared conditions that differed in the number of items but not the number of locations, we found weak-to-moderate evidence for no difference in alpha power, though it is possible that a set-size effect is present in alpha power and we lacked the statistical power to detect it. One possible interpretation for these findings is that alpha power tracked the shift in spatial attention required in the different location trials. However, we propose that the dominant signal in alpha power tracks the number of locations occupied by the memoranda. This latter interpretation is in line with previous work suggesting that alpha power at posterior electrodes also tracks the number of relevant locations when all items are presented simultaneously (Diaz et al., 2021; Fukuda et al., 2015). Thus, the effects of set size and location condition might both be explained by variations in the number of attended locations.

**Multivariate analysis of voltage.** Besides parieto-occipital alpha power, the number of items in working memory can be decoded using multivariate analysis of EEG voltage, or mvLoad (Adam et al., 2020; Thyer et al., 2022). However, the extant work with this approach has always presented distinct items in unique locations, producing a confound between working memory load and the number of relevant positions in the display. Thus, the central goal of Experiment 1 was to test whether the mvLoad approach is sensitive to the total number of items stored when the confound between load and number of locations is eliminated. We accomplished this by presenting each half of the memory array sequentially, with the second half either in the same position or in different positions from the first half of the array. This design allowed us to manipulate the number of relevant positions while the number of stored items was held constant.

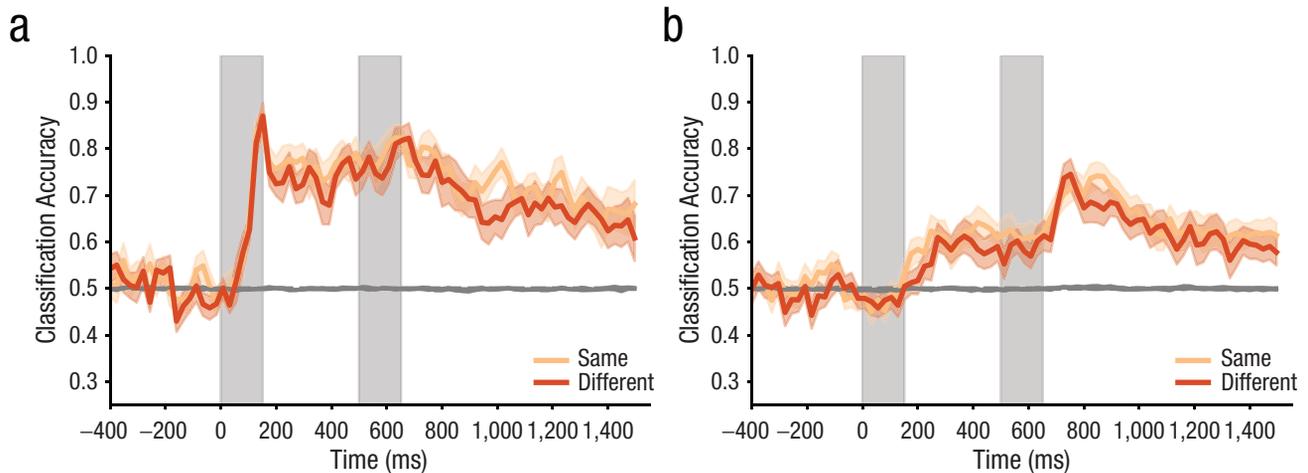
We used a logistic regression model to classify working memory load using baselined EEG. For each

condition of interest. Trials were divided into groups of 20 and then averaged with the resulting matrix (Electrodes  $\times$  Time Points)—subsequently referred to as a trial for ease of reference. The trials (−400 ms to 1,550 ms) were divided into 50-ms time windows with a sliding window of 25 ms. Data were averaged within each time window so that each trial was represented by a  $30 \times 80$  matrix (Electrodes  $\times$  Time Windows). Finally, classification analyses were performed at each time window for each participant.

First, we investigated whether mvLoad could decode the number of items encoded into working memory. To this end, we trained and tested separate classification models to discriminate between set size 2 and 4 within the same-location and different-location conditions. We tested whether accuracy was above chance by averaging performance over the delay period and applying a repeated-measures  $t$  test comparing accuracy using true labels to accuracy using shuffled labels (Fig. 3). In Experiment 1a, we could classify set size across the delay period in both location conditions—same:  $t(9) = 9.0$ ,  $p = 9e-6$ ,  $d = 4.1$ ,  $BF_{10} = 2,370.29$ ; different:  $t(9) = 6.67$ ,  $p = 9e-5$ ,  $d = 3.0$ ,  $BF_{10} = 308.63$ . The results from Experiment 1b follow the same pattern—same:  $t(19) = 6.48$ ,  $p = 3e-6$ ,  $d = 2.0$ ,  $BF_{10} = 6,067.20$ ; different:  $t(19) = 5.77$ ,  $p = 1.5e-5$ ,  $d = 1.7$ ,  $BF_{10} = 1,568.96$ .

Although these findings are consistent with a working memory load signal that does not depend on the number of relevant locations, this analysis is inconclusive because load was still confounded with the number of relevant locations. To examine whether there is a load signature that does not depend on the number of relevant positions, we tested whether the multivariate signature of load in the same-location condition generalized to that in the different-location condition. If the mvLoad analysis was classifying load on the basis of the number of locations occupied by the memory items, then a classifier trained on the same-location conditions should be biased toward a higher set size when tested with data from the different-location condition. Likewise, a reliance on the number of occupied positions would lead a classifier trained on the different-location conditions to be biased toward a lower set size when it is tested on data from the same-location condition. The following results show that this was not the case.

We obtained a measure of the classifiers' confidence for each decision made about the test sample (*decision\_function* in Scikit-learn). This confidence score reflects the signed distance of the test sample to the hyperplane in arbitrary units. Positive scores indicated that the trial was classified as set size 4 (same or different, depending on the analysis) with higher scores reflecting stronger evidence for this decision.



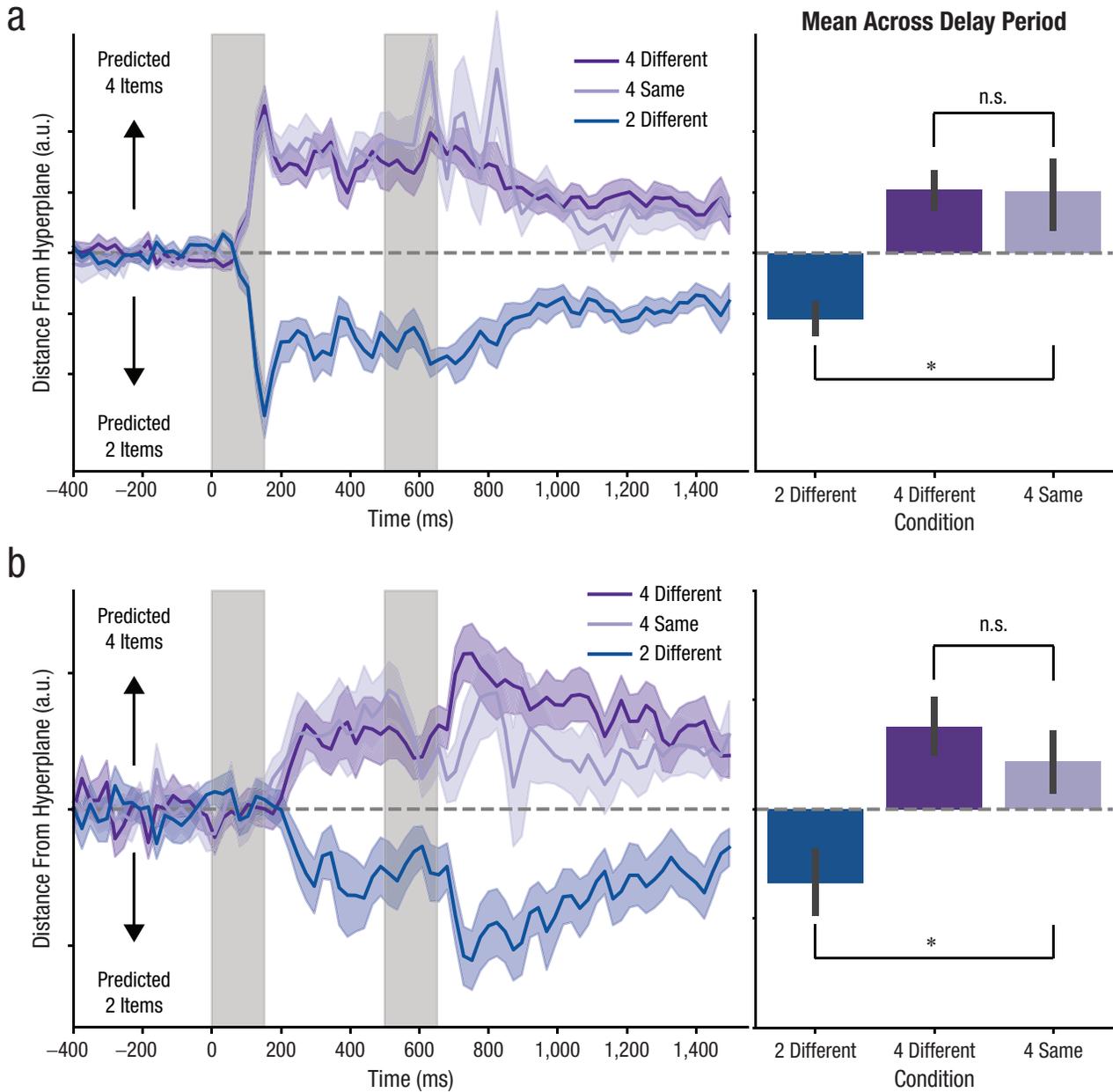
**Fig. 3.** Classification accuracy over time for set size for same locations (light orange) and set size for different locations (dark orange) in Experiments 1a (a) and 1b (b). Shaded regions indicate the duration of stimulus arrays. Gray lines varying about 0.5 indicate accuracy for shuffled labels.

Meanwhile, negative scores indicated a set size 2 (same or different) classification, with lower scores reflecting stronger evidence. Our time window of interest was the delay period (starting at 650 ms after the initial sample onset, excluding the last 100 ms to match our alpha power analyses above), given that our comparisons of interest relied on the total set size. To increase the signal-to-noise ratio (SNR), we averaged these distance values across the time window of interest and tested for differences using a repeated-measures *t* test.

First, we discuss the analysis that used training data from the different-location condition. As demonstrated above, multivariate analysis could distinguish the patterns of activity between set size 2 (different location) and set size 4 (different location). The key question was whether or not the model trained on the different-location condition would generalize to data from the same-location condition. Recall that set size 4 (same-location condition) had the same number of items as set size 4 (different-location condition) but had the same number of relevant locations as set size 2 (different-location condition). Despite this, in both experiments the classifier trained on different-location data (Fig. 4) consistently classified set size 4 same-location trials as significantly different from set size 2 different-location trials—Experiment 1a:  $t(9) = 5.57$ ,  $p = .0003$ ,  $d = 2.85$ ,  $BF_{10} = 98.89$ ; Experiment 1b:  $t(19) = 5.75$ ,  $p = 1.5e-5$ ,  $d = 1.75$ ,  $BF_{10} = 1,489.00$ —but not significantly different from set size 4 different-location trials, with weak to moderate evidence of the null—Experiment 1a:  $t(9) = 0.119$ ,  $p = .91$ ,  $d = 0.04$ ,  $BF_{10} = 0.31$ ; Experiment 1b:  $t(19) = 1.62$ ,  $p = .123$ ,  $d = 0.55$ ,  $BF_{10} = 0.71$ . Thus, the patterns of activity between two items in two locations and four items in two locations were discernibly and reliably different during the delay

period, whereas the patterns of activity between four items in four locations and four items in two locations were not.

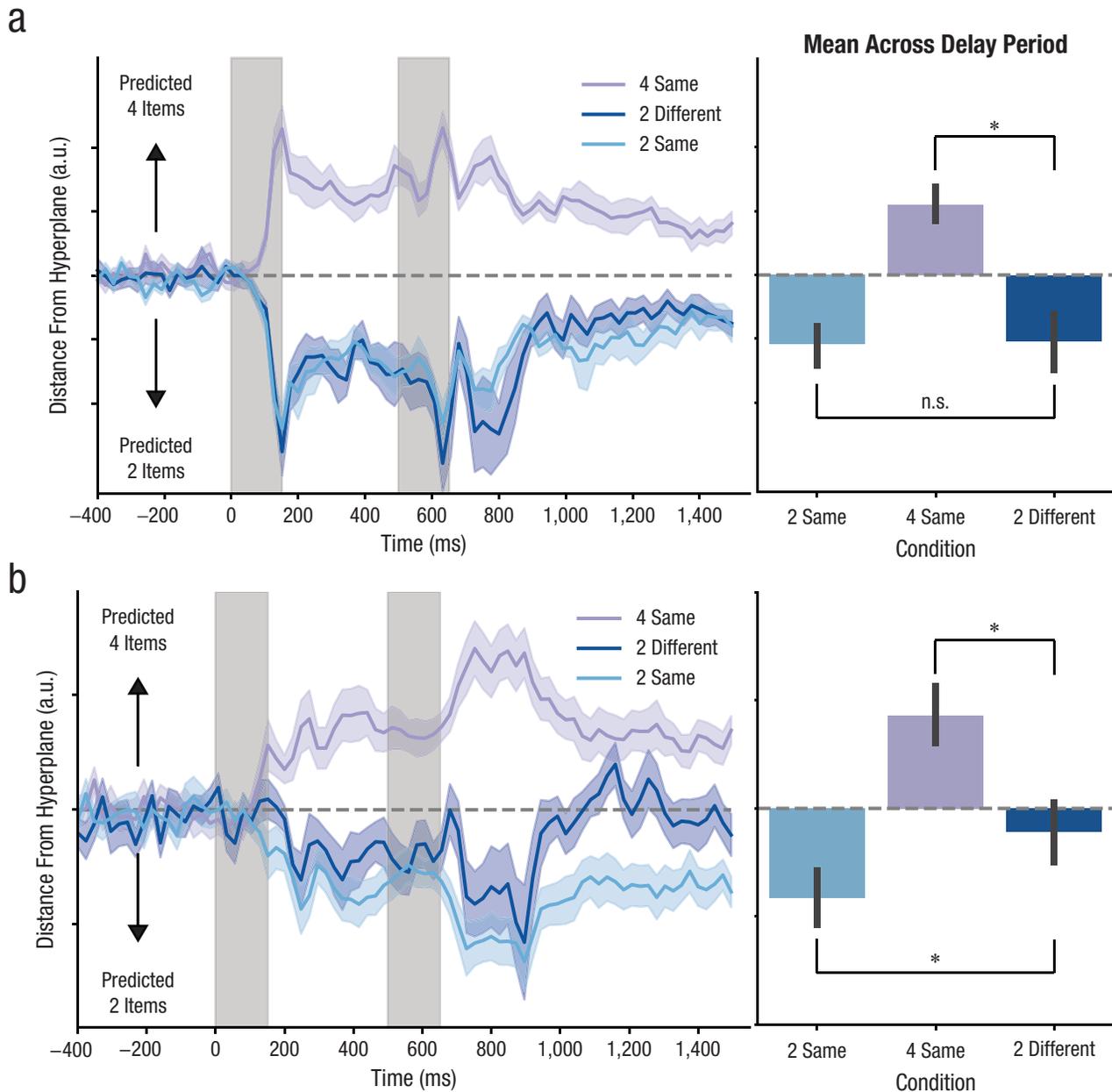
We conducted a complementary analysis after training the load classifier on data from the same-location condition. In this case, we tested the model on data from the set size 2 different-location condition, which had the same number of individuated items as the set size 2 same-location condition but occupied the same number of locations as the set-size 4 same-location condition. Figure 5 shows that the mvLoad analysis could distinguish the patterns of activity between the set size 2 same-location condition and the set size 4 same-location condition, just as seen in the initial decoding accuracy results. Moreover, when the same load model was tested on the set size 2 different-location condition, the condition was more likely to be classified as the set size 2 same-location condition throughout the delay period in Experiment 1a, even though the number of locations in the set size 2 different-location condition contained more locations than did the set size 2 same-location condition (Fig. 5a). The distances from the hyperplane were not significantly different between the set size 2 different-location condition and the set size 2 same-location condition, with weak-to-moderate evidence in favor of the null,  $t(9) = 0.249$ ,  $p = .81$ ,  $d = 0.08$ ,  $BF_{10} = 0.32$ , suggesting that the patterns of voltage activity reflecting working memory load were equivalent for two items in two locations and two items in one location (delay). In contrast, the distances from the hyperplane for the set size 2 different-location condition and the set size 4 same-location condition were significantly different,  $t(9) = 6.50$ ,  $p = .0001$ ,  $d = 3.53$ ,  $BF_{10} = 260.63$ . In Experiment 1b, the distances from the hyperplane for set size 2



**Fig. 4.** Distance from the classification hyperplane for set size 2 different-location trials, set size 4 different-location trials, and set size 4 same-location trials across time for Experiments 1a (a) and 1b (b). Classification was trained on set size 2 and 4 different-location trials and tested on all three conditions. The dashed gray line indicates the hyperplane. At left, distances across time are shown; at right, mean distances during the delay period are shown, with significance markers comparing the set size 4 same-location distances to those of the training conditions. Note that the distances are in arbitrary units (a.u.). Significance markers (\*) indicate significant differences ( $p < .05$ ) in the distances from the hyperplane between pairs of conditions.

different-location condition was different from both the set size 4 same-location condition,  $t(19) = 4.76$ ,  $p = .0001$ ,  $d = 1.67$ ,  $BF_{10} = 209.65$ , and the set size 2 same-location condition,  $t(19) = 3.66$ ,  $p = .0017$ ,  $d = 0.96$ ,  $BF_{10} = 23.54$ . From the decoding results across time (Fig. 5b), it appears that set size 2 different-location distances were initially similar to set size 2 same-location distances through the beginning of the delay period,

before moving toward the hyperplane. It is unclear whether this inconclusive finding reflects the set size 2 different-location condition becoming different from both of the training conditions or more similar to the set size 4 same-location condition. However, given that set size 2 is within capacity for most individuals, it is possible that participants inadvertently processed placeholders presented in the same location as a



**Fig. 5.** Distance from the classification hyperplane for set size 2 same-location trials, set size 4 same-location trials, and set size 2 different-location trials across time for Experiments 1a (a) and 1b (b). Classification is trained on set size 2 and set size 4 same-location trials and tested on all three conditions. The dashed gray line indicates the hyperplane. At left, distances across time are shown; at right, mean distances during the delay period are shown, with significance markers comparing the set size 2 different-location distances to those of the training conditions. Note that the distances are in arbitrary units (a.u.). Significance markers (\*) indicate significant differences ( $p < .05$ ) in the distances from the hyperplane between pairs of conditions.

preceding target on set size 2 different-location trials, resulting in an intermediate load signal. Experiment 2 controls for stimulus energy and spatial attention without placeholders, avoiding this potential confound.

Together, it appears that the mvLoad analysis can discriminate the number of individuated items stored in working memory, regardless of the number of positions occupied by those items, though this decoding is not conclusive in all conditions.

### Discussion of Experiment 1

In Experiment 1, we found that alpha power was sensitive to the number of attended locations but not to the number of items stored in working memory. In contrast, mvLoad decoded the total number of items stored, regardless of the number of attended locations. In Experiment 2, we extended this pattern of results with a refined modeling approach to quantify both the locus

and precision of covert spatial attention. This allowed a more sensitive test of the influence of spatial attention on our multivariate measure of working memory load. We employed a spatial change-detection task using dot clouds as stimuli. Previous work has shown that people can approximately enumerate overlapping sets of dots in parallel, with a capacity of around three sets (Halberda et al., 2006). Therefore, we anticipated that participants would be able to identify overlapping dot clouds as distinct objects to be maintained in working memory. These stimuli also removed confounds found in previous work using *mvLoad*. The total number of dots could be controlled to equate stimulus energy without requiring irrelevant placeholders. In turn, eliminating placeholders avoids a potential confound between the number of relevant targets and the number of irrelevant objects to be ignored. Finally, the dot clouds afforded strong variations in the spatial extent and position of each item, enabling a clear test of whether substantial changes in covert spatial orienting had any influence on multivariate measures of working memory load.

## Experiment 2

### Method

**Participants.** Participants were recruited from the University of Chicago and the surrounding community. Thirty-three participants (25 female, 8 male; mean age = 25.6 years,  $SD = 3.7$  years) were recruited for this study and were compensated \$20 per hour for their participation. Four participants ended the session early because of scheduling conflicts, leaving 29 who completed a full data acquisition session. Of these, 6 participants were excluded for excessive eye movements or artifacts (see preprocessing and artifact rejection below), and 1 was excluded for chance performance in set size 2, leaving a total of 23 participants whose data was included for analysis (17 female, 6 male; mean age = 25.7 years,  $SD = 3.1$  years). Our target sample was 20 participants, as we assumed the difference in precision of spatial attention when attending single narrow clouds compared to attending single broad clouds would be the smallest effect size. This comparison is a conceptual replication of Feldmann-Wüstefeld and Awh's 2020 work; their Experiment 1 data included 22 participants. We overshot our target, as we completed preprocessing and rejection in batches.

Participants were between 18 and 35 years old, reported normal or corrected-to-normal visual acuity, and provided informed consent according to procedures approved by the University of Chicago Institutional Review Board. Participants were recruited via

online advertisements and flyers posted on the university campus.

**Apparatus.** Participants were tested in a dimly lit, electrically shielded chamber. For the first eight data sessions, stimuli were generated using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). For the remaining 25 data sessions, stimuli were generated using PsychoPy (Peirce et al., 2019). Participants viewed the stimuli on a gamma-corrected 24-in. LCD monitor (refresh rate = 120 Hz, resolution = 1,080 × 1,920 pixels) with their chins on a padded chin rest at a viewing distance of approximately 80 cm.

**Stimuli.** Displays consisted of a neutral gray background (RGB values = 127.5, 127.5, 127.5) containing a central fixation point and one or two sets of colored dots, referred to as *dot clouds*. In the initial acquisition sessions, the fixation point was a black cross (RGB values = 0, 0, 0) spanning .3 visual degrees. In the remaining sessions, a modified version of a validated fixation point (Thaler et al., 2013) was used containing a black circle spanning .5° degrees of visual angle across the diameter with a fixation cross matching the background gray overlaid on top and a central black circle spanning .15° visual degrees.

To generate the dot clouds, locations were drawn from an imaginary circular grid consisting of 40 radial columns and 10 circular rows. The innermost row began 1.75° from the center of the screen, and the outermost row ended 5.25° from the center of the screen. The 40 columns were divided into eight location bins, each containing five columns and spanning 45°. Dot clouds spanned either 1 or 3 bins (i.e., either 45° or 135°). In the case of a dot cloud spanning three bins, its location is defined as the central bin. For a given dot cloud, individual dot locations are chosen by randomly sampling cells from the circular grid in the location bins, with the following constraints. To ensure that each cloud spanned the bin completely, we assigned one dot each to a row in the outermost columns of its assigned bin, and the remaining dots were randomly distributed across the remaining cells. For the first cloud to be drawn (regardless of set size), a cell was reserved from each column that could potentially be an outermost column for the second cloud to be drawn. For the second cloud in set size 2 trials, those reserved cells were used if they were in fact on an outermost column. Otherwise, the above process was repeated for the second cloud, excluding already selected cells, and without reserving any additional cells. The two cloud locations and sizes were independently drawn on each set size 2 trial.

Dots in the innermost row were .25° and grew by .01° for each row further from fixation. Within their

cells, dots were randomly slightly jittered with the constraint that they could not cross into an adjacent cell.

The number of dots for a given cloud was drawn from a uniform distribution on each trial. For set size 1 trials, the number ranged from 12 to 48 dots inclusive; in set size 2 trials, the number ranged from 12 to 24 dots inclusive for each cloud (a total of 24 to 48 dots). Dot clouds were either blue (RGB values = 0, 0, 255) or a luminance-matched green (0, 52, 0). The color was randomly chosen on set size 1 trials, and the color of the probed cloud was randomly chosen on set size 2 trials, with the unprobed cloud being the other color.

Although encoding displays could consist of 1 or 2 dot clouds, probe displays consisted of only a single cloud. On no-change trials, the probed dot cloud was redrawn identically to the encoding display. On change trials, the dot cloud could change in two ways. First, the cloud could shift by one bin, either clockwise or counterclockwise, while maintaining its spatial configuration. Alternatively, the cloud could change in size, shifting from one bin width to three bin widths, or vice versa. In this latter case, the cloud was redrawn using the procedure described above, with the same number of dots as during the encoding display.

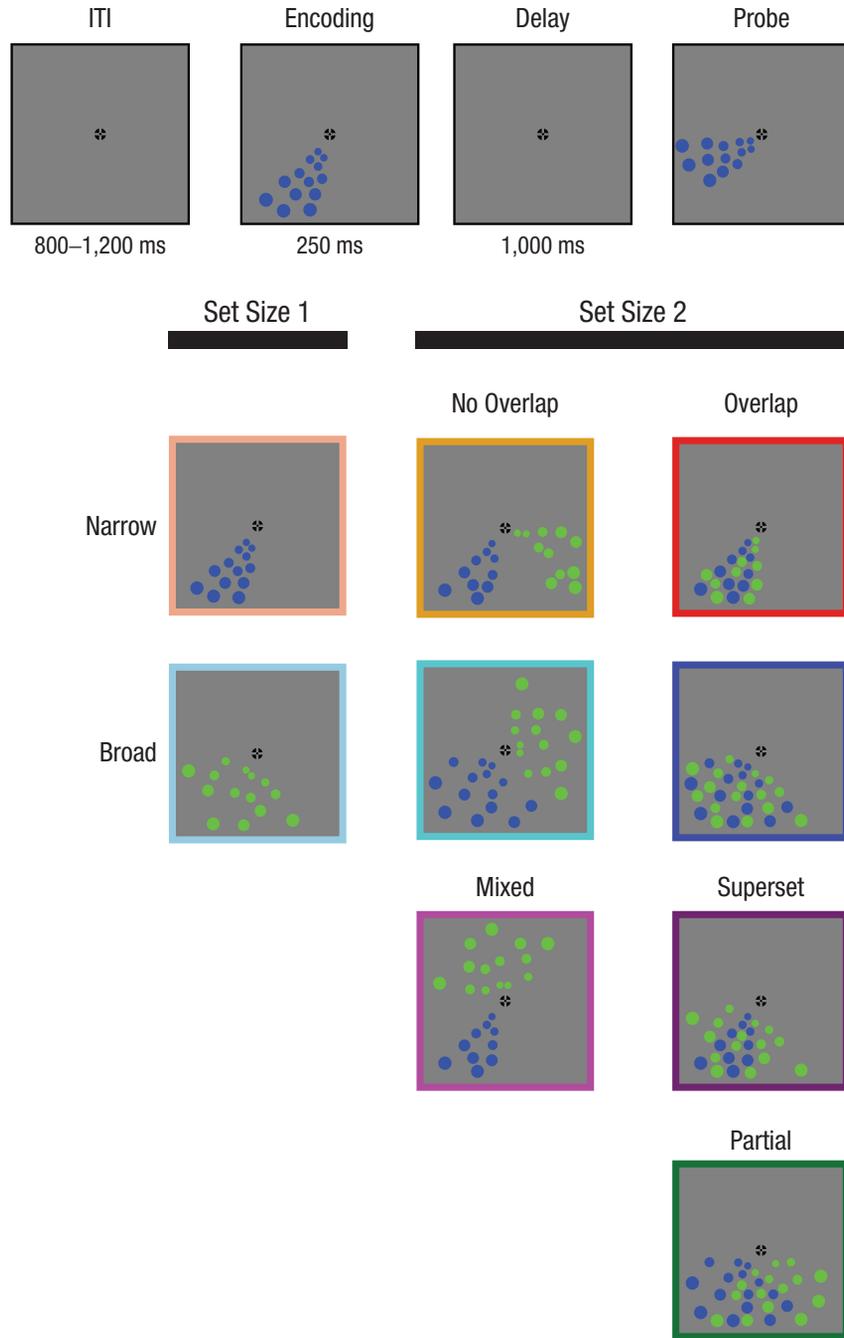
**Task procedure.** Participants completed a change-detection task based on the spatial configuration of dot clouds (Fig. 6). Trials consisted of a 250-ms encoding display in which participants would see one or two colored dot clouds placed among an imaginary ring encircling a fixation point (see above). Dot clouds were centered on one of eight 45° bins dividing the ring and spanned either one bin (45°) or 3 bins (135°). In set size 2 conditions, the dot clouds' locations were independent, allowing for no, partial, or complete spatial overlap, depending on the location and width of the two clouds. During overlap conditions, the individual dots of the clouds were interleaved. Clouds would either be luminance-matched blue or green.

A delay period lasting 1,000 ms would follow the encoding display, after which participants would be probed on the spatial configuration of a cloud. On 50% of trials there was no change, and the probed cloud would be redrawn identically to its configuration in the encoding display. On the remaining 50% of trials, the cloud could change in one of two ways. On 50% of change trials, clouds could shift by one bin, either clockwise or counterclockwise. On the remaining 50% of change trials, clouds could change in width, either broadening from one bin to three bins, or narrowing from three bins to one bin. Participants used the left and right arrow keys to respond, with the mapping to change and no-change being randomly assigned for each participant.

Participants began with practice blocks consisting of 10 trials each in which they were given immediate feedback after each response, and they completed practice blocks until they comprehended the task. During the test phase, participants completed 16 blocks of 96 trials for a total of 1,536 trials. After the first eight acquisition sessions, participants received real-time feedback for eye movements made during the encoding and delay periods of the trial (see below). In cases in which an eye movement was detected, the trial was aborted and a version of the trial (containing the same high-level information, but with the exact dot locations randomly redrawn) was appended to a set of makeup blocks at the end of the session. Acquisition sessions lasted approximately 4 hours, including preparation.

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

**Eye tracking.** We monitored gaze position using a desk-mounted EyeLink 1000 Plus infrared eye-tracking camera (SR Research). Gaze position was sampled at 1000 Hz. According to the manufacturer, this system provides spatial resolution of .01° of visual angle and average accuracy of 0.25° to 0.50° of visual angle. We calibrated the eye tracker before each test block and between trials during the blocks if necessary. For participants analyzed using PsychoPy, we ran a drift-correction procedure every six trials. (For the first two participants using PsychoPy, this drift-correction procedure was every 10 trials.) We drift corrected the eye-tracking data for each trial by subtracting the mean gaze position measured during a 200-ms window from -250 ms to -50 ms before the encoding display.



**Fig. 6.** Illustrative schematic trial from the dot cloud change-detection task (top) and exemplars for each of the nine conditions (bottom). In the dot cloud change-detection task, the cloud changed in position by shifting one bin clockwise. The alternative-change cases shift one bin counterclockwise and change in width (from spanning one bin to three bins, or vice versa). The exemplars shown at bottom are for each of the nine conditions across cloud size, set size, and forms of overlap in set size 2 conditions. The colored borders correspond to those used in the IEM and hyperplane results below. For illustrative purposes, the dots have been enlarged and the contrast between the colors has been amplified, relative to their actual display for the experiment. ITI = intertrial interval.

**Artifact rejection.** Our preprocessing pipeline combines functions from EEGLab (Delorme & Makeig, 2004), ERPLab (Lopez-Calderon & Luck, 2014), and custom

MATLAB scripts. We segmented the EEG data into epochs time-locked to the onset of the memory array (from 250 ms before until 1,650 ms after stimulus onset). We

baseline corrected the EEG data by subtracting mean voltage during a 200-ms window (from  $-250$  ms to  $-50$  ms before the encoding display) and examined the epoch for artifacts from the beginning of the baseline period ( $-250$  ms) to the onset of the test display ( $1,250$  ms). Eye movements, blinks, blocking, drift, and muscle artifacts were first detected by applying automatic criteria. After automatic detection, we visually inspected the segmented EEG data for artifacts (amplifier saturation, excessive muscle noise, and skin potentials) and the eye-tracking data for ocular artifacts (blinks, eye movements, and deviations in eye position from fixation), and we discarded any epochs contaminated by artifacts. Participants were excluded if they lacked 15 trials in any of the held-out test conditions (see below), resulting in four exclusions. On average, 9.5% of trials were rejected in the remaining 24 participants.

**Eye artifacts.** Using eye-tracking data, we rejected trials that contained ocular artifacts such as blinks, saccades, and deviations from fixation. Blinks were identified as any time points without a value and rejected. Saccades were identified as a  $.5^\circ$  shift in location between the first half and the second half of a 80-ms window, which slid across the epoch in steps of 10 ms. Movements from fixation were identified as a  $1^\circ$  shift from fixation at any time point. If both eyes were being tracked, an artifact had to be identified in both eyes. (The first 8 participants had only one eye tracked, so in those cases an artifact needed to be identified only for that eye.)

For 1 participant, eye-tracking data were not available, so EOG was used. Saccades and blinks were identified via a  $20\text{-}\mu\text{V}$  difference between the first half and the second half of a 150-ms window, which slid across the epoch in steps of 10 ms. Large movements were identified via a  $50\text{-}\mu\text{V}$  difference from baseline at any time point.

**EEG artifacts.** We checked for drift (e.g., skin potentials) by fitting a line to each channel. Trials were excluded if the line had a slope greater than a certain threshold (slope = 75, minimal  $r^2 = .3$ ). To check for muscle artifacts, we excluded trials with peak-to-peak activity greater than  $100\text{ }\mu\text{V}$  in a 250-ms window that was moved across the epoch in 20-ms steps. We also excluded trials that differed by  $60\text{ }\mu\text{V}$  between the first half and the second half of a 250-ms window. Last, we excluded trials with any value beyond a threshold of  $100\text{ }\mu\text{V}$ .

**Analyses.** To decode working memory load and spatial attention, we relied on two machine-learning-based approaches: mvLoad (Adam et al., 2020; Thyer et al., 2022) and inverted encoding models (IEMs). Both correct and incorrect trials were used during training and testing

(Feldmann-Wüstefeld & Awh, 2020; Thyer et al., 2022). For all analyses, set size 1 trials containing less than 24 dots were excluded. This ensured that the set size 1 and set size 2 trials examined in these analyses had the same range of dots (24–48) and central tendency (mean and median = 36), making the number of dots nondiagnostic of set size. Analyses were run using Python's Scikit-learn package (Pedregosa et al., 2011), along with custom Python scripts.

**Behavior.** As this was an unspeeded task, we focused on accuracy measures. To support comparison with the neural results below, we excluded trials containing artifacts and set size 1 trials containing fewer than 24 dots. We began by testing whether the mean choice accuracy was above chance in both set size conditions, using a two-tailed  $t$  test. All between-condition comparisons were completed as two-tailed repeated-measures  $t$  tests. As in Experiment 1, we computed BFs (with a standard Cauchy scale of .707) for all  $t$  tests to provide complementary evidence against the null ( $\text{BF}_{10}$ ), the reciprocal of which reflects evidence in favor of the null.

**IEMs.** To examine how spatial attention was affected across conditions, we applied inverted encoding models. Previous work has fitted IEMs to the topography of alpha power across the scalp to reconstruct a location-selective channel response function (CRF) reflecting the allocation of attention across space (Foster et al., 2017). This technique has been used to decode the location spatially attended during the delay period of working memory tasks, even when participants must internally transform the location in response to auditory cues (Günseli et al., 2022). In addition, researchers have found that the shape of the tuning function broadens with increasing number of relevant locations (Sutterer et al., 2019) and with increasing size of the relevant location (Feldmann-Wüstefeld & Awh, 2020).

We began by replicating these latter two findings. Both analyses relied on comparing the precision of the reconstructed CRFs between two conditions, so the overall architecture of the two analyses is the same. For a given pair of conditions (e.g., set size 1 vs. set size 2), we performed the following model-fitting and -testing procedure. First, we isolated activity within the alpha band (8–12 Hz) for each trial by applying a finite impulse response (FIR) filter implemented by the MNE-Python package (Gramfort et al., 2013). We then extracted the complex analytic signal of the isolated activity for each via a Hilbert transform and squared its complex magnitude to compute alpha power. Before this process, we set time points from the test period to 0 to avoid contaminating the end of the delay period. Within each trial, we grouped time points into

nonoverlapping 25-ms windows, and we averaged alpha power within each time window to increase SNR. We completed the following analysis at each time window independently, using all electrodes.

We grouped trials within each condition by the location bin of the to-be-probed cloud. For clouds spanning three bins, this was the middle bin. Across the 16 groupings (eight locations, two conditions), we identified the grouping with the fewest trials and randomly trimmed the other groupings to match the minimum. This equated the number of trials across locations and conditions. From this, we randomly sorted the data for each condition into three folds, each containing one third of the trials at each location for each condition, and we averaged the data within each fold. This produced a matrix of shape (Number of Locations  $\times$  Number of Electrodes  $\times$  Number of Time Windows) for each of the three folds for each condition.

Following the standard IEM approach, we assumed that alpha power at each electrode is the result of weighted contributions from various subpopulations of cells (hereafter referred to as *channels*). Each channel has a preferred location with a graded response profile, so its activity is highest at its preferred location and decreases with increasing distance from its preferred location. We modeled each channel's response function as a half cosine raised to the 25th power, with one channel for each location bin. From this, we computed a response for each channel to each location. This produced a matrix of shape (Number of Locations  $\times$  Number of Channels), containing the response of each channel for each of the locations.

In a  $k$ -fold procedure, we fitted a linear regression model to two folds for each condition (four folds total) to predict alpha power at each electrode from channel responses for each location at each time point. At a given time point, this took the following mathematic form—

$$P = WC \quad (1)$$

—where  $P$  is the matrix of alpha power across the four folds of shape (Number of Electrodes  $\times$  Number of Locations  $\times$  4),  $C$  is a matrix describing the channel responses of shape (Number of Channels  $\times$  Number of Locations  $\times$  4), and  $W$  is a weight matrix of shape (Number of Electrodes  $\times$  Number of Channels), which is found via the linear regression fitting procedure.  $W$  describes the individual contributions of each channel to alpha power at each electrode. Once the weights were fitted, we inverted the weights of the model ( $W^{-1}$ ) to produce a matrix that produces channel responses from the topography of alpha power. We applied  $W^{-1}$  to the held-out fold of each condition separately to reconstruct CRFs at each location in the held-out folds.

Within a condition, we rotated the CRFs for each location to be centered on  $0^\circ$  and averaged them together. We repeated this process for each of the three folds and averaged the resulting CRFs together. Because some trials were randomly trimmed to balance the groupings, we repeated this whole process 100 times, with random trimming and random grouping into folds, and averaged the CRFs across the 100 permutations together.

We then computed the slope of each condition's CRF. To do so, we folded the CRFs in half and averaged pairs of values equidistant from the center (e.g., the estimated responses for channels preferring locations  $+45^\circ$  and  $-45^\circ$  from 0 were averaged together). We fitted a line to this folded CRF and recorded the slope. To increase power, we averaged slope estimates across the delay period, excluding the final 100 ms to avoid bleed-over from the censored test time points. Slopes were compared with a repeated-measures  $t$  test, along with a BF to provide complementary evidence against or for the null.

In addition to replicating the previous two findings, we tested the slopes of two conditions that varied in set size, but not in the total attended area. Specifically, we repeated the above process to compare the CRF slope between SS1 trials containing a broad cloud and SS2 trials containing one broad cloud (SS1 broad overlap), with one narrow cloud superset within it (SS2 superset), both of which contained clouds covering three bins. (We note that although SS2 broad overlap also contains clouds spanning three bins, there is much more data for SS2 superset compared to SS2 broad overlap, allowing us to produce much more reliable slope estimates.)

**Multivariate load decoding (*mvLoad*).** Set size was decoded within participants. First, we replicated previous work decoding set size, while ignoring conditional differences like cloud size and overlap. To increase the signal-to-noise ratio, we randomly assigned trials of the same set size into groups of 20 and averaged within each group. These grouped and averaged time series were the foundation of our testing and training procedure. As above, we divided the epoch into nonoverlapping 25-ms windows and found the average voltage for each electrode in each window.

We tested classification using a cross-validation procedure at each time point. At a given time point, we selected 80% of the grouped time series as training data and reserved the remaining 20% for testing, stratifying by set size. We trimmed the training data to equate the number of grouped time series for each set size; we then  $z$ -standardized the training set. We fitted a logistic regression model to the training data to predict set size.

We then rescaled the test data using the mean and standard deviation of the training set before applying the fitted model and recording the accuracy of its predictions. Last, we generated an empirical null accuracy by testing the model on the same test data, but with randomly shuffled labels. This procedure was repeated 1,000 times (including the random grouping and averaging to improve SNR and the trimming to balance the training data) for each time point, for each participant, and the average accuracy and shuffled accuracy were recorded.

For the main decoding approach, we tested classification accuracy for each 25-ms time window using a one-tailed  $t$  test, comparing the test accuracy against the shuffled test accuracy. To correct for multiple comparisons, we applied the Benjamini-Hochberg procedure to control the false-discovery rate (FDR) at .05. We also computed the BF at each time window. However, as the BF<sub>10</sub>s are aligned with the statistical results, we do not present them for simplicity.

### ***Predicting load regardless of spatial conditions.***

To assess whether load can be decoded regardless of the number and shape of spatial envelopes, we fitted a second logistic regression model at each 25-ms window. For this analysis, trials were labeled both by set size and by spatial conditions—specifically, by the size of the dot clouds and overlap. Set size 1 trials were labeled as either narrow or broad. Set size 2 trials were labeled by cloud size (both narrow, both broad, or mixed), and whether there was complete overlap (both clouds the same size and in the same location), superset overlap (one cloud is broad and subsuming a second, narrow cloud), partial overlap (both clouds are broad and overlap in one or two bins), or no overlap. The no-overlap trials were distinguished by the size of the clouds: Both clouds could be narrow, both clouds could be broad, or they could consist of one narrow cloud and one broad cloud.

We randomly selected groups of 15 trials to average together within each condition. For training, we selected set size 1 trials with broad clouds (SS1 broad overlap) and set size 2 trials in which one cloud subsumed another (SS2 superset). Both of these trial types spanned the same spatial area (three bins) and contained the same number of dots on average, so differences in these conditions should be driven by the number of perceived dot clouds held in working memory. Note that this comparison would also work for SS1 broad overlap trials and SS2 trials with both clouds being broad and perfectly overlapping (SS2 broad overlap), but there were many fewer SS2 broad overlap trials than SS2 superset trials, making model fits much less stable.

To train the model, we randomly reserved one group each from SS1 broad overlap and from SS2 superset for

testing; we used the remaining groupings for each of the two conditions for training. If one condition contained more groups, then we would trim a random subset of groupings to balance the two conditions before  $z$ -standardizing the final training set and fitting the model. We applied the fitted model to the two held-out groupings along with all groupings for the remaining conditions (e.g., SS1 narrow overlap, SS2 no overlap), standardized using the mean and standard deviation of the training set. For each test grouping, we recorded the distance from the model's classification hyperplane set by the training data. This process was repeated 1,000 times (including the random grouping and averaging within each condition), and the final distances were recorded for each participant in each time window.

To increase power, results were averaged across the delay period, using the same time window as above (250 ms to 1,150 ms after stimulus onset). The distances of the training conditions from the hyperplane were compared against 0 with a two-tailed  $t$  test. Each held-out condition was compared against each training condition using a two-tailed repeated-measures  $t$  test. We also computed BFs describing evidence against the null (BF<sub>10</sub>) for each of these tests.

***Representational similarity analysis.*** Although decoding methods can be applied to distinguish between pairs of conditions, it can be difficult to determine what signals the model is seizing upon. For example (as will be discussed in the Results section), it is possible that SS2 superset contains an additional signal reflecting a cognitive mechanism (or set of mechanisms) responding to overlapping clouds (i.e., a separation process, or interference processing). A model trained to separate SS1 broad overlap from SS2 superset could make use of both (or either) the set size signal and this overlap-related signal to separate the conditions. To address this issue, we applied RSA to simultaneously estimate the influence of multiple signals on EEG voltage patterns.

RSA provides a means for exploring the similarity structure across conditions of an experiment, and the factors that contribute to said structure (Kriegeskorte et al., 2008). The base assumption is that conditions that are more similar to each other should also produce more similar neural signatures, measured via some distance measure. To return to our example above, we can make distinct predictions on the structure of distances between conditions on the basis of our two hypothetical signals. If there is a pure set-size signal, we should expect all pairs of conditions that differ in set size to be more neurally distinct from each other and all pairs of conditions with the same set size to be more neurally similar to each other. If there is some overlap-related process, then pairs of conditions that either lack or

contain overlap should be neurally similar to each other, and pairs of trials in which one contains overlap and the other does not should be more neurally distinct. Critically, the influence of different factors is not mutually exclusive, and RSA provides a way to simultaneously estimate the contributions of multiple factors to the similarity structure of the neural signals. Thus, we tested for the presence of these signals using the following RSA procedure.

To increase SNR, we binned trials in nonoverlapping 50-ms windows and averaged the values within each window. The distance between each pair of conditions was measured using a linear discriminant contrast within each participant (also known as *cross-validated mabalinobis distance*). We chose linear discriminant contrast, rather than another measure, for two reasons. First, linear discriminant contrast has been shown to be highly reliable (Walther et al., 2016). Second, although previous work applying RSA has used correlation as a distance measure (Kiat et al., 2022), correlation is scale invariant; in other words, correlation would not distinguish between a vector representing some process ( $v$ ) and a vector representing twice that process ( $2 * v$ ). However, it is very possible that changes in EEG activity between set sizes are reflected by the scaling of a given pattern, rather than a shift in topography. For example, the contralateral delay activity becomes increasingly negative with set size, at least until it plateaus around three or four items (Vogel & Machizawa, 2004). In addition, previous work using mvLoad has identified increasing negativity in posterior electrodes and increasing positivity in frontal electrodes with increasing set size (Thyer et al., 2022). Because our experiment is subcapacity, it seems reasonable to assume that changes in set size could be reflected in scaling. Linear discriminant contrast is a measure of the euclidean distance between two conditions, and therefore can still detect changes in scale.

To compute the representational dissimilarity matrix (RDM) for a given participant and time point, we first randomly split the data into training and test sets. For each of the nine conditions, half of the trials were assigned as training data and half as test. For each condition, we found the mean train-set value and the mean test-set value at each electrode. To adjust for the covariance between electrodes, we computed the covariance within the training set by first demeaning each trial by its condition’s mean and then computing the covariance across trials. This covariance estimate was regularized using the Ledoit-Wolf procedure (Ledoit & Wolf, 2004; Walther et al., 2016). The distance between each pair of conditions ( $i$  and  $j$ ) was computed as

$$d_{LDC}^2(i, j) = (m_i - m_j)_{train} * \Sigma_{train}^{-1} * (m_i - m_j)_{test}, \quad (2)$$

where  $m_i$  and  $m_j$  are the vectors of mean values for conditions  $i$  and  $j$  in the train or test sets and  $\Sigma_{train}^{-1}$  is the inverse of the regularized covariance matrix (Walther et al., 2016). To produce stable distance estimates, we repeated this procedure with 10,000 train-test splits, and we averaged across the resulting RDMs for each participant. We computed an RDM for each participant at each time window.

We next compared each participant’s empirical RDM against a set of four theoretical RDMs, enabling a quantification of the amount of variance in the EEG signal explained by each (putative) cognitive process. First, our set-size RDM predicted that trials with the same set size should look more similar to each other, and distinct from trials with the other set size. Our overlap-related process RDM predicted that trials containing overlap should look similar to each other, trials without overlap should look similar to each other, and pairs of trials crossing these sets should look different. We also included two additional RDMs related to spatial attention. First, we included an RDM of the absolute differences in the total amount of attended area between each pair of conditions. This RDM reflects the assumptions that trials with the same attended area should look similar to each other, trials with different attended areas should look different from each other, and that that difference should grow with the difference in size of attended areas (e.g., trials covering six bins and trials covering one bin should look more different from one another than trials covering four bins and trials covering three bins). Note that attended area also scales with the average density of dots (because the number of dots was equated across set sizes), so this RDM could also be considered to capture density. Last, we added an RDM for the number of noncontiguous spatial locations or spatial envelopes attended. This assumed that both set size 1 and set size 2 conditions with overlap contained only one relevant spatial location. For set size 2 conditions without overlap, we computed the expected number of noncontiguous locations on the basis of the sizes of the two clouds. For example, with two narrow clouds, they would be contiguous (number of locations = 1) on 2/7 of trials and noncontiguous (number of locations = 2) on 5/7 of trials, so the expected value is 12/7. This RDM was computed as the absolute difference in the expected number of noncontiguous attended locations between each pair of conditions. These four theoretical RDMs were minimally related, with the largest positive correlation ( $r = .23$ ) between the attended area RDM and the number of attended locations RDM, and the largest negative correlation ( $r = -.17$ ) between the attended area RDM and the overlap-related process RDM.

**Table 1.** Mean Accuracies and Standard Deviations for Each Condition

Set Size 1		Set Size 2						
Narrow	Broad	Narrow overlap	Broad overlap	Superset overlap	Partial overlap	No overlap, mixed	No overlap, narrow	No overlap, broad
97.0 (5.1)	96.5 (5.7)	91.9 (11.3)	90.0 (12.4)	86.2 (9.5)	83.6 (9.1)	89.4 (8.6)	92.0 (7.8)	88.5 (8.6)

At each time point, we used a rank regression procedure to account for the independent contribution of each theoretical factor (Iman & Conover, 1979; Kiat et al., 2022). For each condition, we computed the semipartial rank correlation by comparing the  $R^2$  of a full model predicting the rank of empirical distances from the ranked distances of all theoretical factors and comparing this against the  $R^2$  of a model excluding the condition. We then multiplied the square root of the difference in  $R^2$ s by the sign of the condition's coefficient in the full model to produce the condition's semipartial correlation. As in Kiat et al. (2022), we choose rank correlation because we did not assume a linear relationship between any of our theoretical factors and the observed condition distances. We tested these correlations at the group level using a Wilcoxon signed-rank test against zero. We tested each time bin with FDR correction using the Benjamini-Hochberg procedure. We also tested each factor's semipartial correlation after averaging across the delay period for maximal power. We also computed BFs for the Wilcoxon tests using the procedure outlined in van Doorn et al. (2020) and the publicly available R code. For the tests at each time window, we sampled 5,000 points from the posterior for each factor. Because these were highly aligned with the statistical tests, we omit them for simplicity. For the tests using data averaged across the delay period, we sampled 10,000 points from the posterior for each factor.

**Controls for eye movements.** We examined whether our IEM analysis was detecting changes in eye position rather than the deployment of covert attention to the dot clouds. This analysis relied on the 22 participants with eye-tracking data. For each participant, eye gaze was drift corrected using baseline data from -250 ms to 50 ms before stimulus onset. The average gaze location during the trial was found and converted to a location bin. We then repeated the set size 1 versus set size 2 IEM comparison, using the eye-gaze bins as the target locations rather than the cloud locations. Because eye-gaze bins are not controlled by experimenters, it is possible that individuals expressed a large amount of bias, favoring some locations over others. This would result in some

bins rarely being set as the target and could also result in a smaller amount of data being used in the model after trimming. This lack of data could potentially explain any unreliable CRF reconstructions. To account for this, we also repeated the set size 1 versus set size 2 IEM comparison while setting the number of trials per cell to be equal to that of the eye-gaze model.

We also examined whether load could be decoded across the delay period using eye gaze information from participants. For each participant, we fitted a logistic regression model to classify set size using all eye data available for that participant. The process was identical to the set-size decoding process above. We set a delay-period decoding threshold of 60% accuracy for whether a participant might have informative eye movements; this process identified 4 participants (mean delay accuracies: 71.7%, 65.8%, 62.0%, and 62.2%). We then repeated all of the above analyses excluding those 4 participants to confirm that no patterns of results changed.

## Results

**Behavior.** Participants performed the task well, with an average accuracy of 91.9% ( $SD = 6.42\%$ ). Participants were more accurate for set size 1 trials (mean accuracy = 96.7%,  $SD = 5.34\%$ ) than for set size 2 trials (mean accuracy = 88.7%,  $SD = .08\%$ ; difference = 8.1%),  $t(22) = 6.67$ ,  $p = 1e-6$ ,  $BF_{10} = 17,280$ ,  $d = 1.171$ .

In addition, within set size 2 trials, participants were worse for trials containing overlap (mean accuracy = 86.1%,  $SD = 8.7\%$ ) than for trials without overlap (mean accuracy = 90.2%,  $SD = 8.1\%$ ),  $t(22) = 5.29$ ,  $p = 2.6e-5$ ,  $BF_{10} = 911.6$ ,  $d = 0.482$ . It is worth noting that these trial types differ in a few ways. In addition to the presence or absence of overlap, they also differ in the total attended area (because clouds without overlap span more bins on average than those with overlap) and in the number of noncontiguous spatial locations that should be attended to. We will return to these differences in the sections on multivariate load decoding and representational similarity analysis. Table 1 provides a description of the mean accuracies and standard deviations across each condition.

**Inverted encoding models.** We began by replicating previous work, which found that spatial attention, as measured by CRF slope, is deployed less precisely as the number of items stored increases (Sutterer et al., 2019). We fitted encoding models to the total alpha power SS1 and SS2 trials with an equal number of trials per location bin and set size. For SS2 trials, the to-be-probed dot cloud was arbitrarily selected as the relevant location. We then inverted the models and applied them to held-out trials for each condition separately, and we reconstructed the CRFs. We folded each CRF in half and fitted a line, recording the slope as a measure of precision of spatial attention. To increase power, we averaged slopes across the delay period within each participant, excluding the last 100 ms to avoid bleedover of information from the test period. We found that CRFs for SS1 trials had a significantly higher slope than SS2 trials,  $t(22) = 5.66$ ,  $p = 1.1e-5$ ,  $BF_{10} = 2,033.46$ ,  $d = 0.872$  (see Fig. 7, top).

We next assessed whether our task design was sensitive enough to capture previously identified changes in precision of spatial attention in response to the size of the relevant spatial area (Feldmann-Wüstefeld & Awh, 2020). We repeated the above analysis, but compared trials containing set size 1 narrow clouds (SS1 narrow overlap) against trials containing set size 1 broad clouds (SS1 broad overlap). In line with past findings, CRFs during the delay period were more precise with one narrow cloud than with one broad cloud,  $t(22) = 3.32$ ,  $p = .003$ ,  $BF_{10} = 13.20$ ,  $d = 0.375$  (see Fig. 7, middle).

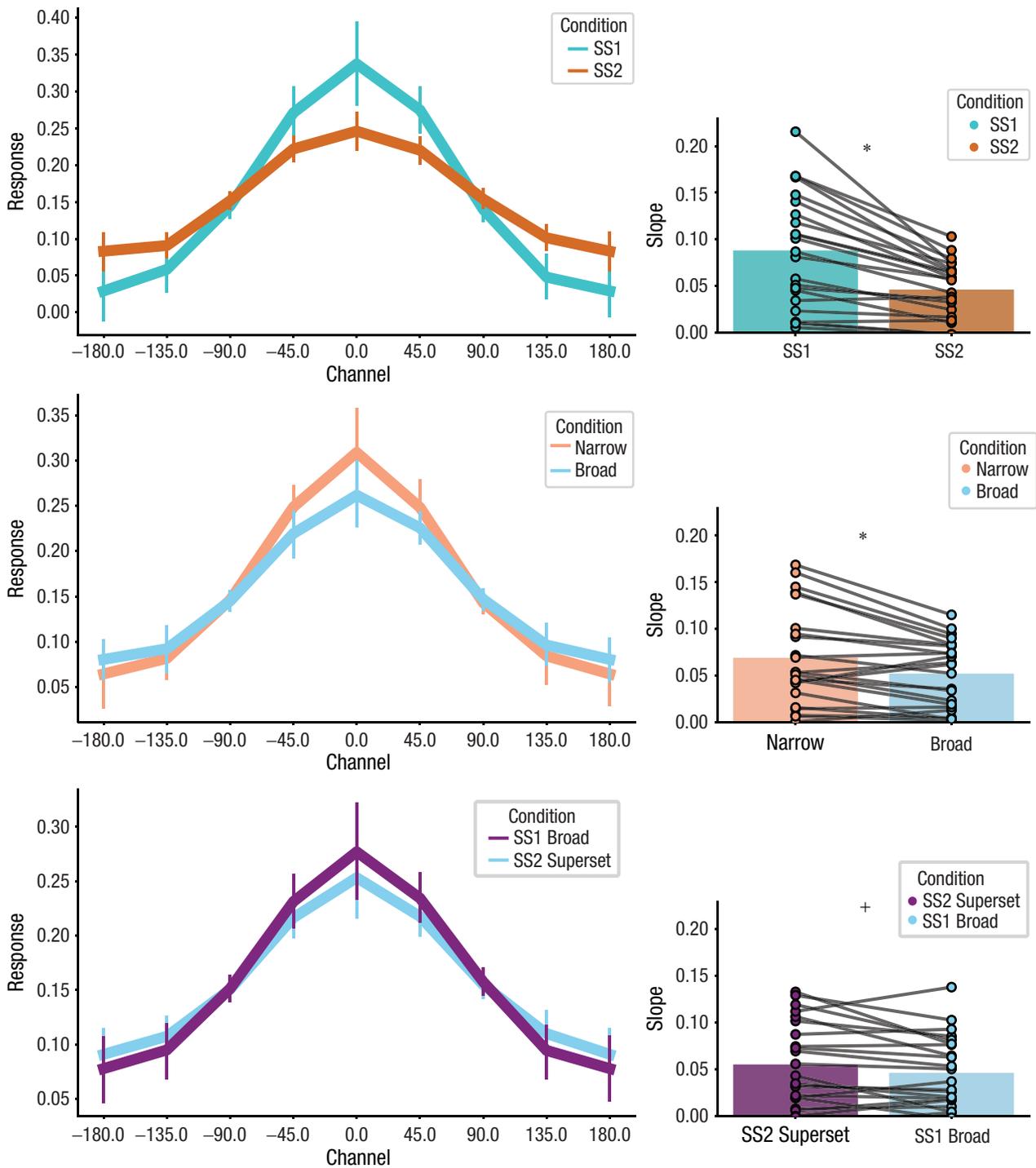
One possibility is that this difference in the precision of CRFs is driven by cross-hemifield effects. Specifically, the broad clouds span the vertical midline when centered on location bins next to the midline (four of the eight possible locations) and may be more likely to recruit cross-hemisphere processing, whereas the narrow clouds never do. To test this, we split the CRFs for both conditions into locations for which broad clouds would span the vertical midline and locations for which they would not; we then aligned, averaged, and computed the slope (see Fig. S1 in the Supplemental Material available online). An ANOVA found a main effect of condition,  $F(1, 22) = 10.34$ ,  $p = .004$ ,  $\eta_g^2 = .033$ , but no main effect of location nor an interaction (both  $F$ s  $< 3.1$ ,  $p > .05$ ). A Bayesian repeated-measures ANOVA also supported a model with only a main effect for condition ( $BF_M = 2.54$ , next highest = 2.24;  $BF_{10} = 12.26$ ). Thus, broad clouds produce broader CRFs than narrow clouds at every location.

Although previous analyses have suggested that spatial attention is deployed with less precision with increasing set size, the work has typically conflated the number of relevant objects with the number of relevant locations. In the current study, however, we had the opportunity to compare a pair of conditions that varied in set size while the relevant spatial envelope was held

constant. We compared SS1 broad overlap trials against set size 2 trials containing a superset overlap pattern—that is, one broad cloud with one narrow cloud within one of the three bins occupied by the broad cloud (SS2 superset). We followed the above analyses, with the center of the broad cloud being the target location for the SS2 superset condition. These conditions are matched on the total relevant spatial area and differ only in the number of perceivable objects. Therefore, we predicted that IEMs would show no difference between these two conditions. We found that there was a trend toward SS2 superset slopes being steeper than SS1 broad overlap slopes, but this effect was not significant,  $t(22) = 2.04$ ,  $p = .053$ ,  $BF_{10} = 1.26$ ,  $d = 0.211$  (Fig. 7, bottom). Note that the size of this marginal effect was very small, equaling about half the effect found when comparing set size 1 narrow and broad clouds and less than a quarter of the effect found when comparing set size 1 and set size 2 across all conditions.

Though the slopes between SS1 broad overlap and SS2 superset were not significantly different when averaged across the whole delay period, visual inspection of the slopes suggests that the conditions started to diverge toward the end of the delay period (Fig. 8). To test this, we divided the delay period into first and second halves and ran a repeated-measures ANOVA with two factors: condition (SS1 broad overlap and SS2 superset) and delay half (first and second). Although there was a main effect of delay half,  $F(1, 22) = 4.82$ ,  $p = .039$ ,  $\eta_g^2 = .025$ , the effect of condition was only trending toward significance,  $F(1, 22) = 4.28$ ,  $p = .051$ ,  $\eta_g^2 = .010$ , as was the interaction between delay half and condition,  $F(1, 22) = 3.13$ ,  $p = .091$ ,  $\eta_g^2 = .0033$ . A Bayesian repeated-measures ANOVA provided different results, supporting a model with main effects for delay half and condition, though the evidence was weak ( $BF_M = 1.65$ , next highest = 1.41;  $BF_{10} = 2.92$ ). Together, it appears that the precision of spatial attention is minimally affected by the number of dot clouds within a fixed spatial area. Thus, although multiple past studies have documented reduced spatial selectivity as the number of items encoded into working memory increases (e.g., Sprague et al., 2016; Sutterer et al., 2019), our findings suggest that this empirical pattern may have been driven by changes in the spatial extent of relevant positions rather than by the effects of storing additional items. This is in line with past findings of a tight relationship between working memory and attention (Awh et al., 2006; Awh & Jonides, 2001; Chun, 2011).

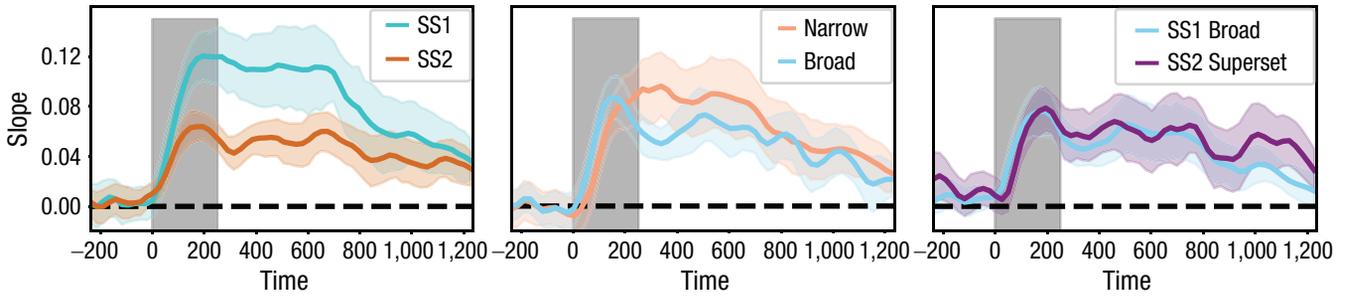
**Multivariate load decoding.** As in previous work, we found reliable decoding of set size across the delay period (Fig. 9; mean delay accuracy = 0.68,  $SD = 0.052$ ). This decoding accuracy was significant despite matching the number of dots and luminance of the two colors,



**Fig. 7.** IEM results, averaged across the delay period (250 ms–1,150 ms). At left are reconstructed CRFs; at right an average slope is illustrated, with individual participants superimposed as dots with lines connecting a given individual’s CRF slopes across conditions. Set size 1 and set size 2 are shown at the top; narrow and broad clouds within set size 1 in the middle; and set size 1 broad clouds and set size 2 superset clouds at the bottom. Error bars on the reconstructed CRFs are 95% bootstrapped confidence intervals of the response for each channel. Asterisks (\*) and crosses (+) indicate significance values  $p < .05$  and  $p < .1$ , respectively.

minimizing the impact of stimulus energy. However, other confounds remained. For example, SS2 trials covered a larger spatial area on average, and the classifier could make use of this difference to separate the two conditions.

To control for this, we tested whether a classifier trained on conditions with matched spatial areas could still decode set size, and we compared its predictions to held-out conditions varying in spatial area. First, we



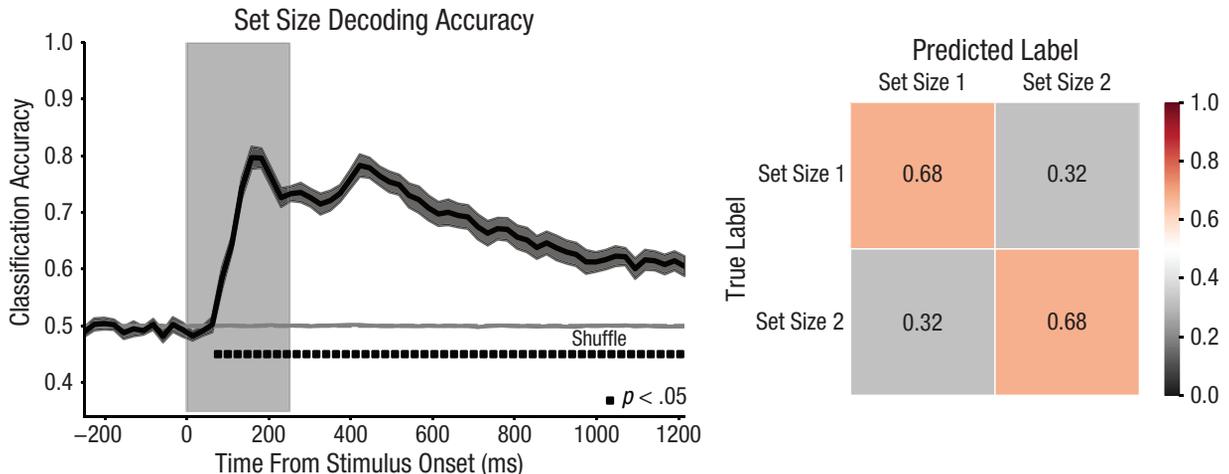
**Fig. 8.** Channel response function (CRF) slopes across time. At left, set size 1 (SS1) is compared with set size 2 (SS2); in the middle, SS1 narrow overlap is compared with SS1 broad overlap; and at right, SS1 broad overlap is compared with SS2 superset. The light gray shaded region indicates the duration of stimulus array. The colored shaded intervals reflect bootstrapped 95% confidence intervals of the slopes. Note that comparisons between conditions are within-participant comparisons, so the confidence intervals presented here do not indicate whether conditions are significantly different from one another.

fitted a model to predict load from SS1 broad overlap and SS2 superset conditions, both of which spanned three location bins. We then found the predictions of the model for held-out trials of these conditions by computing their distance from the model's decision boundary, and we averaged these values across the delay period (Fig. 10). We found that held-out trials for both conditions were a significant distance from the decision boundary and in opposite directions of the boundary—SS1 broad overlap:  $t(22) = -8.55$ ,  $p = 1.9e-8$ ,  $BF_{10} = 69,260.0$ ,  $d = 1.78$ ; SS2 superset:  $t(22) = 8.10$ ,  $p = 4.79e-8$ ,  $BF_{10} = 29,730.0$ ,  $d = 1.69$ . We next tested the predictions of the model to held-out conditions by finding their distance to the hyperplane, averaged across the delay period. We tested the following conditions: SS1 narrow overlap, SS2 with two overlapping narrow clouds (SS2 narrow overlap), SS2 with two overlapping broad clouds (SS2 broad overlap), SS2 with two broad, partially overlapping clouds (SS2 partial overlap), and SS2 with no overlap, broken into three size conditions:

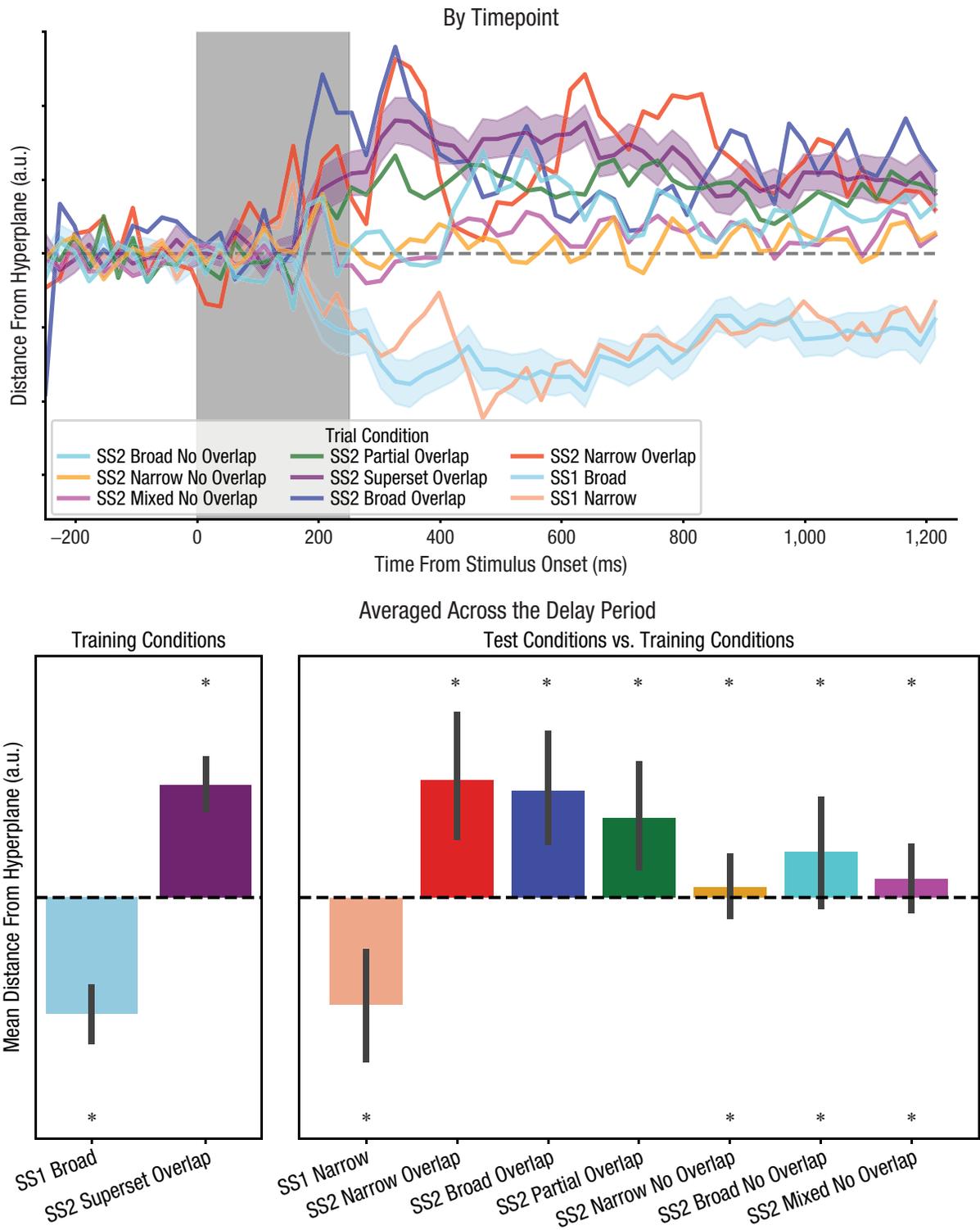
both broad (SS2 broad no overlap), both narrow (SS2 narrow no overlap), or one broad and one narrow (SS2 mixed no overlap). We compared the distance estimates for each of these held-out conditions against the test distances of the two training conditions.

We found that the mean distance across the delay period for SS1 narrow trials was significantly different from that of SS2 superset,  $t(22) = 5.67$ ,  $p = 1.1e-5$ ,  $BF_{10} = 2,072.12$ ,  $d = 2.03$ , but not significantly different from that of SS1 broad overlap,  $t(22) = 0.38$ ,  $p = .71$ ,  $BF_{10} = 0.233$ ,  $d = 0.085$ . In contrast, the distances for SS2 broad overlap, SS2 narrow overlap, and SS2 partial overlap showed the opposite pattern: Each was significantly different from SS1 broad overlap (all  $t$ s  $> 5$ , all  $p$ s  $< .0001$ ,  $BF_{10}$ s  $> 1,000$ ), and not significantly different from SS2 superset (all  $t$ s  $< 1.6$ , all  $p$ s  $> .1$ ,  $BF_{10}$ s  $< 1$ ).

This set of results suggests that the trained model can decode set size regardless of the size of the attended area. However, SS2 conditions without overlap were not as cleanly decoded. All three conditions without



**Fig. 9.** Set size decoding accuracy across the trial. At left, the decoding accuracy across time is shown with significance values for each time window. The light gray shaded region indicates the duration of stimulus array. The dark gray shaded interval about the average time course reflects the bootstrapped 95% confidence interval. The gray line varying about 0.5 indicates accuracy for shuffled labels. At right, the confusion matrix for the model is averaged across the delay period.



**Fig. 10.** Distance from the hyperplane across conditions. At top, the time series of distances to test trials across conditions is shown. For the two conditions that were used as the training set, bootstrapped 95% confidence intervals are included. The light gray shaded region indicates the duration of stimulus array. At bottom, the average distances from the hyperplane across the delay period are shown. On the left side, asterisks indicate significant differences from the hyperplane (0). On the right side, asterisks indicate significance differences from the training conditions on the opposite side of the hyperplane. Asterisks above indicate differences between a given condition and SS1 broad overlap (i.e., it looks more like SS2). Asterisks below indicate significant differences between a given condition and SS2 superset (i.e., it looks more like set size 1). Note that the distances are in arbitrary units (a.u.).

overlap were numerically on the SS2 side of the hyperplane, but all three conditions were significantly different from both training conditions (all  $t$ s > 2.5 and  $p$ s < .02). The pattern of distances was qualitatively identical between the first half (250–700 ms) and the second half (700–1,150 ms) of the delay period. Below, we offer a straightforward explanation for why these conditions were less confidently classified as SS2, with an eye toward EEG signals that track spatial overlap.

Overall, we found that that classifier could separate set size regardless of attended area throughout the delay period. However, the distances of some held-out set size 2 conditions were closer to the decision boundary than we predicted and significantly different from both training conditions. There are a few possible reasons for a condition's predictions being near the hyperplane. First, a condition may have too few trials, making it difficult to produce a reliable distance estimate; however, this is unlikely. Conditions with the fewest number of trials (SS2 narrow overall, and SS2 broad overlap, each making up 24 trials or 1.5625% of all trials before artifact rejection) were as far from the hyperplane as the training conditions, which contained many more trials, whereas the two conditions closest to the hyperplane accounted for 168 trials and 240 trials, or 10.9375% and 15.625%.

One might argue that these results could be driven by differences in effort across trials. Under this explanation, participants used less effort for SS1 broad overlap than for SS2 superset trials, and the remaining conditions are sorted on the basis of the amount of effort participants expended. Although we cannot measure effort directly, we can look at task performance as a measure of the difficulty of each condition, which may modulate effort (and if it does not, then effort is not a confound). To assess whether difficulty (and therefore effort) affected these results, we ran the following post hoc analysis. The assumption is that, if the hyperplane results are driven by effort, then conditions with similar hyperplane distances should also have similar accuracies. We compared participants' accuracy between each training condition and those conditions whose distances from the hyperplane were not significantly different from that training condition. SS1 broad overlap and SS1 narrow overlap did not significantly differ in distance from the hyperplane (see above) nor in accuracy,  $t(22) = 1.36$ ,  $p = .19$ ,  $BF_{10} = 0.493$ ,  $d = 0.0996$ —which is potentially in line with the possibility that effort and difficulty are driving the hyperplane results—but this hypothesis does not bear out when considering comparisons against SS2 superset. In the hyperplane distances, SS2 superset was not significantly different from SS2 narrow overlap ( $p = .87$ ,  $BF_{10} = 0.219$ ), SS2 broad overlap ( $p = .84$ ,  $BF_{10} = 0.305$ ), nor SS2 partial overlap ( $p = .14$ ,  $BF_{10} = 0.570$ ). However, accuracy between SS2 superset and these conditions was either

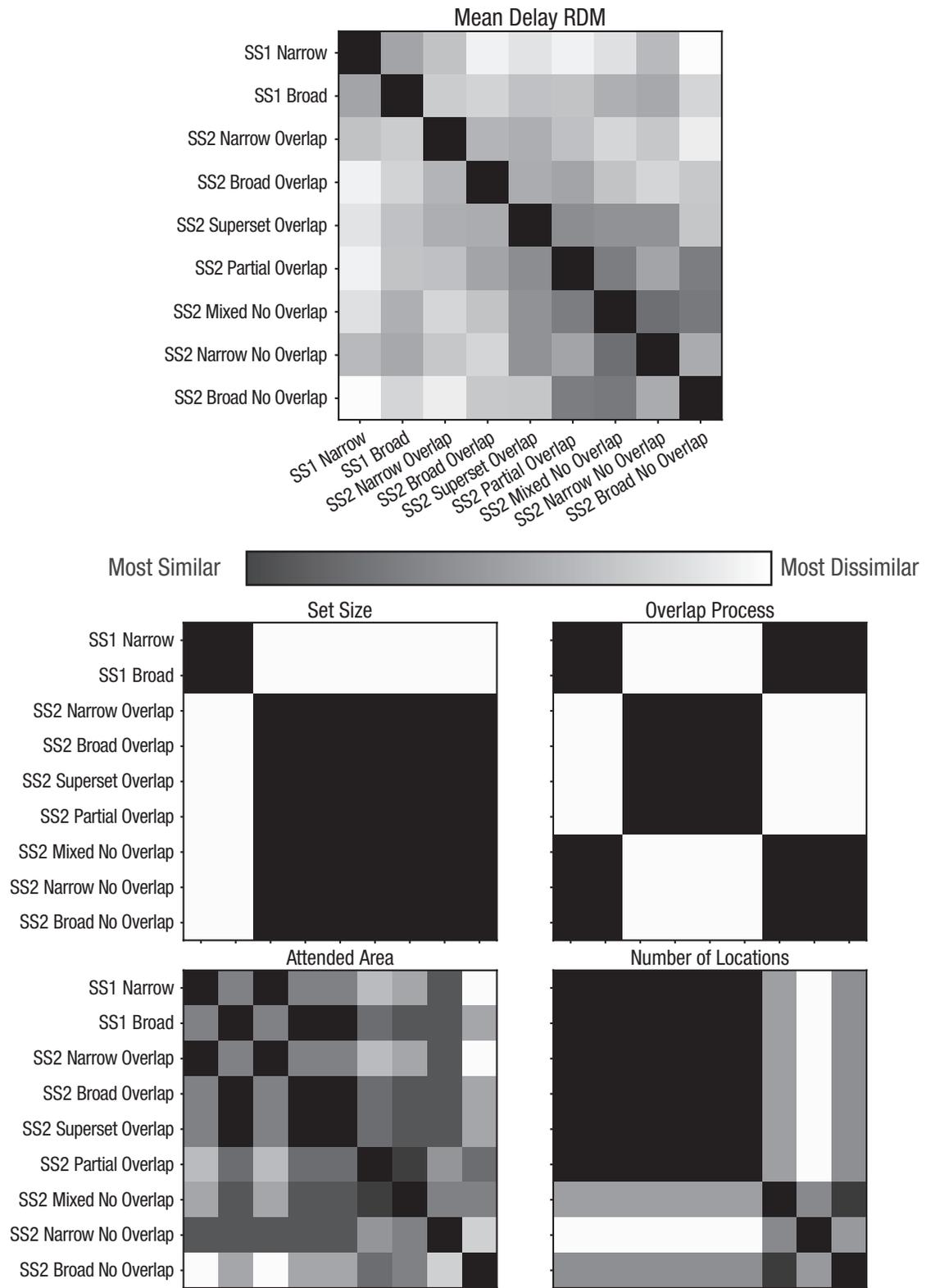
trending toward significantly different—against SS2 partial overlap,  $t(22) = 1.89$ ,  $p = .072$ ,  $BF_{10} = 0.99$ ,  $d = 0.275$ —or was significantly different—against SS2 narrow overlap,  $t(22) = -2.52$ ,  $p = .02$ ,  $BF_{10} = 2.84$ ,  $d = 0.542$ , and against SS2 broad overlap,  $t(22) = -2.3$ ,  $p = .03$ ,  $BF_{10} = 1.93$ ,  $d = 0.351$ . Thus, it does not appear that the hyperplane results can be explained by variations in difficulty or effort.

Ruling out imprecision due to low trial count and the combination of effort and difficulty, two possibilities remain to explain why the set size 2 conditions without overlap were less confidently classified as set size 2. The first is that the conditions lie in a space that is largely orthogonal to the hyperplane drawn by the classifier. The second is that the conditions do lie along the long axis separating the two training conditions and are falling somewhere in the middle. Both cases would imply that even if there is an axis that maximally separates set size, our classifier has not clearly identified it.

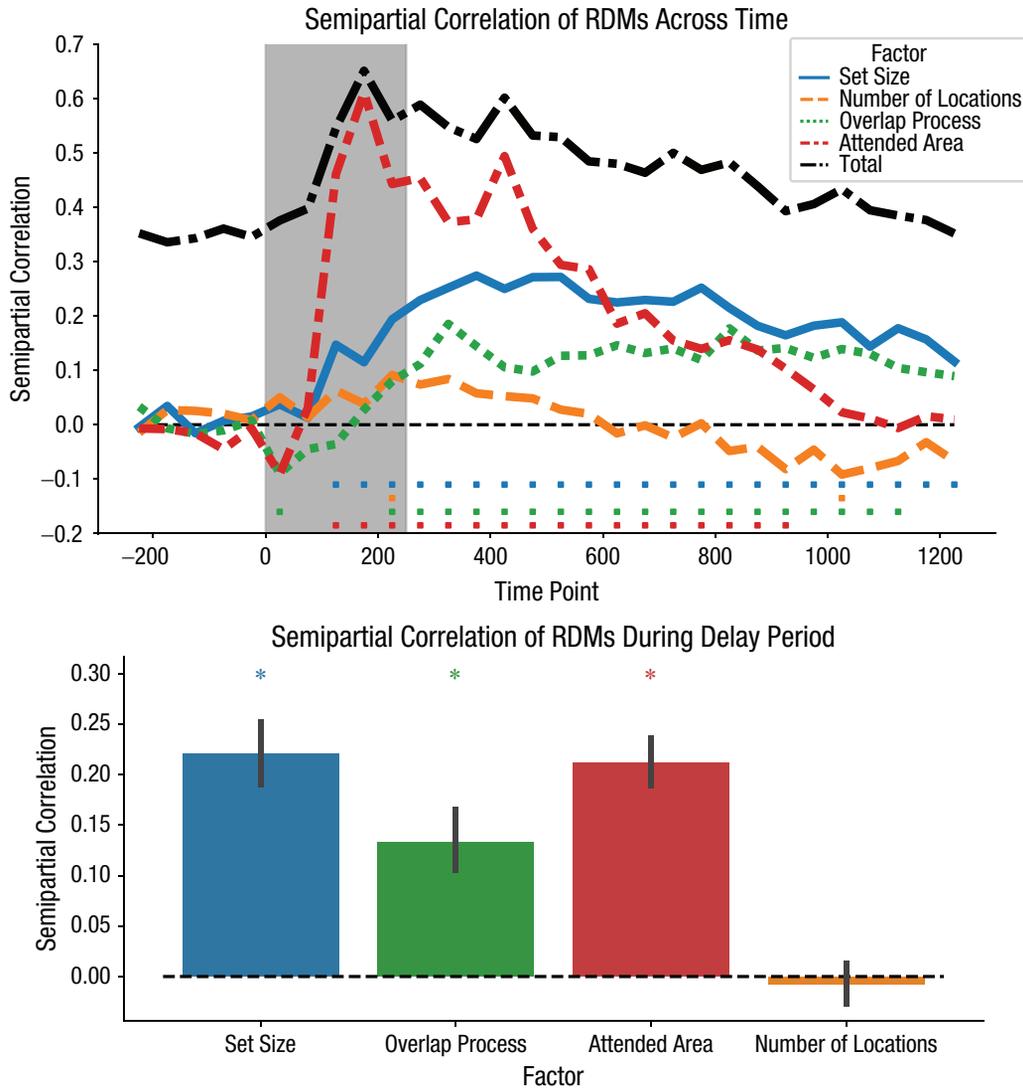
We hypothesized that an additional signal may be present in the training set the classifier is relying on. Specifically, there may be an additional process that is engaged when the two clouds overlap in space. This signal would be present in the SS2 superset training condition and could contribute to separation from the SS1 broad overlap training condition. The pattern of results could then be explained by both set size and this additional overlap-related process. Conditions containing both two objects and an overlap-related process, like SS2 narrow overlap, SS2 broad overlap, and SS2 partial overlap, would be most like the training SS2 condition, and the condition containing only one object and no overlap-related process, like SS1 narrow overlap, would be most like the training SS1 condition. Conditions containing two objects but no overlap-related process would fall in the middle. Although our mvLoad analysis could not directly assess these possibilities, the following section reports how we used RSA to identify the independent contributions of these signals to our decoding results.

**RSA.** Using RSA, we investigated whether both a set-size signal and an overlap-related process signal could be contributing to these results. In addition to set size and overlap, we also considered the impact of the total attended area and the expected number of noncontiguous attended locations as potential confounds to be controlled for. The aim of this approach was to compare empirical RDMs, which capture the similarity structure across conditions, against theoretical RDMs reflecting how these factors of interest would affect the similarity structure (Fig. 11). Thus, this RSA analysis revealed the degree to which each theoretical process was a reliable influence on the overall similarity structure.

In each nonoverlapping, 50-ms window, we computed the linear discriminant contrast (Walther et al.,



**Fig. 11.** Theoretical and empirical representational dissimilarity matrices (RDMs). The top RDM shows the average ranked distance between each pair of conditions during the delay period, averaged across participants. The bottom two rows reflect the theoretical RDMs.



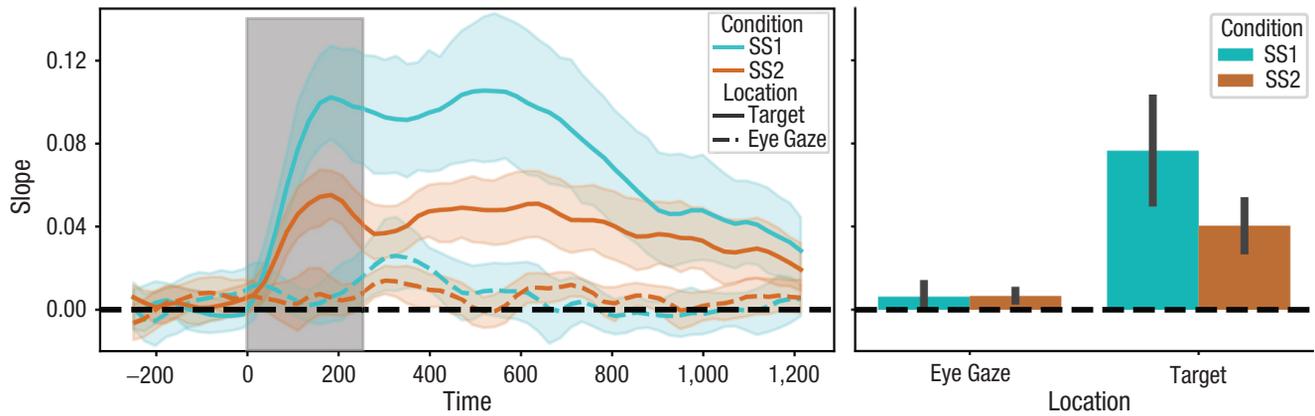
**Fig. 12.** Semipartial rank correlations between theoretical representational dissimilarity matrices (RDMs) and the empirical RDM across the delay period. At top, the semipartial rank correlations across time are shown. The light gray shaded region indicates the duration of stimulus array. Color-matched squares indicate correlations significantly greater than 0, Benjamini-Hochberg corrected for comparisons at each time point. At bottom, the average correlation during the delay period is shown. Error bars reflect the bootstrapped 95% confidence interval. Asterisks (\*) indicate statistical significance ( $p < .05$ ).

2016) for each pair of conditions for each participant using the average activity of the window. This produced an empirical RDM for each participant in each time window. We then compared this empirical RDM against a set of theoretical RDMs by computing the semipartial correlation between the empirical RDM and each factor.

To identify each factor's contribution to the structure of distances between conditions, we computed the semipartial rank correlation for each RDM (Kiat et al., 2022). Using a rank correlation avoids assuming a linear relationship between changes in factor values and changes in the empirical distances. Using a semipartial correlation assures that we are only capturing each

factor's independent contribution to the empirical RDM. Thus, the presence of a significant semipartial correlation can be taken as strong evidence for the presence of the factor. We repeated this procedure for each participant at each time point. We tested these correlations at the group level using a Wilcoxon signed-rank test against zero.

The results are presented in Figure 12. We found that, averaged across the delay period, an overlap-related process independently contributed to differences between conditions (mean semipartial  $r = .133$ ,  $p = .0003$ ,  $BF_{10} = 205.92$ ), as did differences in the total attended area (mean semipartial  $r = .212$ ,  $p = 4.8e-7$ ,  $BF_{10} = 5,173.48$ ). There was no significant contribution



**Fig. 13.** Channel response function (CRF) slopes across time and averaged across the delay period. At left, dashed lines reflect inverted encoding models (IEMs) based on eye gaze, whereas solid lines reflect IEMs based on target conditions, with a matched number of trials. The light gray shaded region indicates the duration of stimulus array. The colored shaded intervals and error bars reflect bootstrapped 95% confidence intervals.

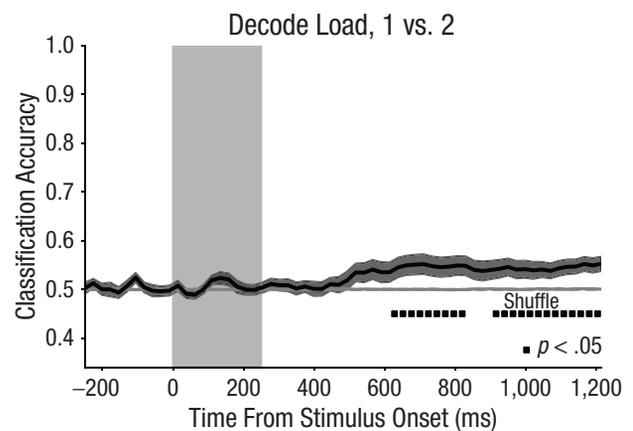
from the number of noncontiguous attended locations, with moderate evidence for the null (mean semipartial  $r = -.007$ ,  $p = .64$ ,  $BF_{10} = 0.246$ ). Critically, set size also produced an independent contribution to the difference between conditions across the delay period (mean semipartial  $r = .22$ ,  $p = 4.5e-6$ ,  $BF_{10} = 1,445.67$ ). This pattern of results was qualitatively identical when subsetting to just the first or second half of the delay period, with one exception. In the second half of the delay period, there was a significant inverse contribution of the number of locations (mean semipartial  $r = -.05$ ,  $p = .048$ ,  $BF_{10} = 2.48$ ). However, this effect was no longer significant when excluding participants with informative eye movements (mean semipartial  $r = -.038$ ,  $p = .145$ ,  $BF_{10} = 0.92$ ), unlike all other effects, which remained significant (and which were significant during both halves of the delay period). Thus, RSA confirmed the presence of a set-size factor that tracked the number of items stored in memory, regardless of variations in the spatial extent of the items or the overlap between those items.

**Controls for eye movements.** We examined whether systematic eye movements could be the driver of these EEG results using two sets of analyses.

First, we asked whether our IEM analysis was influenced by eye position rather than by the locus of covert spatial attention. To this end, we recoded the location for each trial using the average gaze position of the trial. We then compared set size 1 trials with set size 2 trials using these recoded locations. We also reran the original IEM comparison using total alpha power, matching the number of trials per location with that of the eye-gaze locations. The results are plotted in Figure 13. We found extremely small slopes for eye locations, with no distinction based on set size, and moderate evidence for the null,  $t(22) = 0.025$ ,  $p = .98$ ,  $BF_{10} = 0.219$ ,

$d = 0.005$ . In contrast, using a matched number of trials per location, we continued to find clear differences in slope,  $t(22) = 4.90$ ,  $p = 6.8e-5$ ,  $BF_{10} = 387.7$ ,  $d = 1.02$ . Given that this set size contrast produced the largest difference in slopes between conditions, it is unlikely that eye movements are contributing to the differences in conditions that we find here.

Next, we attempted to decode set size (regardless of cloud size and overlap) using just information from eye movements, combining all eye-tracking and EOG data available for each participant. Doing so revealed slight but significant decoding (mean decoding accuracy = 53.3%,  $SD = 1.6\%$ ) beginning around 600 ms after stimulus onset (Fig. 14). This contrasts with the much higher decoding accuracy that is possible when using EEG (mean decoding accuracy = 67.8%,  $SD = 5.5\%$ ), and



**Fig. 14.** Decoding of set size on the basis of eye-movement data available for each participant (electrooculogram and gaze data). The light gray shaded region indicates the duration of stimulus array. The dark gray shaded interval about the average time course reflects the bootstrapped 95% confidence interval. The gray line varying about 0.5 indicates accuracy for shuffled labels.

which begins around 100 ms after stimulus onset. We next identified 4 participants with informative eye movements by setting a threshold of 60% decoding accuracy across the delay period. We replicated all analyses excluding those 4 participants. All significant analyses remained significant or trending toward significance, and in no cases did the pattern of results meaningfully change (but see footnote 3).

Overall, these results suggest that eye movements made little if any contribution to the signals of interest here.

## ***Discussion of Experiment 2***

In Experiment 2, we examined whether neural activity that precisely tracks the locus and breadth of covert spatial attention was tied to neural activity that indexes the number of individuated objects stored in visual working memory. Using dot cloud stimuli to control for variations in spatial extent, we found that the precision of spatially selective alpha activity was unaffected by the number of objects stored in working memory. At the same time, when the spatial extent across different loads was perfectly controlled, a multivariate model still robustly decoded the number of items encoded into visual working memory. Moreover, representational similarity analyses confirmed the conclusion that there was a pure load signal—sustained throughout the delay period—that indexed the number of items in working memory, even while simultaneously documenting the unique variance that was explained by other factors.

## **General Discussion**

Past work has shown a close intertwining between spatial attention and working memory (e.g., Awh et al., 2006; Awh & Jonides, 2001; Chun, 2011). Observers direct spatial attention toward the position of items held in working memory, even when position is irrelevant to the task (Foster et al., 2017), and working memory performance declines when covert orienting toward those positions is hindered (Awh et al., 1998; Williams et al., 2013). Nevertheless, here we found strong evidence that these signals tap into dissociable aspects of voluntary attentional control. In Experiment 1, spatial attention, as measured by posterior alpha power, was sensitive to the number of relevant locations regardless of the number of items occupying those positions, whereas our load signal was sensitive to the number of relevant objects, regardless of the number of locations. In Experiment 2, the precision of spatial attention, as measured by inverted encoding models applied to the topography of alpha power, was sensitive to the size of the attended region but not the number of relevant objects. Simultaneously, we

observed a load signal that was sensitive to the number of relevant objects, regardless of the breadth of the attended region. Thus, although both processes are subject to voluntary control, distinct experimental factors had a selective influence on each aspect of attention. This strongly implies a functional dissociation between covert spatial attention and working memory storage. These results are aligned with past work (Bae & Luck, 2018; Diaz et al., 2021; Fukuda et al., 2015; Günseli et al., 2019; Hakim et al., 2019, 2021; Thyer et al., 2022). Our work shares similar logic to that of Bae and Luck (2018), which separately manipulated and decoded the location and orientation of remembered teardrop stimuli using low-frequency event-related potential signals ( $\leq 6$  Hz) but could only decode location, and not orientation, with alpha power. However, rather than dissociating spatial position from specific feature information (teardrop orientation) maintained in working memory, we are dissociating spatial position from the number of items held in working memory. This rules out spatial attention as a compelling alternative explanation for past observations of working memory load signals that generalized across variations in the type of visual features maintained (Thyer et al., 2022).

Importantly, we are not claiming that alpha power only reflects spatial attentional signals and that raw voltage only reflects signals of working memory load. Although it is true that precise spatial information is predominantly reflected by variations in alpha power (as seen in Experiment 1b; Bae & Luck, 2018; Foster et al., 2017), it is likely that activity in the alpha band indexes other cognitive processes as well. For example, unpublished work in our lab found that robust mvLoad models can be constructed on the basis of the topography of alpha activity. Moreover, RSA in the present work revealed the presence of spatial information in raw voltage. Thus, rather than arguing for a distinction between alpha power and raw voltage, we are focused on the distinction between spatial attention and working memory gating. In that context, alpha power and raw voltage provided the most robust analytic approaches for tracking spatial attention and working memory encoding, respectively.

Although the methods we employed were chosen because they provide a precise measurement of spatial attention and working memory load, they are still limited by the information in the signals to which they are applied. Thus, it is possible that some of our conclusions could change if we were to use alternative methods. For instance, the amplitude of the P1 event-related potential indexes whether spatial attention is allocated toward a position (Mangun & Hillyard, 1991). Future work could examine whether dissociations between the breadth of covert spatial attention and the number of

items encoded into working memory will generalize with alternative measures of these processes.

Although these findings help to refine our taxonomy of voluntary attentional control, they do not affect the importance of interactions between spatial attention and working memory storage. As noted above, spatial attention is consistently oriented toward the position of items in working memory, suggesting that spatial selection could be a prerequisite for working memory encoding, or a component of rehearsal during the delay period (Awh & Jonides, 2001; Williams et al., 2013). Thus, these attentional processes may interact with one another in a mutually reinforcing manner to prioritize the processing of relevant information across stages of processing.

Basic questions remain regarding the specific role (or roles) of spatial attention in visual working memory. Here we consider three possible answers, which are not mutually exclusive. First, space may act as a key organizing dimension in the building and maintenance of object representations, (i.e., for feature binding; Abrahamse et al., 2014; Pertzov & Husain, 2014; Schneegans & Bays, 2017; Treisman & Zhang, 2006). If so, it may be that the allocation of spatial attention is a prerequisite to working memory encoding. Second, the allocation of spatial attention may improve the signal-to-noise ratio of early sensory processing and representations (Martínez et al., 2006), thereby shaping the fidelity of representations available for working memory storage. Finally, spatial attention may be a specific instance of feature-based attention. Previous work suggests that humans can prioritize the maintenance of the relevant feature during the delay period of a working memory task (Serences et al., 2009; Woodman & Vogel, 2008). The task employed in Experiment 2 required participants to remember the spatial location and envelope of each cloud; the reconstructed CRFs could reflect attending to spatial information as a feature of the clouds. As stated above, these possibilities are not mutually exclusive. Attentional modulation of early sensory processing may be a form of feature-based attention (e.g., Müller et al., 2006). Thus, future work can examine whether similar dissociations might be observed between nonspatial forms of feature-based attention (e.g., attention to color or motion) and the gating of working memory representations.

Finally, our proposed dissociation raises interesting questions for the broader literature on attentional control. For example, much past work has examined the neural mechanisms of spatial attention, often focusing on parietal regions (e.g., Bisley & Goldberg, 2003; Fiebelkorn et al., 2019; Karnath & Rorden, 2012). However, this literature often fails to differentiate between the selection of spatial locations or the encoding into working memory of the objects in those locations. Likewise, an emerging body of work has focused on how

experience with distractors of a given color or location can yield increased resistance to interference from those irrelevant stimuli (e.g., Gaspelin & Luck, 2018; Wang et al., 2019). But reduced processing of distractors could be explained either by reduced capture of spatial attention or by a reduced probability of encoding those distractors into working memory. Thus, the dissociation we are suggesting between spatial attention and working memory gating may enable a refined understanding of which aspects of voluntary control are critical in different contexts.

## Transparency

*Action Editor:* Sachiko Kinoshita

*Editor:* Patricia J. Bauer

*Author Contributions*

**Henry M. Jones:** Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing – original draft; Writing – review & editing.

**Gisella K. Diaz:** Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Visualization; Writing – original draft; Writing – review & editing.

**William X. Q. Ngiam:** Conceptualization; Supervision; Writing – review & editing.

**Edward Awh:** Conceptualization; Funding acquisition; Resources; Supervision; Writing – original draft; Writing – review & editing.

*Declaration of Conflicting Interests*

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

*Funding*

This research was supported by National Institute of Mental Health Grant No. ROIMH087214 and Office of Naval Research Grant No. N00014-12-1-0972 to E.A. and a Neubauer Distinguished Scholar Doctoral Fellowship from The University of Chicago to H.M.J.

*Open Practices*

This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



## ORCID iDs

Henry M. Jones  <https://orcid.org/0000-0001-7719-3646>

Gisella K. Diaz  <https://orcid.org/0000-0002-3677-824X>

William X. Q. Ngiam  <https://orcid.org/0000-0003-3567-3881>

Edward Awh  <https://orcid.org/0000-0002-5211-5278>

## Acknowledgments

We thank John Veillette for help optimizing the generation of representational dissimilarity matrices. William X. Q. Ngiam is now affiliated with the School of Psychology, University of Adelaide.

## Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/09567976241263002>

## References

- Abrahamse, E., van Dijk, J.-P., Majerus, S., & Fias, W. (2014). Finding the answer in space: The mental whiteboard hypothesis on serial order in working memory. *Frontiers in Human Neuroscience*, *8*, Article 932. <https://doi.org/10.3389/fnhum.2014.00932>
- Adam, K. C. S., Vogel, E. K., & Awh, E. (2020). Multivariate analysis reveals a generalizable human electrophysiological signature of working memory load. *Psychophysiology*, *57*(12), Article e13691.
- Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126.
- Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 780.
- Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*(1), 201–208.
- Bae, G. Y., & Luck, S. J. (2018). Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials. *Journal of Neuroscience*, *38*(2), 409–422.
- Bisley, J. W., & Goldberg, M. E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science*, *299*(5603), 81–86.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia*, *49*(6), 1407–1409.
- Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 109–127.
- Cowan, N. (1999). An Embedded-Processes Model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge University Press. <https://doi.org/10.1017/CBO9781139174909.006>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.
- Diaz, G. K., Vogel, E. K., & Awh, E. (2021). Perceptual grouping reveals distinct roles for sustained slow wave activity and alpha oscillations in working memory. *Journal of Cognitive Neuroscience*, *33*(7), 1354–1364.
- Feldmann-Wüstefeld, T., & Awh, E. (2020). Alpha-band activity tracks the zoom lens of attention. *Journal of Cognitive Neuroscience*, *32*(2), 272–282.
- Fiebelkorn, I. C., Pinsk, M. A., & Kastner, S. (2019). The medio-dorsal pulvinar coordinates the macaque fronto-parietal network during rhythmic spatial attention. *Nature Communications*, *10*(1), Article 215.
- 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.
- 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.
- Gaspelin, N., & Luck, S. J. (2018). The role of inhibition in avoiding distraction by salient stimuli. *Trends in Cognitive Sciences*, *22*(1), 79–92.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., . . . Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, Article 267. <https://doi.org/10.3389/fnins.2013.00267>
- Günsele, E., Fahrenfort, J. J., van Moorselaar, D., Daoulziz, K. C., Meeter, M., & Olivers, C. N. (2019). EEG dynamics reveal a dissociation between storage and selective attention within working memory. *Scientific Reports*, *9*(1), Article 13499. <https://doi.org/10.1038/s41598-019-49577-0>
- Günsele, E., Foster, J., Sutterer, D., Todorova, L., Vogel, E. K., & Awh, E. (2022). Overlapping neural representations for dynamic visual imagery and stationary storage in spatial working memory. *BioRxiv*, 2022-09. <https://doi.org/10.1101/2022.09.24.509255>
- Hakim, N., Adam, K. C. S., Günsele, 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.
- Hakim, N., Feldmann-Wüstefeld, T., Awh, E., & Vogel, E. K. (2021). Controlling the flow of distracting information in working memory. *Cerebral Cortex*, *31*(7), 3323–3337.
- Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science*, *17*(7), 572–576.
- Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *353*(1373), 1257–1270.
- Iman, R. L., & Conover, W. J. (1979). The use of the rank transform in regression. *Technometrics*, *21*(4), 499–509.
- Karnath, H. O., & Rorden, C. (2012). The anatomy of spatial neglect. *Neuropsychologia*, *50*(6), 1010–1017.
- Kiat, J. E., Hayes, T. R., Henderson, J. M., & Luck, S. J. (2022). Rapid extraction of the spatial distribution of physical saliency and semantic informativeness from natural scenes in the human brain. *Journal of Neuroscience*, *42*(1), 97–108.
- Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, Article 4. doi: 10.3389/neuro.06.004.2008
- Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411.

- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, Article 213. <https://doi.org/10.3389/fnhum.2014.00213>
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E. J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*, 1–17.
- Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*(6601), 616–618.
- Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience & Biobehavioral Reviews*, *62*, 100–108.
- Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(4), 1057–1074.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.
- Martínez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., Wong, E. C., Hinrichs, H., Heinze, H. J., & Hillyard, S. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, *2*(4), 364–369.
- Martínez, A., Teder-Sälejärvi, W., Vazquez, M., Molholm, S., Foxe, J. J., Javitt, D. C., Di Russo, F., Worden, M. S., & Hillyard, S. A. (2006). Objects are highlighted by spatial attention. *Journal of Cognitive Neuroscience*, *18*(2), 298–310.
- Müller, M. M., Andersen, S., Trujillo, N. J., Valdes-Sosa, P., Malinowski, P., & Hillyard, S. A. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. *Proceedings of the National Academy of Sciences, USA*, *103*(38), 14250–14254.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1–9.
- Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, *592*(7855), 601–605.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Pertsov, Y., & Husain, M. (2014). The privileged role of location in visual working memory. *Attention, Perception, & Psychophysics*, *76*, 1914–1924.
- Schneegans, S., & Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *Journal of Neuroscience*, *37*, 3913–3925.
- 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.
- Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring latent visual working memory representations in human cortex. *Neuron*, *91*(3), 694–707.
- Sutterer, D. W., Foster, J. J., Adam, K. C. S., Vogel, E. K., & Awh, E. (2019). Item-specific delay activity demonstrates concurrent storage of multiple active neural representations in working memory. *PLOS Biology*, *17*(4), Article e3000239. <https://doi.org/10.1371/journal.pbio.3000239>
- Thaler, L., Schütz, A. C., Goodale, M. A., & Gegenfurtner, K. R. (2013). What is the best fixation target? The effect of target shape on stability of fixational eye movements. *Vision Research*, *76*, 31–42.
- Thyer, W., Adam, K. C. S., Diaz, G. K., Velázquez Sánchez, I. N., Vogel, E. K., & Awh, E. (2022). Storage in visual working memory recruits a content-independent pointer system. *Psychological Science*, *33*(10), 1680–1694.
- Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition*, *34*, 1704–1719.
- van Doorn, J., Ly, A., Marsman, M., & Wagenmakers, E. J. (2020). Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and Spearman's  $\rho$ . *Journal of Applied Statistics*, *47*(16), 2984–3006.
- Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1656. <https://doi.org/10.1037//0096-1523.24.6.1656>
- Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, *9*(4), 739–743.
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751.
- Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*, 188–200.
- Wang, B., van Driel, J., Ort, E., & Theeuwes, J. (2019). Anticipatory distractor suppression elicited by statistical regularities in visual search. *Journal of Cognitive Neuroscience*, *31*(10), 1535–1548.
- Williams, M., Pouget, P., Boucher, L., & Woodman, G. F. (2013). Visual-spatial attention aids the maintenance of object representations in visual working memory. *Memory & Cognition*, *41*, 698–715.
- Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review*, *15*(1), 223–229.
- Woodman, G. F., Wang, S., Sutterer, D. W., Reinhart, R. M., & Fukuda, K. (2022). Alpha suppression indexes a spotlight of visual-spatial attention that can shine on both perceptual and memory representations. *Psychonomic Bulletin & Review*, *29*(3), 681–698.