Representing Context and Priority in Working Memory

Quan Wan[1], Adel Ardalan[2], Jacqueline M. Fulvio[1], and Bradley R. Postle[1]

Abstract

■ The ability to prioritize among contents in working memory (WM) is critical for successful control of thought and behavior. Recent work has demonstrated that prioritization in WM can be implemented by representing different states of priority in different representational formats. Here, we explored the mechanisms underlying WM prioritization by simulating the double serial retrocuing task with recurrent neural networks. Visualization of stimulus representational dynamics using principal component analysis revealed that the network represented trial context (order of presentation) and priority via different mechanisms. Ordinal context, a stable property lasting the duration of the trial, was accomplished by segregating representations into orthogonal subspaces. Priority, which changed multiple times during a trial, was accomplished by separating representations into different strata within each subspace. We assessed the generality of these mechanisms by applying dimensionality reduction and multiclass decoding to fMRI and EEG data sets and found that priority and context are represented differently along the dorsal visual stream and that behavioral performance is sensitive to trial-by-trial variability of priority coding, but not context coding. ■

INTRODUCTION

One of the hallmarks of working memory (WM) is its ability to flexibly prioritize among its contents in the service of the current behavioral goal. For example, say that you have just completed a talk at a conference and you see two people simultaneously approaching each of two microphones to ask a question. You turn to the moderator and wait for them to indicate who will ask the first question, and based on this, your shift of gaze is guided by your memory of the location of the cued microphone. To study prioritization in WM, one line of work has made extensive use of the double serial retrocuing (DSR) task, in which two sample items are initially presented and "remembered," followed by a blank "no-action" delay, and then a retrocue indicating which of the two memorized items will be tested by an impending memory probe (see Figure 1A for an example). This item is said to take on the status of prioritized memory item (PMI). Because the item that was not cued may be tested later in the trial, however, it cannot be dropped from memory (i.e., "forgotten"), so it takes on the status of unprioritized memory item (UMI) until the PMI is tested. Subsequently, the priority status of both items resets to neutral until a second retrocue indicates, unpredictably, which will be tested by a second memory probe; thus, either item can take on the status of PMI during the second half of the trial. An initial set of studies applying multivariate pattern analysis (MVPA) decoding to fMRI and EEG data from participants performing the DSR task failed to find evidence for an active representation of the UMI, giving rise to the idea that it might be held in an "activity-silent" state (LaRocque, Riggall, Emrich, & Postle, 2017; Rose et al., 2016; Larocque, Lewis-Peacock, & Postle, 2014; Lewis-Peacock, Drysdale, Oberauer, & Postle, 2012). More recently, however, studies using variants of the DSR task (with fMRI; Yu, Teng, & Postle, 2020; van Loon, Olmos-Solis, Fahrenfort, & Olivers, 2018) and the 2-back WM task (with EEG; Wan, Cai, Samaha, & Postle, 2020) have provided evidence for an active trace of the UMI that undergoes a transformation relative to the representational format of the PMI. Specifically, the UMI can produce significantly below-baseline MVPA decoding (van Loon et al., 2018) and "opposite" reconstruction with multivariate inverted encoding modeling (IEM; Wan et al., 2020; Yu et al., 2020).

As an initial step toward better understanding the priority-based representational transformations observed in neuroimaging data (Wan et al., 2020; Yu et al., 2020; van Loon et al., 2018), we had trained recurrent neural networks (RNNs) with a long short-term memory architecture to perform the 2-back WM task (Wan, Menendez, & Postle, 2022). Visualization of long short-term memory hidden layer activity using principal component analysis (PCA) had confirmed that stimulus representations in RNNs also undergo representational transformations when transitioning between priority states. Specifically, demixed (d) PCA of these data had identified two representational trajectories, one within a UMI-specific subspace and the other a PMI-specific subspace, both undergoing a reversal of stimulus coding axes. Having thus observed similar priority-based transformational dynamics in the human brain and in RNNs, we speculated that this type of transformation might be a computationally rational way to meet the competing demands of retaining information in WM

[1]University of Wisconsin–Madison, [2]Princeton Neuroscience Institute

© 2024 Massachusetts Institute of Technology

Journal of Cognitive Neuroscience 36:7, pp. 1374–1394
https://doi.org/10.1162/jocn_a_02166

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

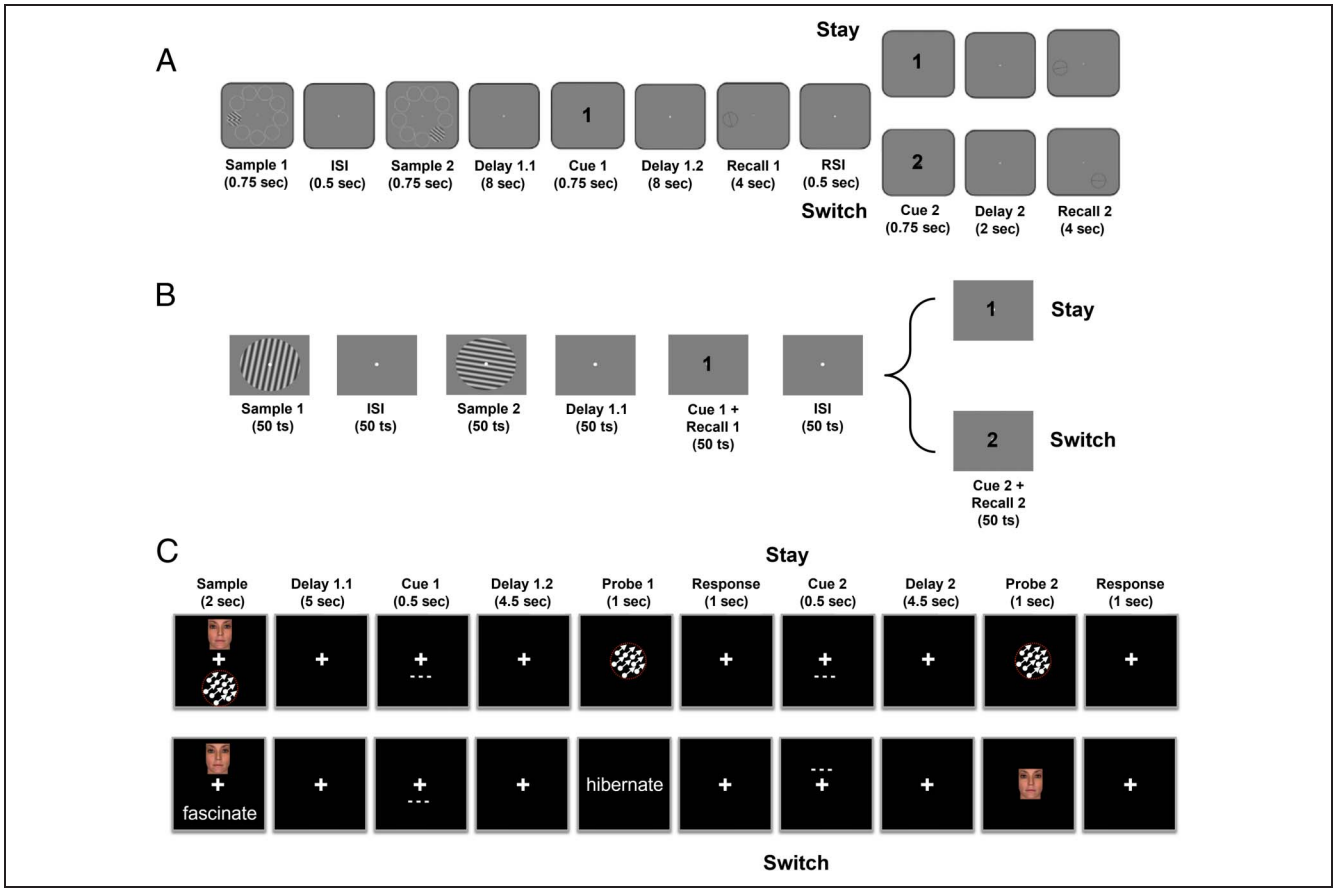

Figure 1. Experimental procedure for (A) the fMRI task, (B) the RNN task, and (C) the EEG task. Figures adapted, with permission, from Yu, Teng, and Postle (2020; A), and Fulvio and Postle (2020; C).

while simultaneously preventing it from interfering with concurrent behavior (Wan et al., 2022).

Whereas in Wan and colleagues (2022) we simulated the 2-back task, the work presented here begins with RNN simulation of the DSR task. This was important to do because although the *n*-back task has been important for the study of many aspects of WM, it is poorly suited for the study of the flexible control of behavior with WM. This is because the *n*-back is a continuous performance task in which each item follows the same functional trajectory. For the 2-back, for example, each item *n* first serves as a memory probe against which to compare one's memory for item $n - 2$, then transitions to UMI (whereas $n + 1$ is compared with the memory of $n - 1$), then transitions to PMI (for its comparison with item $n + 2$), and then becomes no longer relevant and can be dropped from WM. The DSR, in contrast, does require online, flexible control, because the identity of the two retrocues cannot be predicted before their onset.

At the beginning of each trial of DSR, sample items can either be presented simultaneously or serially. When items are presented simultaneously, they necessarily each appear at a different location, and it is an item's unique location that is used by the retrocue to designate it the PMI. Thus, the location at which an item appears serves as a critical, trial-specific context. When items are presented serially, they can appear at the same or different locations, but if the retrocue designates the prioritized item by referring to the order in which it was presented (i.e., "first" or "second"), then it is the item's ordinal context that is critical for successful performance. Note here the fundamental distinction between an item's *identity* and the *context* in which it was presented, and the necessary role played by both. In the DSR task of Yu and colleagues (2020), for example, stimuli were drawn from a pool of nine oriented gratings. On any given trial *n*, it would not be sufficient to remember that a stimulus with the identity of, for example, 0° was one of the two presented, because the 0° stimulus may have already appeared on several previous trials. To successfully interpret the cue on trial *n*, it is necessary to also remember that this stimulus had been presented first or second on this trial. This latter property is the trial-specific context in which the item appeared, and it is the binding of an item's identity to its trial-specific context that is fundamental to that stimulus being in the state of being "in WM" on that trial (Oberauer & Lin, 2017).

The initial motivation for the RNN simulation of the DSR task that we report here was to better understand the priority-based transformations summarized above (Wan et al., 2020; Yu et al., 2020; van Loon et al., 2018). However, as we report here, these simulations also yielded the

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

unexpected finding that the representation of the first sample item underwent a dramatic transformation upon the onset of the second item. That is, before the designation of priority (which would be indicated by the retrocue), the first item underwent a context-based transformation. Specifically, whereas it had been represented in a subspace defined by the first two principal components (PCs) of a PCA applied to the hidden layer of the RNN, the representation of the first item was displaced from this subspace upon the presentation of the second item, and shunted to a new subspace defined by the third and fourth PCs of the PCA. This finding caused us to reconsider our interpretation of the transformational dynamics observed in the 2-back task (Wan et al., 2022), because an item's functional trajectory during that task confounded priority with context. The aims of this report, therefore, were twofold. One was to explore, at the algorithmic level, how context-based representational transformations may differ from priority-based transformations. This was carried out via RNN simulations. The second was to leverage what was learned from the RNNs to assess how the encoding of these two properties, context and priority, might differ in the way they influence behavior and in the way they are represented in the brain.

METHODS

The data presented here derive from three sources: RNN simulations of a DSR task, reanalysis of data from an EEG study of DSR, and reanalysis of an fMRI study of DSR.

Participants

EEG

The EEG data set is from 12 healthy young adults (5 female participants, average age = 21.7 ± 3.2 years, all right-handed), as described in detail in Fulvio and Postle (2020). This *N* was double that of a previous EEG study for which MVPA decoding results yielded informative prioritization effects (Rose et al., 2016), and so was deemed satisfactory for the analyses to be carried out here.

fMRI

The fMRI data set is from 13 healthy young participants (10 female participants, average age = 21.1 ± 4.5 years, all right-handed), as described in detail by Yu and colleagues (2020). Because IEM analyses of these fMRI data had yielded informative prioritization effects, this *N* was deemed satisfactory for the analyses to be carried out here.

Behavioral Tasks

DSR Procedure

At a generic level of description, a trial in the DSR task entails testing WM for the sample stimuli with two successive tests, with a cue preceding each test that indicates, with 100% validity, which sample item will be tested. The first half of a trial begins with the presentation of two sample items, followed by a delay period ("*Delay 1*") during which a retrocue is presented at the halfway point, followed by a memory test. The retrocue ("*Cue 1*") indicates with 100% validity which of the two sample items will be tested at the end of Delay 1, and this "*Test 1*" can be either a recognition probe or a recall interface (for orientations, a recall dial). Because this is the first delay period in the trial, the portion of the trial that spans between the offset of the sample stimuli and the onset of Test 1 is considered "*Delay 1*." In addition, however, because the retrocue changes the nature of how the information being held in WM is processed, it is useful to distinguish the precue portion of Delay 1 as "*Delay 1.1*" and the postcue portion of Delay 1 as "*Delay 1.2*." Indeed, it is the transformation of an item's status from neutral (during Delay 1.1) to either PMI or UMI, during Delay 1.2, that is of primary interest. Upon completion of Test 1, the two items in WM return to a neutral status during the response–stimulus interval separating Test 1 from the second retrocue ("*Cue 2*"), because it is equiprobable that either of them will be designated by Cue 2, and subsequently, after "*Delay 2*, " tested by Test 2 (Figure 1B).

RNN Models

The training task (Figure 1B) was modeled after the task performed in the fMRI study (Yu et al., 2020; Figure 1A), including the fact that the two sample items were presented serially. (Different from the fMRI study, however, the location of the stimulus presentation was not modeled, and so the RNN simulations were carried out as though both sample stimuli were presented at the same location.) The training task also deviated from the generic structure of the DSR in a few ways that accommodated idiosyncrasies and constraints of RNN modeling, including the facts that a "cue" input unit had to input information to the network at each timestep and that the network had to output information at each timestep. One deviation was that each distinct epoch in the trial was the same length: 50 timesteps. A second deviation was that the cue-input unit input a value of 0 during each timestep when the priority status of the two items was neutral (i.e., during sample presentation and the equivalents of the ISIs and of Delay 1.1), and it input a value of 1 or −1 during each timestep when either the first or the second sample, respectively, had the priority status of PMI. A third deviation was that the RNN training task did not include a postcue delay period (i.e., a Delay 1.2 or a Delay 2); instead, Delay 1.1 was followed by an epoch that combined the roles of Cue 1 and Recall 1, and the second ISI was followed by an epoch that combined the roles of Cue 2 and Recall 2 (i.e., recall responses were made beginning coincident with the onset of each cue).

Stimuli were randomly drawn from a pool of oriented gratings that covered the continuous range from [0°,

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

180°] interval (Sample 1: φ and Sample 2: θ). Stimulus location was not simulated, and it was possible for φ and θ to take on the same orientation. Each trial began with the presentation of Sample 1 (50 timesteps) followed by an ISI (i.e., blank delay; 50 timesteps) followed by Sample 2 (50 timesteps) followed by Delay 1.1 (50 timesteps) followed by Cue 1/Recall 1 (50 timesteps; the response window was the duration of Cue 1). Next came another ISI (50 timesteps) followed by Cue 2/Recall 2 (50 timesteps). Cue 2 matched ("stay") or did not match ("switch") Cue 1, unpredictably, and equal number of times.

fMRI: DSR with Ordinal and Location Context, and Recall Probes

Stimuli were drawn from a pool of nine oriented gratings that evenly covered the range from 0° to 179°, and could be presented at one of nine locations that, each at a distance of 8° of visual angle from central fixation, evenly covered the range of possible locations from 0° to 359° of polar angle. Each trial began with the presentation Sample 1 (.75 sec) followed by an ISI (.5 sec), followed by Sample 2 (.75 sec), followed by Delay 1.1 (8 sec), followed by a centrally presented digit ("1" or "2," Cue 1; .75 sec). After the ensuing Delay 1.2 (8 sec), a recall dial appeared at the location that had been occupied by the PMI, and the participant had 4 sec to rotate it to match their memory of that item's orientation (Recall 1). Subsequently, after a brief unfilled interval (.5 sec), a second centrally presented digit ("1" or "2," Cue 1; .75 sec) indicated the item to be tested, after Delay 2 (2 sec), at Recall 2 (4 sec). The critical independent variable was continuous error of recall. Cue 2 matched ("stay") or did not match ("switch") Cue 1, unpredictably, an equal number of times (Figure 1A; see Appendix Figure A1 for the task timeline mapping task events to time in seconds or TRs).

Because the location of the recall dial indicated the item to be recalled, a possible strategy would be to ignore the cues and simply behave based on the location of the recall dial. However, this strategy was discouraged because of an important detail of the procedure. On each trial, the orientation and the location of each stimulus were selected at random (with replacement) and independently. Thus, on each trial, there was a $p = .11$ chance that the second sample would have the same orientation as the first and, independently, a $p = .11$ chance that the second sample would appear at the same location as had the first. These contingencies encouraged participants to not wait for the onset of the recall dial to recall the orientation of the PMI and, indeed, patterns of priority-related transformation of the UMI during Delay 1.2, as assessed by IEM, confirmed that participants used the ordinal cue to guide their behavior (Yu et al., 2020).

EEG: DSR with Location Context and Recognition Probes

Each trial began with the simultaneous presentation of two sample items, one drawn from each of two out of three possible categories (faces, direction of dot motion, and words), one appearing above and one below central fixation (2 sec; Figure 1C). The samples were replaced by a central fixation symbol ("+") during an initial delay (Delay 1.1; 5 sec), followed by a dashed line appearing at one of the two sample locations (.5 sec), indicating that that item would be the first to be tested (Cue 1). After a second delay (Delay 1.2; 4.5 sec), during which the cued item had the status of PMI and the uncued item the status of UMI, an image serving as a recognition probe appeared centrally and was either identical to the PMI ("match," $p = .5$), drawn from the same category but a different exemplar than the PMI ("nonmatch," $p = .3$), or identical to the UMI (also "nonmatch," $p = .2$; Probe 1; 1 sec). Probe 1 was replaced by the fixation symbol (Recall 1; 1 sec), and a response was required during the 2 sec spanning Probe 1 and Recall 1. Next, a dashed line appeared at one of the two sample locations (Cue 2; .5 sec), thereby designating the PMI for the following Delay 2 (4.5 sec), then Probe 2 (1 sec), and then Recall 2 (1 sec). Intertrial interval varied from 2 to 4 sec.

Data were collected during three sessions, each on a separate day, with each session comprising eight 30-trial blocks, alternating between blocks of DSR and a single retrocue task (results from single retrocue task not presented here). During each block, Cue 1 appeared unpredictably at the "up" or "down" location an equal number of times and, orthogonal to Cue 1 location, Cue 2 appeared, unpredictably, at the same ("Stay"; Figure 1C, top row) or opposite ("Switch"; Figure 1C, bottom row) location as had Cue 1 an equal number of times. Balanced across cue conditions, single pulses of transcranial magnetic stimulation (spTMS) was delivered 2–3 sec after the offset of Cue 1 on 50% of trials and, orthogonally, after the offset of Cue 2 on 50% of trials. Note that the EEG data used for the "transformation variability analyses" (see the Analysis Procedures section below) include epochs with and without spTMS. Although a key finding from Fulvio and Postle (2020) was a selective effect of spTMS on one subtype of nonmatching probe, the analyses carried out for this study would collapse across this factor. Nonetheless, to confirm that spTMS did not affect behavior at a more general level, we fit a linear mixed-effects model to the behavioral accuracy with main effects of (i) probe position (two levels: first and second = specified as an ordinal variable) and (ii) spTMS (two levels: delivered and not delivered = specified as a categorical variable). In addition, the interaction between probe position and spTMS delivery was included in a first model and omitted in a second model. As random effects, the models included an intercept for each participant. p Values were obtained by F tests for each term in the models. The second model without the interaction term resulted in a better fit to the data yielding a lower Akaike information criterion. This model revealed a significant main effect of probe position (Probe 2 responses were overall less accurate than Probe 1 responses; $F(1, 45) = 5.5479$, $p = .023$, and no main effect of spTMS, $F(1, 45) =$

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

2.8385, $p = .099$). Thus spTMS did not influence performance at the level at which it was considered here.

Experimental Procedures

RNN

Stimulus orientations were fed into the network via 32 orientation-tuned input units whose preferred orientations spanned the full 180° range and whose response properties were based on V1 orientation-selective neurons (Teich & Qian, 2003; Figure 2). A 33rd input unit was used for retrocue, inputting "0" on each noncue timestep and a "1" or "−1" (indicating "1st" or "2nd," respectively) during each cue timestep. The two output units were trained to produce $\cos(2x)$ and $\sin(2x)$ where x was either θ or φ depending on the cue) so that the 0° orientation had the same output as the 180° orientation (Figure 2).

Our network had 100 fully connected recurrent units, and the dynamics $u_i(t)$ of each recurrent unit were governed by the following standard continuous-time RNN equations:

$$\tau \frac{dx_i(t)}{dt} = -x_i(t) + \sum_{j=1}^{N^{rec}} W_{ij}^{rec} u_j(t) + \sum_{k=1}^{N^{in}} W_{ik}^{in} I_k(t) + b_i$$

$$u_i(t) = f(x_i(t)) + \xi_i(t)$$

for $i = 1, \ldots, N^{rec}$. We introduced nonlinearity using the rectified linear unit function $f(x) = \max(0,x)$. Each recurrent unit received input from other units via recurrent connections with weights specified by the matrix W^{rec}, initialized orthogonally (Saxe, McClelland, & Ganguli, 2013). In addition, these units received external input $I(t)$ to the RNN via weights specified by the matrix W^{in}. Each unit carried two sources of bias: (1) b_i, learned during training, and (2) $\xi_i(t)$, which represented intrinsic noise in the network and took the form of white Gaussian (sampled independently at each timestep) with zero mean. Adding noise to the RNN during training has been shown to stabilize the results and help find a more canonical solution (Cueva, Ardalan, Tsodyks, & Qian, 2021). (Note that noise was not added at testing.) We simulated the approximate network dynamics using the Euler method for $T = 350$ timesteps, each having a duration $\tau/10$ (Mante, Sussillo, Shenoy, & Newsome, 2013). We chose $dt/\tau = 0.1$ similar to (Cueva et al., 2021), for example, $dt = 10$ msec and $\tau = 100$ msec, which would make the time scale of our simulations close to that of the fMRI experiment. The outputs $y_j(t)$ were then generated by combining the activities of the recurrent units based on:

$$y_j(t) = g\left(\sum_{i=1}^{N^{rec}} W_{ji}^{out} u_i(t)\right)$$

where g is the tanh activation function.

We optimized the network parameters W^{in}, W^{rec}, b, and W^{out} to minimize the mean squared error between the target outputs and the network outputs:

$$E = \frac{1}{MTN^{out}} \sum_{m,t,j=1}^{M,T,N^{out}} \left(y_j(t,m) - y_j^{target}(t,m)\right)^2$$

Parameters were updated with the Adam stochastic gradient descent algorithm (Kingma & Ba, 2014) via backpropagation through time (Rumelhart, Hinton, & Williams, 1986), and each network was trained for 10,000 epochs. (See Cueva et al., 2021, for more methodological detail.)

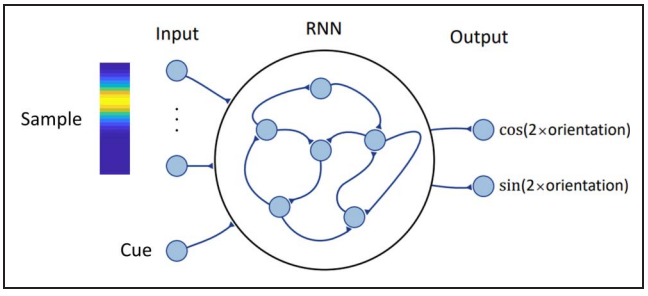

Figure 2. RNN input and architecture. Top left illustrates input of a stimulus (either Sample 1 or Sample 2) with an angular value corresponding to the peak magnitude of this 32-dimensional vector; bottom left illustrates that at each timestep the value of the input to the cue input unit was 0, 1, or −1.

fMRI

For each participant, ROIs were defined, both anatomically and functionally, for eight regions: early visual cortex (EVC, V1 and V2 merged), IPS0-through-IPS5 (six ROIs), and FEF (all ROIs cover both hemispheres). First, anatomical ROIs were defined by extracting masks from the probabilistic atlas of Wang, Mruczek, Arcaro, and Kastner (2015) and warping them to each participant's structural scan in native space. To identify a task-related activity, we modeled each epoch of the task with six boxcar regressors in a general linear model—sample (2 sec), Delay 1.1 (8 sec), Delay 1.2 (8 sec), Recall 1 (4 sec), Delay 2 (2 sec), and Recall 2 (4 sec) convolved with a canonical hemodynamic response function, and we also included covariates to control for motion. We proceeded to create anatomically constrained functional ROI for bilateral EVC by selecting the 500 voxels inside the EVC anatomical ROI with the strongest loading on the sample regressor and for bilateral IPS0–5 and FEF by separately selecting the 500 voxels inside each of IPS0–5 and FEF anatomical ROIs, with highest loading on the Delay 1.2 regressor. We selected these ROIs to facilitate comparison with the original analyses of this data set (i.e., Yu et al., 2020), and because they cover regions known to be important for the delay-period representation and control of visual information in WM.

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

EEG Data Collection

The experimental procedure from the experiment reported by Fulvio and Postle (2020) entailed recording the EEG with concurrent delivery of spTMS on half of the delay periods of the DSR task. However, because the original report only included behavioral results (with and without spTMS), here, we detail the EEG methods.

EEG was recorded with a 60-channel cap and TMS-compatible amplifier, equipped with a sample-and-hold circuit that held amplifier output constant from 100 μs before stimulation to 2 msec after stimulation (NexStim eXimia). Electrode impedance was kept below 5 kΩ. The reference electrode was placed superior to the supraorbital ridge. Eye movements were recorded with two additional electrodes, one placed near the outer canthus of the right eye and one underneath the right eye. The EEG was recorded between 0.1 and 350 Hz at a sampling rate of 1450 Hz with 16-bit resolution.

Data were processed offline using EEGLAB (Delorme & Makeig, 2004) with the TMS-EEG signal analyzer open-source EEGLAB extension (Mutanen, Biabani, Sarvas, Ilmoniemi, & Rogasch, 2020; Rogasch et al., 2017) and Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2010) toolboxes in MATLAB (The MathWorks). The pipeline followed the TMS-EEG analysis pipeline (https://nigelrogasch.github.io/TESA/). Then, electrodes exhibiting excessive noise were removed and the data were epoched to −12 sec to 8 sec around the first spTMS event tag (Delay 1.2) and −4.5 sec to 4.5 sec around the second spTMS event tag (Delay 2). The data were downsampled to 500 Hz. To minimize the TMS artifact in the EEG signal, the data were interpolated using a cubic function from −2 to 30 msec around the TMS pulse, and this interpolation was also carried out on delay periods on which TMS was not delivered. (For delay periods for which no spTMS was delivered ["spTMS-absent"], a dummy spTMS event tag was added at a latency that matched the most recent spTMS-present delay period.) The data were bandpass filtered between 1 and 100 Hz with a notch filter centered at 60 Hz. Independent component analysis was used to identify and remove components reflecting residual muscle activity, eye movements, blink-related activity, residual electrode artifacts, and residual TMS-related artifacts. A spherical spline interpolation was applied to electrodes exhibiting excessive noise. Finally, the data were rereferenced to the average of all electrodes that were included in the independent component analysis.

The present analyses included EEG data from all delay periods (i.e., averaging data from spTMS-present and spTMS-absent trials and ignoring this factor).

Analysis Procedures

PCA Visualization of the RNN Hidden Layer Activity

We extracted from each network the activity of the 100 recurrent units from all 1000 testing trials (no noise added to the RNN during testing) and used PCA to project these 100-dimensional activity patterns onto the four dimensions accounting for the most variance across all training trials separately for each timestep. We then visualized the representations of each Sample 1 and Sample 2 by plotting the dimensionality-reduced activity across the 350-timestep time course of a trial, and coloring the activity patterns according to stimulus identity, separately, in three 2-D plots (PC1–2, PC2–3, and PC3–4).

In addition, we plotted the effective dimensionality (ED) of the data at each timepoint, which is the equivalent number of orthogonal dimensions that would produce the same overall pattern of covariation (Del Giudice, 2021). It is calculated using the following formula:

$$ED = \frac{\left(\sum_{i=1}^{N} \lambda_i\right)^2}{\sum_{i=1}^{N} {\lambda_i}^2}$$

where λ_is are the eigenvalues of the covariance matrix of the N recurrent units' activities at a certain time point.

Transformational Variability Analyses on EEG and fMRI Data

The PCA visualizations of RNN activity revealed representational dynamics, such that stimulus information was represented differently as a function of ordinal context (first, second) and as a function of cue identity 0, 1, or −1, corresponding to priority status of neutral, PMI, or UMI/IMI). To assess the functional relevance of these two coding schemes for human behavior, we assessed the trial-by-trial variability of context-based and priority-based transformations, and determined for each whether this variability related to variability in behavior.

For the representation of context, we first calculated a template stimulus representational format for each participant by averaging the neural activity for each context status ("1st" or "2nd" for fMRI; "up" or "down" for EEG) over a time window corresponding to Delay 1.1, across all trials. (For the remainder of this section, for simplicity, we will only refer to ordinal context.) To these two windowed averages, we applied demixed PC analysis (refer to Wan et al., 2022, and Kobak et al., 2016, for methodological details) to derive the first two demixed PCs, thereby constructing a Sample 1 template subspace and a Sample 2 template subspace. We then projected individual trial activity from the same time window into the template subspaces and calculated the "transformational variability index" (TVI) for that trial's representational transformation into the Sample 1 subspace and its representational transformation into the Sample 2 subspace. TVI was defined as the Euclidean distance between that trial's representation in the subspace and the template representation, normalized by the distance between the two template representations in that subspace. (For example, for trial n, the TVI for the Sample 1 subspace would be the Euclidean distance between the trial representation

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

projected into the Sample 1 subspace and the Sample 1 template [projected into the Sample 1 subspace], divided by the distance between the Sample 2 template projected into the Sample 1 subspace and the Sample 1 template [projected into the Sample 1 subspace].) For the fMRI data, we used repetition time (TR) 5–7 (8–14 sec) to define the Delay 1.1 subspaces, and for the EEG data, we used the entirety of Delay 1.1. (See Appendix Figure A1 for a mapping of epochs of interest to a timeline demarcated in seconds and TRs.)

For priority-based transformations, we followed the same procedures, but used TR 9–11 (16–22 sec) to define the Delay 1.2 subspaces and the entirety of Delay 1.2 period for the EEG data and labeled the data according to priority status (i.e., PMI and UMI).

If the efficacy of a context-based transformation is important for behavior, smaller TVIs, corresponding to lower trial-to-trial variability, should be associated with superior performance. To assess this in the fMRI data, for each participant, we sorted responses separately for Recall 1 and for Recall 2, by median split of angular error, and then calculated, for each response, the average TVI for each type of transformation (e.g., "What was the average TVI for the transformation to *Sample 1* for low-error vs. high error responses to *Recall 1*?"). Then, we performed paired-samples *t* tests between group-average, high-error and low-error TVIs, separately for each subspace, each brain region, and each response (Recall 1 and Recall 2). The analysis procedure was similar for the EEG data except that the comparison was between incorrect and correct responses.

To test how the TVI for UMI and PMI covary, we ran two-sided, Spearman's rank correlations between the two metrics across all trials for each participant and counted the number of participants with correlations reaching the significance level of $\alpha = .05$.

Within- and Cross-label Decoding of RNN and fMRI Data

To assess where (and how) in the brain context and priority are represented, we carried out a series of decoding analyses, the inferential logic of which is detailed in Results, Analyses of fMRI and EEG Data, and Within- and Cross-label Decoding of RNN and fMRI Data sections. To generate RNN data in a format consistent with the fMRI data set, RNN data were generated by testing the trained network on 324 trials of nine possible orientations (counterbalanced across the identities of Sample 1, Sample 2, Cue 1, and Cue 2, to be analogous to the Yu et al. [2020] task), and subsequently extracting the RNN hidden layer activity. For the RNN data, we decoded orientation, and for the fMRI data, we decoded location. (Decoding item location is generally more sensitive than decoding item orientation, and so demonstrations of failures of cross-label decoding of item location would provide stronger evidence for the encoding of the stimulus property of interest.)

For the RNN data and for the fMRI data from each ROI, we trained linear support vector machine (SVM) multiclass classifiers to decode stimulus identity with a *k*-fold, cross-validation procedure and a "one vs. one" coding design (see Appendix Figure A5 for comparisons with results from other decoding methods). For context-based decoding, for each participant and at each timepoint, we trained a classifier with the data labeled as Sample 1 and then tested it on the data labeled as Sample 1 (within-label decoding) and with the data labeled as Sample 2 (cross-label decoding). We then repeated this process by training on Sample 2, and with fMRI data, for simplicity, we averaged the results to generate the overall accuracies for within-label decoding and cross-label decoding. For priority-based decoding, we used the same procedure except that the labels were PMI and UMI, instead of Sample 1 and Sample 2, the PMI/UMI label reassigned at Timestep 301 (for RNN) or TR 15 (28–30 sec; for fMRI) to reflect identity of Cue 2 (i.e., to account for the fact that priority status changed partway through "switch" trials). (The choice of TR 15 as the timepoint to reassign priority status label was based on the following reasoning: (1) TR 15 reflects Cue 2 onset (23.25 sec) after accounting for the hemodynamic delay (~6 sec); (2) data from a different fMRI study of DSR, Teng and Postle (2024), indicates that the IEM time courses for "stay" and "switch" trials intersect at TR 15. See Appendix Figure A1 for a task diagram that maps task epochs to a timeline that is demarcated in seconds and TRs. For the fMRI data, to evaluate the significance of decoding accuracy against chance level (1/9), we performed one-tailed, one-sample *t* tests against 1/9 on decoding accuracies across all participants, and corrected for multiple comparisons using the false discovery rate (FDR) method.

RESULTS

RNN

We trained 10 RNNs using the same training scheme, an initial one for hypothesis generation and then nine more for replication.

Performance

After training, the range of mean response error for the 10 networks was 0.33°–0.89° and the range of standard deviations was 0.29–0.66. (For comparison, for human participants performing this task in the fMRI study [Yu et al., 2020], the mean error was 16.84° [$SD = 4.49$]).

PCA Visualization of Hidden Layer Activity

The dynamical representational patterns observed in all 10 networks were highly consistent, and the results from the third network are reported here in detail. (See appendices for results from other networks.) PCA was carried out

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

on the RNN hidden layer activity across all timepoints from 1000 withheld testing trials with Sample 1 and Sample 2 spanning the [0°, 180°] angular range and the resultant dimension-reduced activity projected into three subspaces that were spanned by PC1–PC2, PC2–PC3, and PC3–PC4, respectively, on a timepoint-to-timepoint basis (Figure 3; see Appendix Figure A2 for the time courses of percent variance explained by each of the

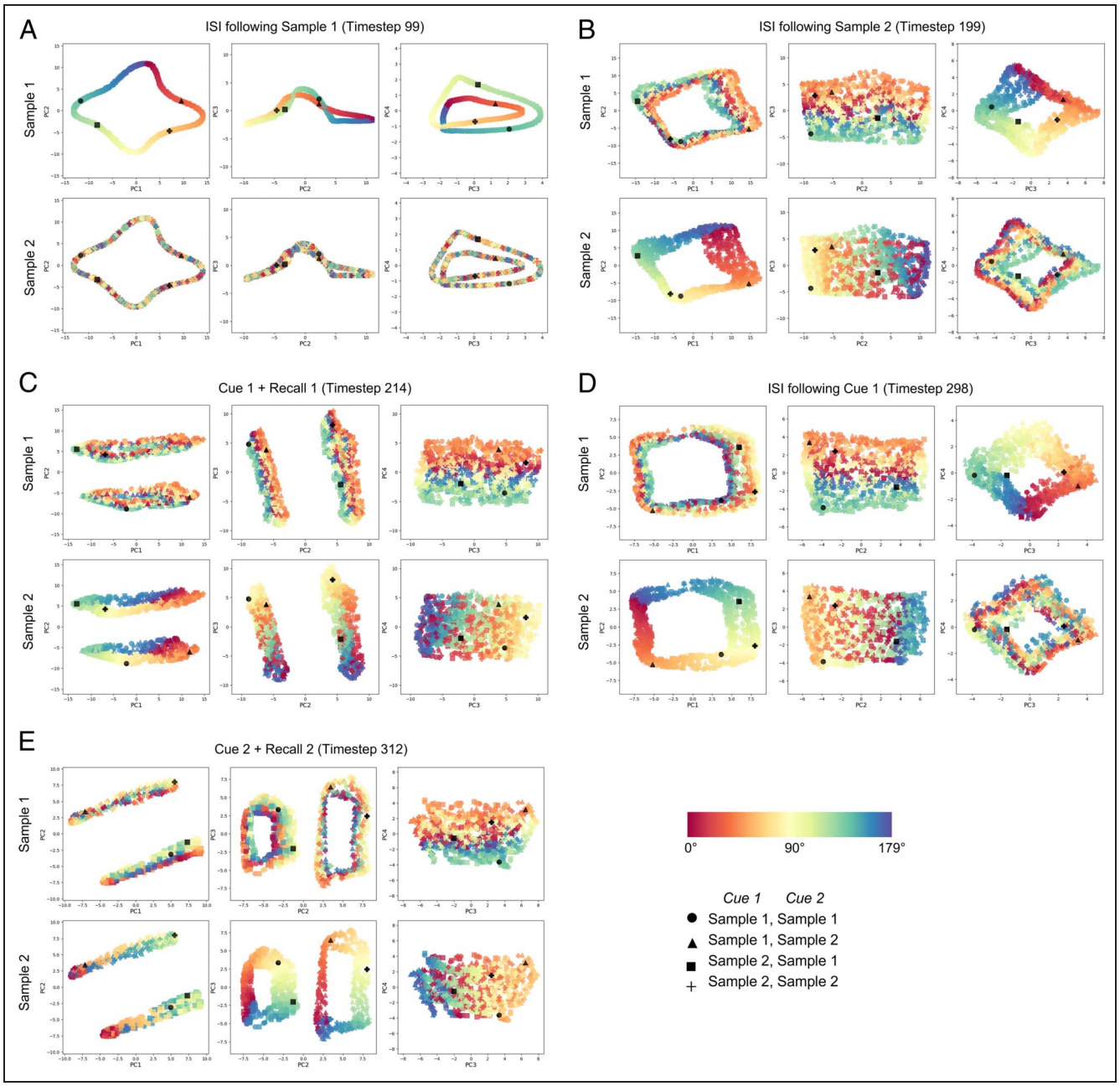

Figure 3. Visualization of representational dynamics embedded in hidden layer of RNN No. 3 at each of five representative timesteps across the DSR task. Each plot contains 1000 data points, one corresponding to each simulated trial, and the symbols indicating that trial's cue configuration: *Cue 1 - > Sample 1, Cue 2 - > Sample 1* (●); *Cue 1 - > Sample 1, Cue 2 - > Sample 2* (▲); *Cue 1 - > Sample 2, Cue 2 - > Sample 1* (■); *Cue 1 - > Sample 2, Cue 2 - > Sample 2* (+). In each plot, an example trial of each cue configuration is colored black for better visualization. For each of the five timesteps, the same data are illustrated in six ways: the top row with the data labeled as Sample 1 and the bottom row with the data labeled as Sample 2, and for each, they are projected into three subspaces. (A) After the presentation of Sample 1 (Timestep 99). Note that because Sample 2 has not yet been presented, the stimulus values are haphazard. (B) After the presentation of Sample 2 (Timestep 199). With both items in WM, but before cuing, Sample 1 is now represented in the PC3–PC4 subspace and Sample 2 in the PC1–PC2 subspace. (C) During the presentation of Cue 1 and generation of Recall 1 (Timestep 214), illustrating a separation-by-priority status in the PC1–PC2 subspace. (A comparable priority-based separation was visible in the PC3–PC4 subspace earlier during this same epoch [not shown].) (D) During the delay between Cue 1 and Cue 2 (Timestep 298). (E) During presentation of Cue 2 and generation of Recall 2 (Timestep 312), again illustrating a separation-by-priority status in the PC1–PC2 subspace but now based on Cue 2. (As with the Cue 1 epoch, a comparable priority-based separation was visible in the PC3–PC4 subspace earlier during this Cue 2 epoch [not shown].)

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

four PCs; movies of network dynamics can be found at https://osf.io/tnh9x/).

Upon the presentation of Sample 1, its representation formed a ring in the subspace spanned by the first two PCs, with relative distances between stimulus values preserved (as shown by the smooth color gradient of the ring; Figure 3A, top left), such that stimulus value can easily be read out from this subspace. Although there are also smooth color gradients in the other two subspaces, their geometry is more complex, making it less clear if they would support readout. The ring structure in the PC1–PC2 subspace was maintained across the ensuing ISI (see Figure 3A and Movie 1 in the OSF repository [https://osf.io/tnh9x/]). After the presentation of Sample 2, Sample 2's identity was represented in the PC1–PC2 subspace, also in the form of a ring with a smooth color gradient (although the ring was somewhat "stretched out" relative to Timestep 99; Figure 3B, bottom left). In parallel, information about Sample 1 emerged in the subspace spanned by PC3 and PC4, in the shape of a ring with a smooth (albeit "stretched out") color gradient (Figure 3B, top right). In effect, whereas Sample 1 was represented in the PC1–PC2 subspace when it was the only item in WM, it was shunted to the PC3–PC4 subspace with the presentation of Sample 2, which replaced Sample 1 in the PC1–PC2 subspace. Thus, before cuing, the RNN encoded the ordinal context of Sample 1 and Sample 2 by segregating them in orthogonal subspaces.

Upon the presentation of Cue 1 (at Timestep 201), the stimulus representations within each subspace separated into two clusters that were defined by priority status. For example, Figure 3C illustrates that in the PC1–PC2 subspace, at Timestep 214, trials for which Sample 1 was cued (denoted by triangle and circle symbols) separated from trials for which Sample 2 was first cued (square and plus-sign symbols). Throughout the Cue 1 epoch, the axis along which this separation occurred rotated in multidimensional space over time. Thus, whereas Timestep 214 was selected for Figure 3C because it clearly shows this separation-by-priority status in the PC1–PC2 subspace; the separation was visible in the PC3–PC4 earlier during this epoch, at Timestep 207 (see Movie 1 in the OSF repository [https://osf.io/tnh9x/]). Thus, the RNN encoded priority status via separation within each subspace.

During the delay between Cue 1 and Cue 2 (Timesteps 251–300), the prioritization clusters merged such that, before the presentation of Cue 2, information about Sample 1 and Sample 2 was again clearly observed in the PC3–PC4 subspace and in the PC1–PC2 subspace, respectively (Figure 3D). Finally, upon the presentation of Cue 2, the network representation once again separated into two priority-defined clusters, this time based on Cue 2's identity, (i.e., trials for which Sample 1 was cued [denoted by circle and square symbols] and trials for which Sample 2 was cued [triangle and plus-sign symbols] separated into two clusters; Figure 3E). Thus, visualization of the representational of the RNN recurrent unit activities revealed that context and priority were represented via different transformational mechanisms, the former via the *segregation* of stimuli to orthogonal subspaces, and the latter via *separation* within each subspace.

ED

During the processing of Sample 1, ED initially rose to a value between 3 and 4 before declining to a value of ≈ 2 during the ensuing ISI (Figure 4). Upon the presentation of Sample 2, ED rose precipitously to a value close to 6 before declining steadily for the remainder of this epoch and the ensuing Delay 1.1 to a value just below 3, which corresponds well to the encoding of a new stimulus and the segregation of subspaces to represent the ordinal context. The three remaining trial epochs were characterized by an initial increase of ED to a value of roughly 4 followed by a decline back to roughly 3. Particularly noteworthy in these results is the increase in ED following the offset of Cue 1. Note that because a similar increase in ED was not observed upon the offsets of the Stimulus 1 or Stimulus 2 epochs, this effect cannot be simply because of a transition from one epoch to the next. Rather, this effect closely resembled those time-locked to the onset of Cue 1 and to the onset of Cue 2, events that each prompted the separation of stimuli into priority-defined clusters (Figure 3C and 3E). Therefore, it may be that the operation of removing from the network the encoding of no-longer-relevant information about priority status related to Cue 1—corresponding to the merging of priority-defined clusters that was observed in the PCA visualization—is also an operation that entails a transient increase in ED.

Interim Discussion

We trained RNNs to perform the DSR task and applied dimensionality reduction to the internal representations of the network. Visualization of the representational dynamics yielded several important insights. First, upon

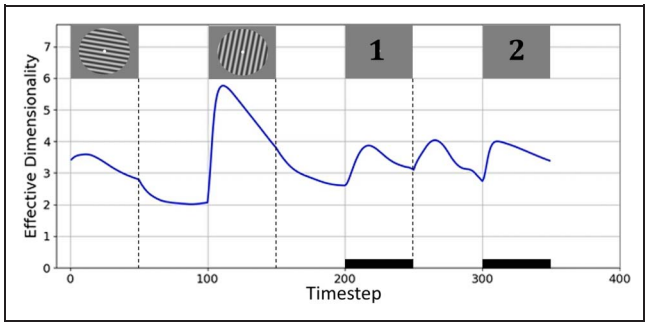

Figure 4. The time course of ED of the hidden layer stimulus representations of RNN No. 3. The rectangular images above the curve denote corresponding task events. The black rectangles along the *x* axis indicate time periods when a response was being made. (See Appendix Figure A3 for ED time courses from other networks.)

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

presentation of the second sample (and, therefore, before the first prioritization cue), information representing the first sample underwent a rotational transformation into an orthogonal subspace, effectively segregating the two representations (cf. Panichello & Buschman, 2021). We will refer to this process as the *segregation* of the two representations, and speculate that it may have served not only to individuate the two but also to encode the distinct ordinal context that the network needed to correctly interpret the cues. Second, the encoding of priority status was accomplished by stratification of each subspace and translation of the stimulus representation within each to one of two "priority-specific" strata, a PMI stratum and a UMI stratum. We will refer to this process as the *separation* of stimulus information as a function of priority, and speculate that only the PMI stratum within each subspace was amenable to readout by the output layer. The first observation is important because it emphasizes the importance of encoding trial-specific context in WM, an operation that has been underemphasized in previous empirical studies of prioritization. The second observation is important because it indicates that the representation of a priority status may be implemented in a different way than is ordinal context—via separation within a subspace (via translation) versus via the segregation of stimulus information to distinct subspaces (via rotation), respectively.

Because ordinal context and priority are orthogonal factors in the experimental design, it might seem intuitive that they would be encoded differently by an RNN. However, for the RNN, they need not represent two qualitatively different factors. Alternatively, they may be construed two dimensions of task context that play out on different time scales during a single trial. To elaborate, in this variant of DSR, one dimension of an item's context is the order in which it was presented. This can be considered the "first-order" context because it uniquely individuates an item for the duration of a trial, and it does not change for the duration of the trial. (It is to such "first-order" context that Oberauer and Lin [2017] refer when they state that the binding of context to a stimulus is fundamental to that stimulus being in the state of being "in WM".) A second dimension of context is priority status, and this differs from first-order context because its status can change multiple times within the trial, between "neutral," "prioritized," and "unprioritized" (indicated by the values of 0, 1, and −1, respectively, that are input by the cue unit). Thus, priority serves as a "second-order" level of context, one that indicates an item's in-the-moment status with respect to the rules of the task and that cannot be interpreted in the absence of information about first-order context. These considerations highlight that to fully understand the flexible control of WM, we need to understand how first-order context is coded in the brain and how it interfaces with higher-order context to guide thought and action.

Recent empirical studies that have manipulated demands on first-order context in WM have implicated regions of frontal cortex and the intraparietal sulcus (IPS; for ordinal context, see Fulvio, Yu, & Postle, 2023; Gosseries et al., 2018; for location context, see Fulvio et al., 2023; Cai, Fulvio, Yu, Sheldon, & Postle, 2020). In Yu and colleagues (2020), a study that also manipulated priority, the location context of differently prioritized orientation stimuli was found to be preferentially coded in IPS, and not EVC, although location information was not directly tested by the task. More recently, Teng and Postle (2024) used the same stimuli and procedure, but flipped the roles of context and content, making orientation the first-order context used to cue memory of an item's location. "Context load" was manipulated via the similarity of orientation of the two sample stimuli, and individual differences in context-load sensitivity of activity in IPS (but not EVC) predicted behavioral sensitivity to this factor. Generalizing across these studies suggests that first-order context in WM may be represented more prominently in areas associated with cognitive control than in areas associated with stimulus representation. The same may not be true for second-order context, because prioritization effects are prominent in EVC (Teng & Postle, 2024; Yu et al., 2020).

An implication of these considerations is that the results from the RNNs have highlighted a distinction in information encoding—the representation of first-order versus second-order context—that has previously been underappreciated in cognitive neuroscience research. (For example, in the 2-back task simulated by Wan and colleagues [2022], when an item had the status of UMI, it also had the contextual status of *item-that-was-presented-most-recently* [i.e., "1-back"], and when it then transitioned to the status of PMI, its context simultaneously transitioned to that of *item-that-was-presented-2-back*.) What follows, therefore, are initial attempts to evaluate the relevance of these RNN results for cognition in the human brain, via reanalyses of an extant fMRI and an extant EEG data set from two previous studies of the DSR task.

Analyses of fMRI and EEG Data

The goal of these reanalyses was not to replicate with neuroimaging data the same dimensionality-reduction analyses from the RNNs. Previous experience indicates that, for example, applying demixed PC analysis to human EEG data does not yield easily interpretable, ring-like, representational structure such as seen in Figure 3 (cf. Wan et al., 2022; Figure 6). This is likely due, in part, to the fact that, unlike RNNs, human brains are concurrently engaged in many processes unrelated to the experimental task (e.g., constantly processing information about the environment that is external to the experimental stimuli). Rather, these reanalyses were intended as initial assessments of the relevance of the RNN results for human cognition. Specifically, they addressed two questions about the putative distinction between the encoding of first-order context versus higher-order context (here

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

operationalized as priority): *Are they dissociable in terms of their effect on behavior?* and *Are they dissociable in terms of where (and how) they are represented in the brain?* The fMRI study used in these reanalyses used a DSR procedure that was most closely matched to that used with the RNN, including the fact that it used stimulus order as first-order context. The fMRI data would also allow for assessment of possible regional differences in the representation of the two types of context. In contrast to the fMRI task, the task used in the EEG study used location as the dimension of first-order context, and so would allow an assessment of generalization of what has been observed for ordinal context (with the RNN and fMRI) to location context. (For ease of exposition, in the results that follow, we will refer to first-order context as "context" and second-order context as "priority," because priority is the only dimension of second-order context that is relevant in the DSR task.)

Transformational Variability

One way to compare the neural representation of context versus priority is to assess their influence on behavior. To do this, we took an individual differences approach, using the variability of trial-by-trial encoding of context and of priority as proxies of the efficacy with which these operations were carried out (i.e., a participant for whom context-based or priority-based transformations were more variable from trial-to-trial might be expected to perform worse on the task).

For context-based transformations, results failed to show evidence that behavior was sensitive to transformational variability. For the fMRI data (ordinal context), TVIs did not differ for low- versus high-error trials, for Recall 1 or Recall 2, in any of the three ROIs (EVC, IPS 0–5, and FEF), all $t(12) < 1.74$, *n.s.* For the EEG data (location context), TVI did not differ for correct versus incorrect trials, $t(11) < 1.37$, *n.s.*

For priority-based transformations, in contrast, there was considerable evidence that behavior was sensitive to transformational variability. For the fMRI data, in EVC, TVI was lower for low-error than high-error trials for the UMI subspace for Recall 1, $t(12) = 1.81$, $p = .048$, and for the PMI subspace for Recall 2, $t(12) = 2.06$, $p = .031$. For IPS0–5, TVI for the PMI subspace was lower for low-error than high-error trials for Recall 1, $t(12) = 2.04$, $p = .032$, and was lower for low-error than high-error trials for both UMI, $t(12) = 3.04$, $p = .005$, and PMI, $t(12) = 3.00$, $p = .006$, subspaces for Recall 2. All other comparisons, including all for FEF, failed to achieve significance, all $t(12) < 1.57$, *ns.* For the EEG data, TVI was lower for correct trials than incorrect trials for Recall 1 in both UMI, $t(11) = 2.17$, $p = .027$, and PMI, $t(11) = 4.28$, $p < .001$, subspaces (Figure 5).

Additional analyses carried out at the single-subject level indicated that, for a subset of participants, trial-by-trial variation in TVI predicted performance (see spTMS A1).

The TVI also offered a metric with which to begin exploring whether the transformation to PMI and the transformation to UMI may share a common component that acts on the two simultaneously (cf. Panichello & Buschman, 2021). Specifically, we correlated trial-by-trial TVI for the PMI with trial-by-trial TVI for the UMI (two-sided), reasoning that evidence of correlation would be expected if the two do share an underlying mechanism. For the fMRI data, in EVC, this correlation was significant at $p < .05$ for 12 out of 13 participants, in IPS 0–5 for 11 participants, and in FEF for 10 participants. For the EEG

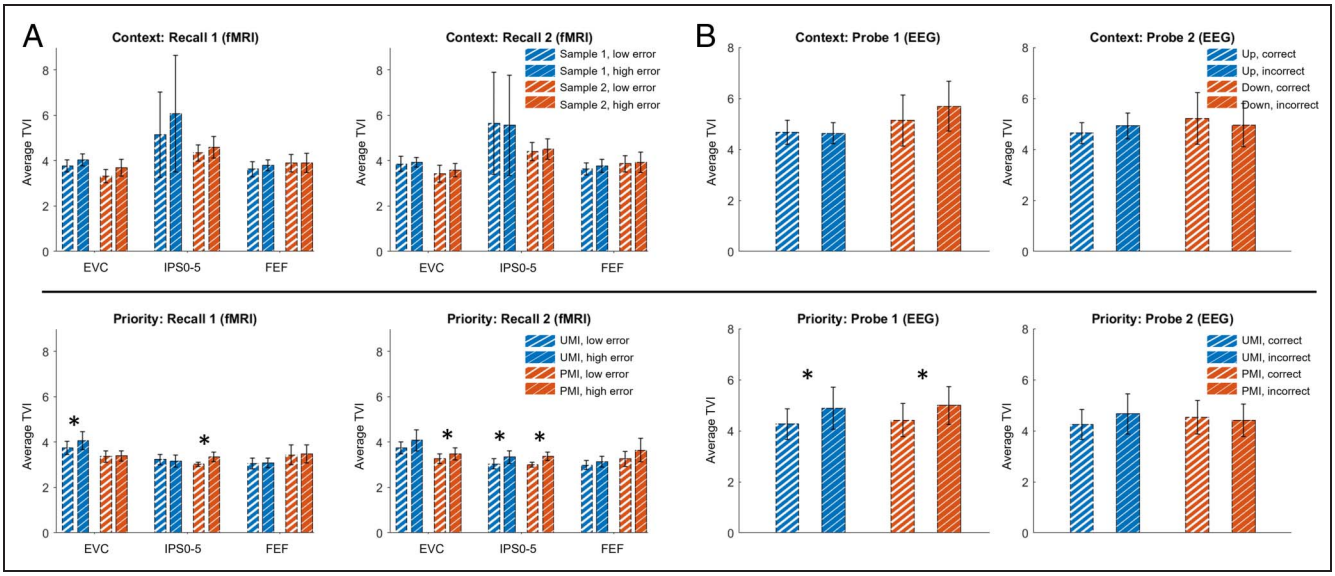

Figure 5. Transformational variability analysis results on fMRI (Yu et al., 2020) and EEG (Fulvio & Postle, 2020) data. (A) Comparisons between average TVI for high-error and low-error trials across participants from the fMRI data set. (B) Comparisons between average TVI for incorrect and correct trials across participants from the EEG data set. Top row: priority-based decoding; bottom row: context-based decoding. The subspace from which the TVI is calculated is indicated in the legends. Asterisks above bars of the same color indicate the significance level of the paired-samples *t* tests comparing the average TVI between each two groups.

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

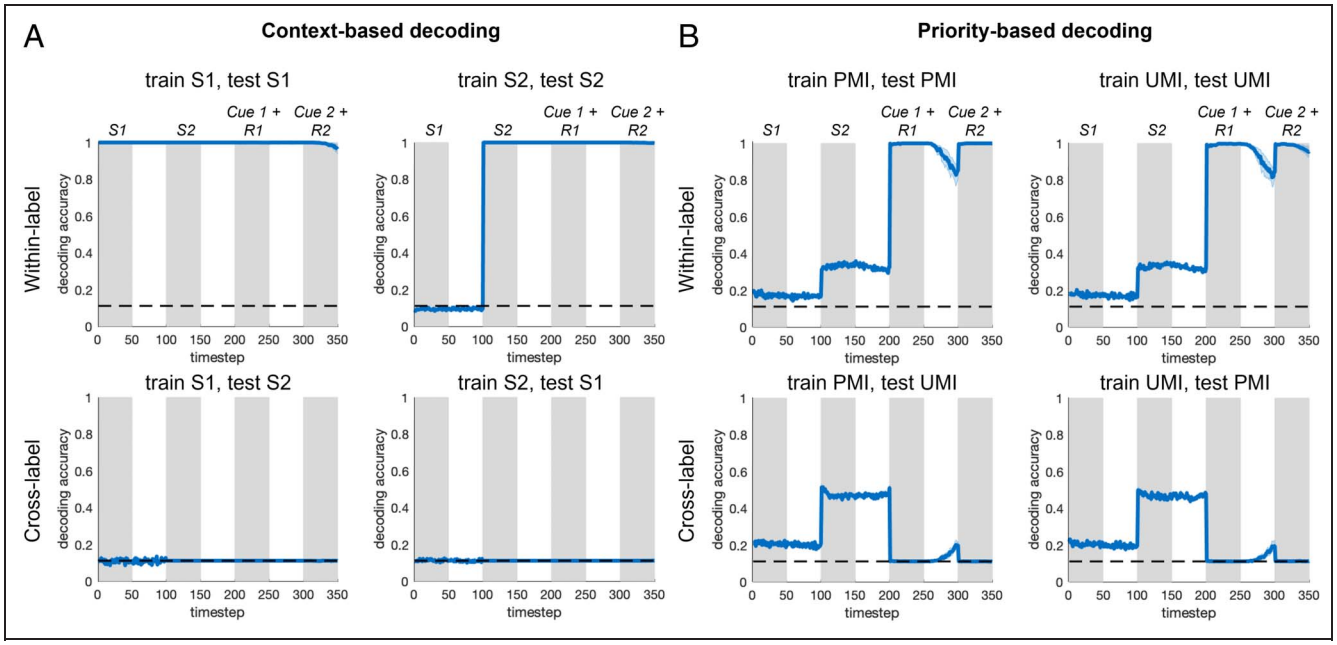

Figure 6. Within- and cross-label decoding of stimulus identity averaged across the 10 RNNs. (A) Context-based decoding. Classifiers were trained on Sample 1/2, then tested on Sample 1/2 (within-label), or tested on Sample 2/1 (cross-label). (B) Priority-based decoding. Classifiers were trained on PMI/UMI, then tested on PMI/UMI (within-label), or tested on UMI/PMI (cross-label). Solid lines correspond to average classifier accuracy; shaded error bands correspond to ± 1 *SEM*. S1 = Sample 1, S2 = Sample 2; R1 = Recall 1; R2 = Recall 2.

data, TVIs for PMI and UMI were significantly correlated for 10 out of 12 participants. All correlations were positive.

Within- and Cross-label Decoding of RNN and fMRI Data

To address the question of how stimulus context and stimulus priority are represented in the brain, we applied the following logic. If a region represents context, any given stimulus item will be represented differently when it has the status of, for example, Sample 1 versus when it has the status of Sample 2. If a decoder applied to data from this region can be successfully trained to classify stimulus identity when the data are labeled as Sample 1 (successful "within-label" decoding), it should fail to decode stimulus identity when the data are relabeled as Sample 2 (unsuccessful cross-label decoding). If stimulus information in this region does not incorporate a representation of context, in contrast, any given item's representational format will not differ as a function of its context status, and so a decoder that can be successfully trained on the data labeled as Sample 1 should succeed at decoding stimulus identity when the data are relabeled as Sample 2 (successful cross-label decoding). (Note that it would not be possible to use this approach to decode an abstract representation of context, independent of stimulus information, in the Yu and colleagues [2020] data set, because a first and a second item were presented on every trial and, furthermore, they were presented within the same TR.)

RNN data. Because RNNs can be trained to perform the DSR task, it is necessarily true that they represent both context and priority, and so we first carried out within- and cross-label decoding of the RNN data to validate the inferential logic of this approach, intended to assess the representation of context and of priority in the fMRI data set. We performed these analyses on the RNN recurrent unit activities across all 350 timesteps from 324 novel, counterbalanced trials of nine different orientations using a linear SVM classifier (Figure 6).

For context-based decoding, we obtained close to perfect decoding accuracy when training and testing on the labels of the same sample throughout the task (note that for train S2, test S2 decoder performance was at chance before Timestep 101, because of the absence of information about Sample 2 at those time steps). For cross-label decoding, however, accuracy was at chance level for the duration of the trial. For priority-based decoding, within-label decoding accuracy for both PMI and UMI was close to chance level before Cue 1. With the onset of Cue 1, for both PMI and UMI, decoder performance rose to close-to-perfect for the remainder of the trial. For cross-label decoding, whereas decoding accuracy for both PMI and UMI was above chance level before Cue 1, for both, it dropped to chance level with the onset of Cue 1 and remained there for the remainder of the trial. Both of these sets of results validated the reasoning that a system that represents context and priority would not support cross-label decoding for either factor.

fMRI data. We investigated the anatomical distribution of the representation of context and priority during the DSR task by carrying out a series of multiclass decoding analyses on the fMRI data set (Figure 7). In general (and

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

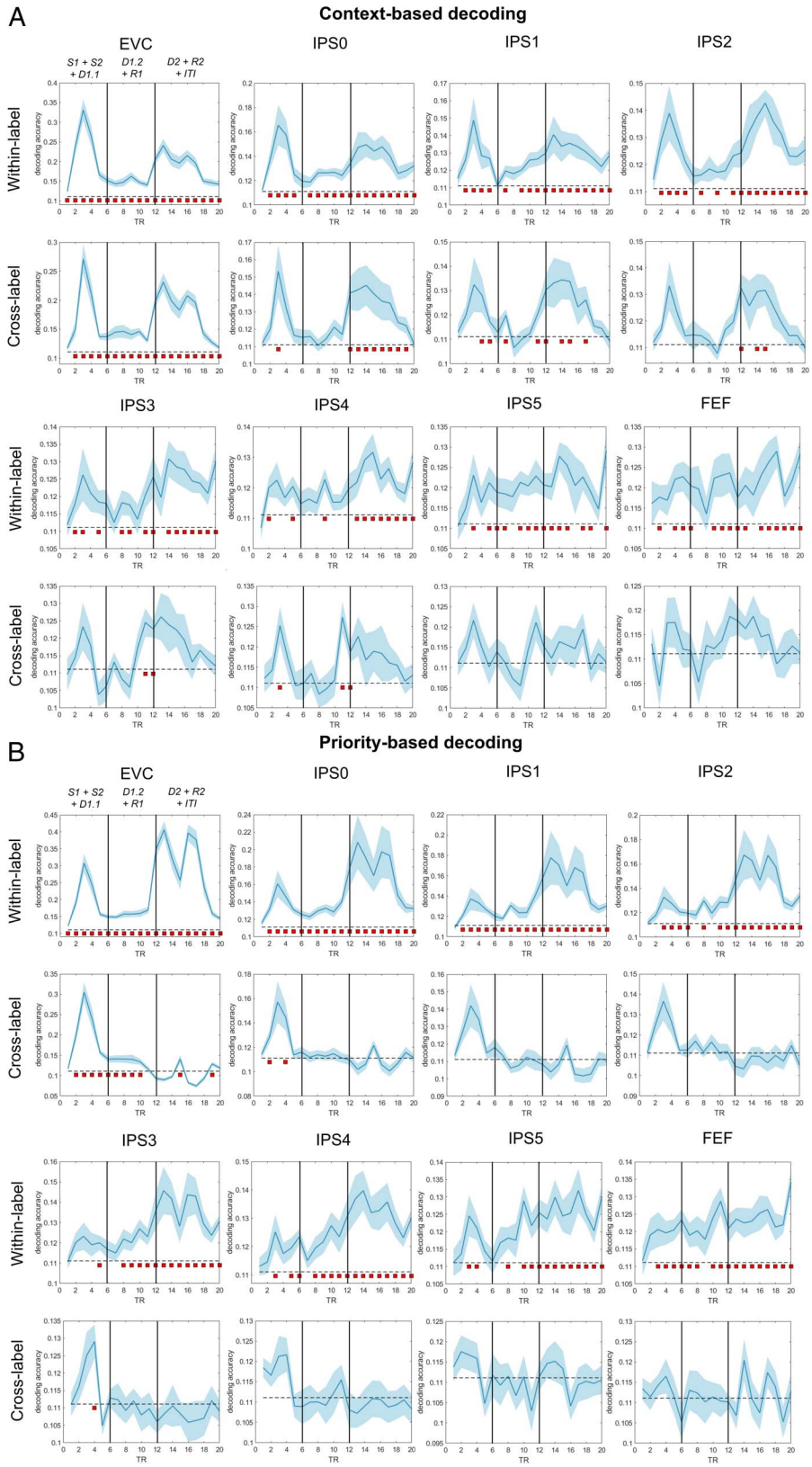

Figure 7. Within- and cross-label decoding analyses from the fMRI data set. (A) Context-based decoding. (B) Priority-based decoding. In each graph, the two vertical solid black lines indicate Cue 1 and Cue 2, respectively. The blue shading around each curve shows the standard error of the mean. The horizontal dashed line indicates the chance-level decoding accuracy of 0.11. Red squares below the dashed line indicate time points with significant above-chance decoding accuracy ($p < .05$, FDR-corrected across all time points). Note that the range of the y axis varies from graph to graph. S = sample; D = delay; R = recall.

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

unlike for the RNNs), decoder performance was far from ceiling and tended to be superior for time points corresponding to trial epochs when stimuli were on the screen. Importantly, however, we were generally able to decode the stimulus identity across the whole time course with above-chance accuracy in every ROI, especially in the time period between Cue 1 and Cue 2, where one stimulus is prioritized over the other in WM (within-label rows of Figure 7). (The one exception was in IPS4 with context-based decoding; the reason for this is unclear.)

For context, cross-label decoding revealed a marked posterior-to-anterior gradient: It was successful for the entirety of the trial in EVC; successful for Cue 2 and Delay 2 epochs for IPS0 and for a smaller number of timepoints for IPS1 and IPS2; successful only for late Delay 1.2 for IPS3 and IPS4, and entirely at chance for IPS5 and FEF (Figure 7A). These results indicate that context was not represented at the earliest stations of the visual system and became progressively more prominent at progressively higher levels of the dorsal stream.

For regions that do represent stimulus context, we can also use results from these analyses to carry out exploratory assessment of alternative accounts of how context is encoded. In particular, with reference to the results from the RNNs, is there evidence that the segregation of stimulus representations by context is accomplished by their anatomical segregation? (That is, in the RNNs, the presentation of Sample 2 prompted the rotation of Sample 1 out of the PC1–PC2 subspace and into the PC3–PC4 subspace [Figure 3B]. Might such an operation be accomplished in the brain by rerepresenting Sample 1 in a different anatomical region than where it had been represented before the presentation of Sample 2?) To assess this, we inspected the results from within-label decoding for train Sample 1, test Sample 1, and train Sample 2, test Sample 2, and looked for regions that supported within-label decoding of one but not the other. The results did not provide strong evidence consistent with a separation-by-region account: Decoding evidence was robust for both Sample 1 and Sample 2 in EVC, IPS0, and IPS1, and progressively weaker for both in more rostral ROIs. In IPS3, IPS4, and IPS5, there was a trend toward more robust within-label decoding for Sample 1 than for Sample 2 during the second half of the trial, but nowhere was there evidence for the opposite trend (Appendix Figure 4A). In the Discussion section, we will return to the question of context-related transformations in the brain.

For priority, cross-label decoding for EVC was successful for the beginning of the trial through late Delay 1.2, after which it dropped to a level numerically below baseline, with the exception of two isolated time points during the second half of the trial. (Note that although planned statistical comparisons were one-tailed, post hoc two-tailed tests suggested that cross-label decoding for several time points beginning with TR 12 were statistically below chance.) For the remainder of the ROIs, cross-label decoding was at baseline for the entirety of the trial. This suggests that priority is represented in every ROI that we investigated, albeit taking longer to manifest in EVC (Figure 7B). (See Appendix Figure 4B for within- and cross-label, priority-based decoding broken out for PMI and for UMI.)

DISCUSSION

In this study, we initially set out to investigate the mechanisms underlying prioritization on a task in which changes of priority were not predictable—the DSR task—via visualization of representational dynamics of RNNs trained to perform the task. Unexpectedly, results from the RNNs called to our attention the importance of also understanding the representation of an additional dimension of trial-specific information, the first-order context that uniquely individuates each item during the trial. Across model training, validation, and testing, we saw that the encoding of first-order context was accomplished via the segregation into orthogonal subspaces of the representation of the first and second items to be presented. Unlike first-order context, higher-order context can change within a trial, a property that is often manipulated with instructional cues. In the DSR, priority status is the second-order context, and it is specified, then removed, and then specified a second time, during the course of each trial. The encoding of priority was accomplished via the stratification of each context-encoding subspace into priority-based strata and the concomitant translation of the stimulus representation to the appropriate one. Thus, the RNN indicated that first- and second-order contexts are encoded via distinct mechanisms, *segregation* (via rotation) to orthogonal subspaces versus *separation* (via translation) within a subspace, respectively. Furthermore, an ED analysis suggested that the operation of resetting second-order context (as happens during the ISI separating Cue 1 and Cue 2 in the RNN version of the task) may make computational demands that are comparable (in terms of requiring additional dimensions) to those needed to establish it.

Because of their architecture, it was necessarily the case that RNNs would represent both stimuli, and both context-dependent subspaces, within the same population of 100 units in the hidden layer. In a mammalian brain with multiple distinct regions, however, this need not be the case. It could be possible, for example, that whereas the first item to be presented in the DSR task is initially represented in EVC, its representation would get "rewritten" to a different region (e.g., in the IPS) upon the presentation of the second item. The results from reanalyses of the fMRI data set, however, do not support this account for either first- or second-order context. When considering first-order (i.e., ordinal) context, it is first important to note that the cross-label decoding results failed to find evidence for its incorporation into stimulus representations in EVC. In IPS0–2, however, successful within-label decoding for both Sample 1 and Sample 2 is inconsistent with a

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

separation-by-region model. For second-order context (i.e., priority), the combination in EVC of the failure of cross-label decoding for much of trial plus successful within-label decoding for both the PMI and the UMI suggests that both states of priority are represented at this earliest level of visual representation. Indeed, the pattern of statistically negative cross-label decoding during Delay 1.2 and Delay 2 may reflect the same factors that produced "opposite" results with multivariate analyses in previous studies of DSR (Wan et al., 2020; Yu et al., 2020; van Loon et al., 2018). Consistent with the present results, previous studies have reported the simultaneous representation, within the same population of neurons, of items in subspaces associated with different contexts (e.g., in auditory cortex of the mouse (Libby & Buschman, 2021) and in pFC of the nonhuman primates (Panichello & Buschman, 2021).

Consistent with the distinct dynamics observed with RNNs, reanalyses of an fMRI and an EEG data set established that the processing of first- and second-order contexts has distinct behavioral and neural profiles for humans performing the DSR. To assess relations to behavior, dimensionality reduction was applied to the neural data and TVI derived for each participant for each of the two levels of first-order context and for each the two levels of second-order context (i.e., priority). Correlations with behavior failed to show any evidence that performance is sensitive to variation in TVI for first-order context, whether defined by ordinal position (fMRI study) or location (EEG study). For second-order context (i.e., priority for both the fMRI and EEG studies), in contrast, there was considerable evidence that larger TVIs (indicating higher trial-to-trial variability) corresponded to poorer performance. In the fMRI data set, the anatomical distribution of the representation of order and priority also differed, with the former absent from EVC and becoming progressively more robust in more rostral ROIs, whereas the latter was evident in every ROI that we investigated. Thus, our results suggest that not only are representational transformations corresponding to first-order versus second-order context implemented via different mechanisms, they also differ according to their influence on behavior and to their distribution in the brain.

These results share some similarities and some differences with a recent study that recorded neuronal activity from several brain areas of nonhuman primates performing a single-retrocue WM task (Panichello & Buschman, 2021). In that study, dimensionality reduction revealed that, before the retrocue, the two stimuli were represented in orthogonal subspaces that corresponded to the location at which each had been presented (i.e., to their first-order context). Upon cueing, stimulus information transformed into different "postselection" subspaces that retained first-order context and now also represented selection status (selected/nonselected; i.e., second-order context). Notably, the representations of "selected upper" and "selected lower" items were no longer orthogonal. The degree of cue-triggered representational transformation was highest in dorsolateral pFC and progressively weaker in more posterior regions, weakest in extrastriate visual area V4. One similarity of those results to those reported here is the initial encoding of first-order context into orthogonal subspaces. A notable difference between the two is the nature of the postcue transformations. This difference is also seen in a modeling study that simulated the Panichello and Buschman (2021) study using an RNN architecture similar to the one presented in this report (Piwek, Stokes, & Summerfield, 2023). Specifically, and at variance with what we report here, Piwek and colleagues (2023) found that the representation of the uncued item became more compressed, and thus less discriminable, in comparison to its initial state. This discrepancy is most likely because of the differing demands of the single-retrocuing task (Piwek et al., 2023; Panichello & Buschman, 2021) versus the DSR, in that only for the DSR does it remain possible that the initially uncued item might be needed later in the trial. Importantly, this observation reinforces a central point of the present work, which is that the computational problem of deprioritization requires an algorithmic solution that is different from compression or other types of inhibition. In other regard, all of these studies report a pattern of orthogonalized precue and parallel postcue item representations, suggesting similar mechanisms for item maintenance and selection.

In the DSR, the representational transformation of one item into a PMI and the other into a UMI are prompted by the same cue, a design feature that allows for direct comparison of the two processes. For the majority of participants in the EEG study, and in the majority of ROIs in the majority of participants in the fMRI study, trial-by-trial variation in the TVIs for the transformation to PMI and for the transformation to UMI were correlated, a result consistent with the idea that a common factor underlies both. There are at least two possible accounts for this pattern of results that will require future research to adjudicate. One is a parallel mechanism whereby a single signal is "split" so as to trigger the simultaneous output gating of one item into the PMI state and of the other item into the UMI state. A second is a serial process akin to biased competition (cf. Desimone & Duncan, 1995) whereby a control signal first selects the cued item, and a consequence of this item's transformation to PMI is that it "pushes" the other item into the UMI state. Importantly, the correlation of TVIs reported here rules out what had been a third possibility, which was a "passive" account of the transformation to UMI whereby the withdrawal of attention would allow the relaxation of the representation into a default state such that the relaxation process would not be influenced by the active PMI transformation. Along with the application of second-order context that is prompted by the prioritization cue, the time course of ED of the RNN suggests that the resetting of second-order context partway through the trial may be a process that requires as much active control as does its initial application.

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

APPENDIX

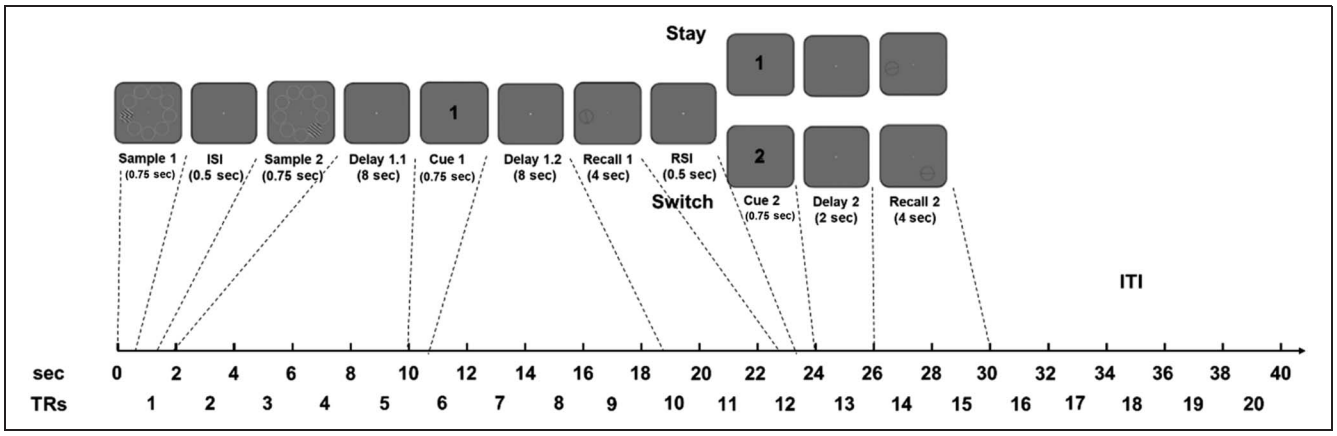

Figure A1. Timeline for the Yu and colleagues (2020) DSR task mapping task events to time in seconds and TRs.

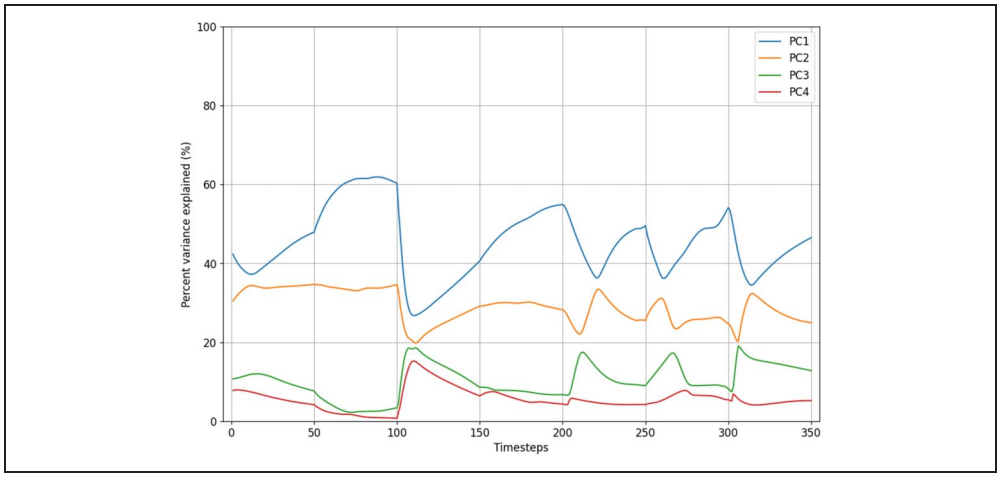

Figure A2. Time courses of percent variance explained by each of the first four PCs from the PCA performed on the recurrent unit activities for RNN No. 1.

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

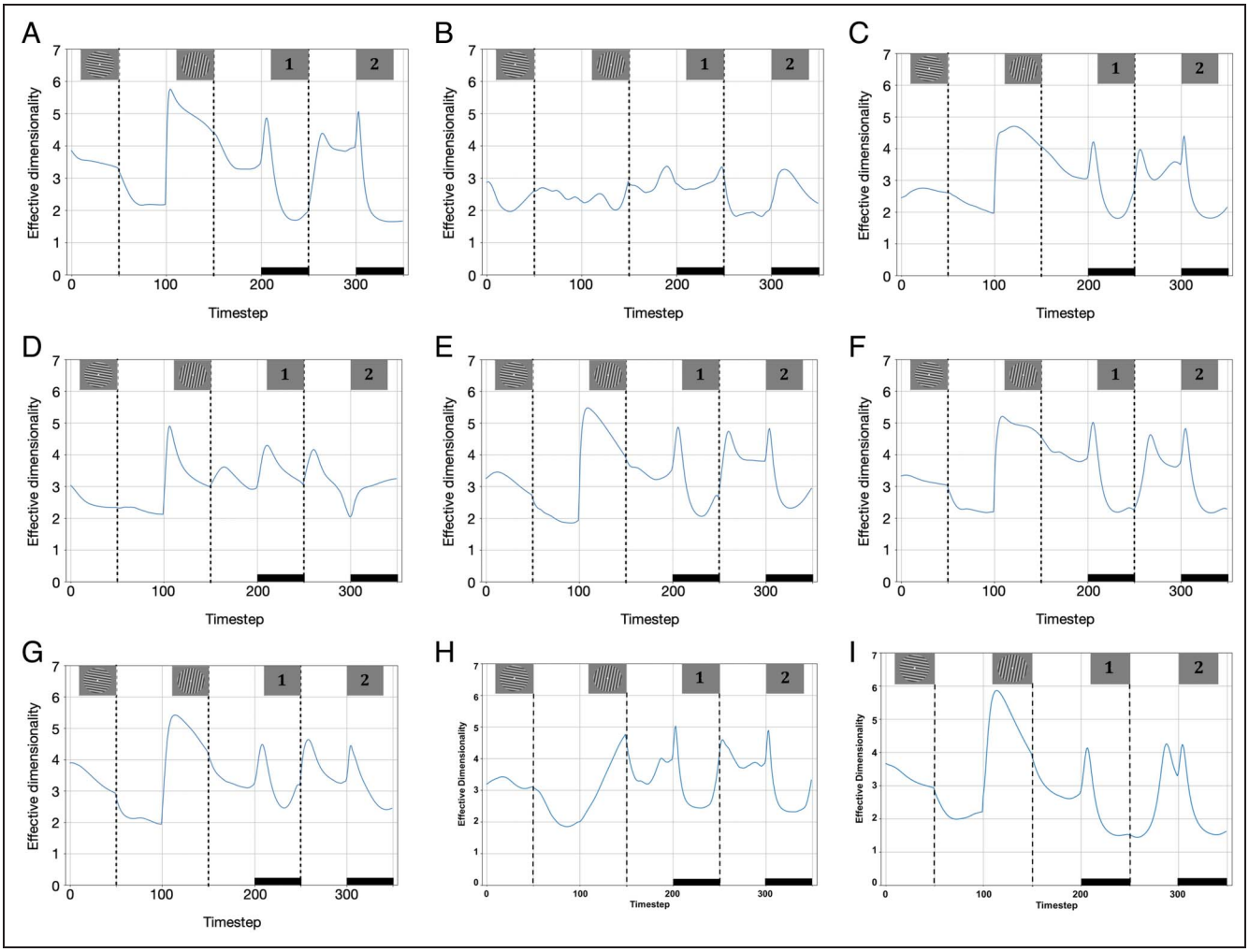

Figure A3. The time course of ED of the hidden layer stimulus representations of replication RNNs. (A) RNN No. 2. (B) RNN No. 3. (C) RNN No. 4. (D) RNN No. 5. (E) RNN No. 6. (F) RNN No. 7. (G) RNN No. 8. (H) RNN No. 9. (I) RNN No. 10.

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

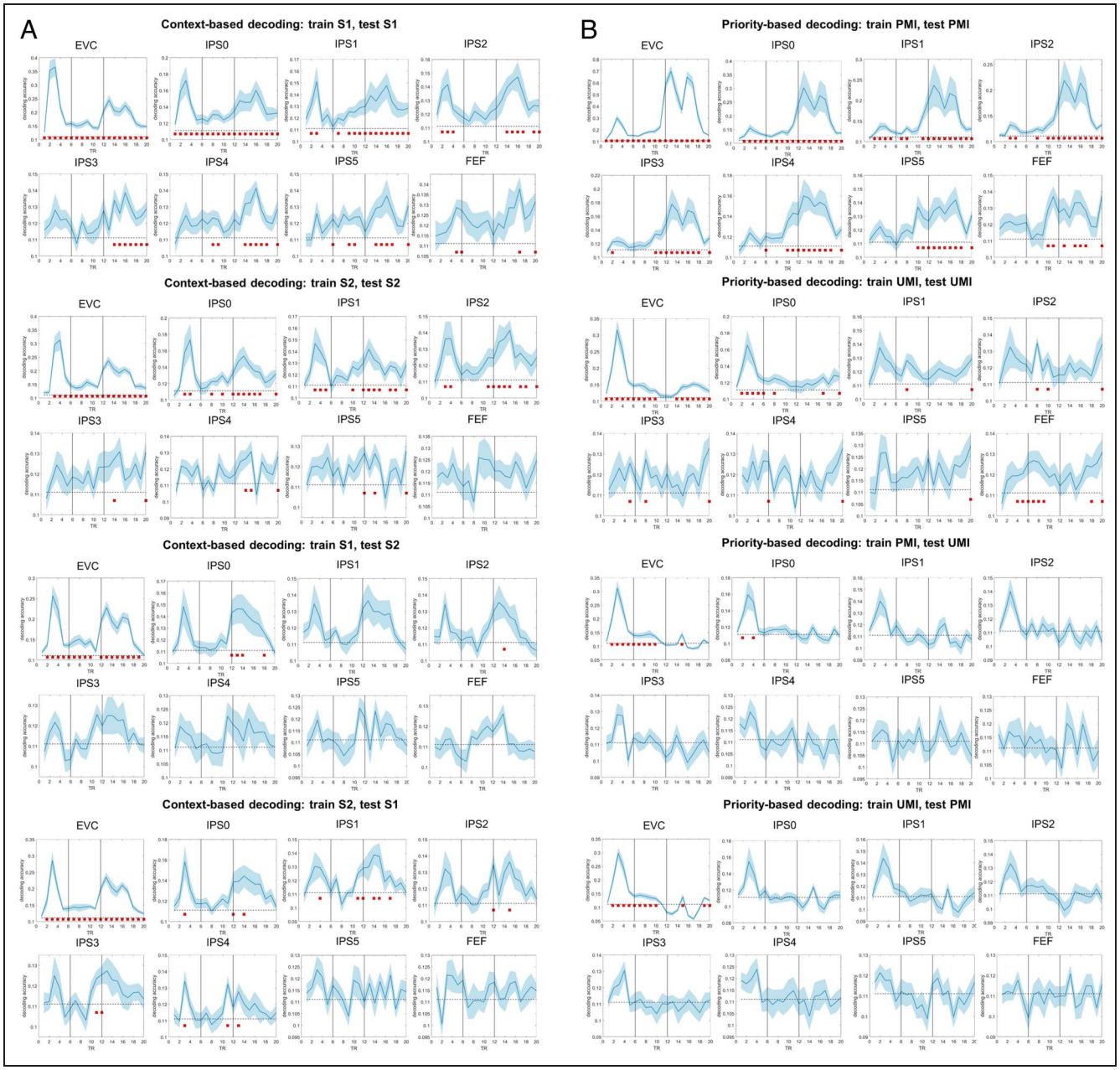

Figure A4. Within- and cross-label decoding of fMRI data separated by training and testing on each of the two items (i.e., Sample 1/2 for context and PMI/UMI for priority).

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

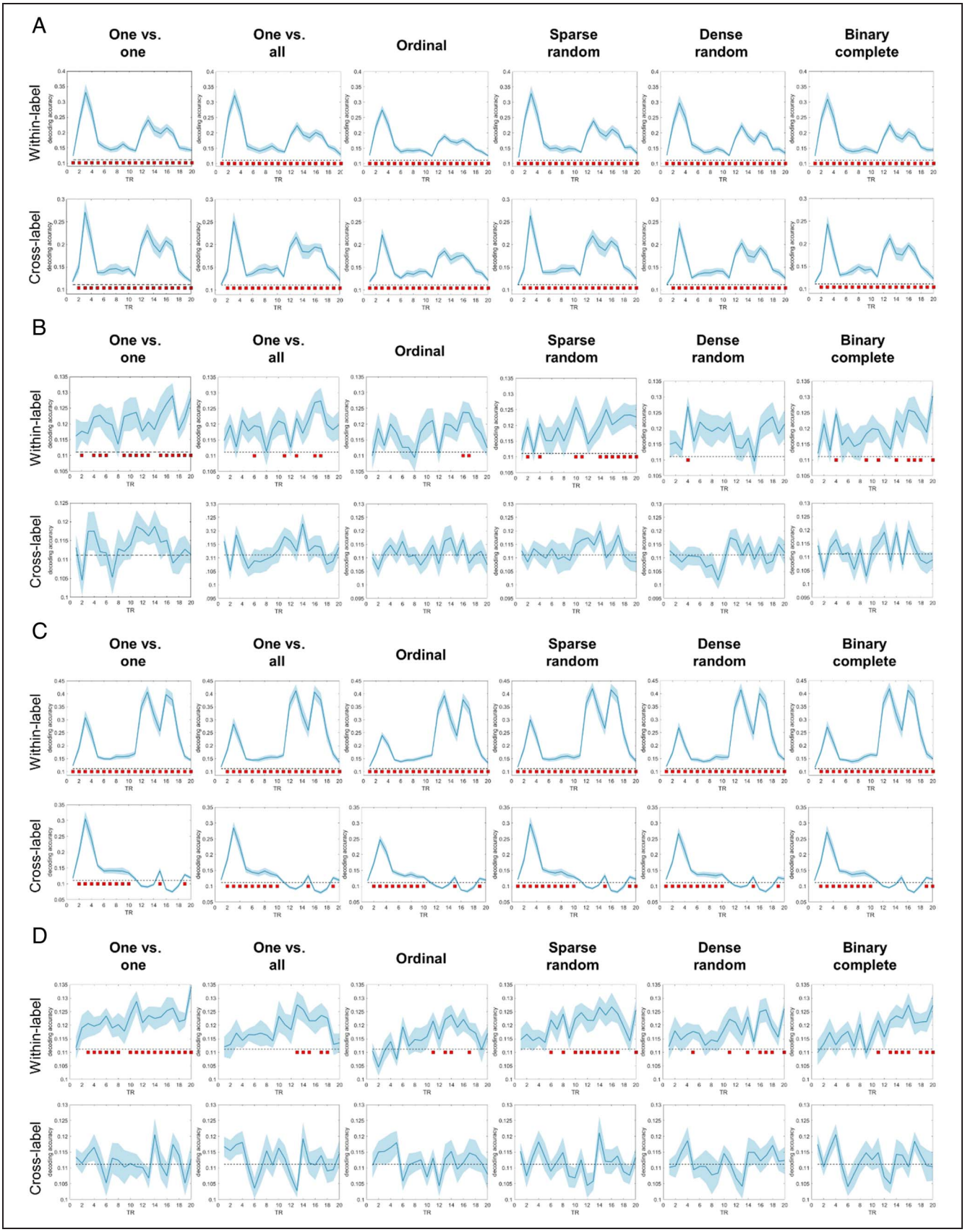

Figure A5. Comparisons of within- and cross-label decoding from the fMRI data set across various SVM coding designs in MATLAB. (A) Context-based decoding for EVC. (B) Context-based decoding for FEF. (C) Priority-based decoding for EVC. (D) Priority-based decoding for FEF. In each graph, the blue shading around each curve shows the standard error of the mean. The horizontal dashed line indicates the chance-level decoding accuracy of 0.11. Red squares below the dashed line indicate time points with significant above-chance decoding accuracy ($p < .05$, FDR-corrected across all time points).

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

Table A1. Additional TVI-behavior Correlation Analysis Results

		Context		*Priority*	
Region	*Subspace*	*Recall 1*	*Recall 2*	*Recall 1*	*Recall 2*
EVC	PMI/Sample 1	2	3	3	2
	UMI/Sample 2	1	2	1	2
IPS0–5	PMI/Sample 1	2	1	5	2
	UMI/Sample 2	5	1	0	1
FEF	PMI/Sample 1	3	1	1	0
	UMI/Sample 2	3	1	3	3

To evaluate whether trial-by-trial variation in TVI can predict recall error on the DSR task (Yu et al., 2020), we ran Spearman's rank correlations, separately for each participant, between the TVI (derived from Sample 1/2 subspaces for context, and PMI/UMI subspaces for priority) and recall error (for Recall 1 and 2) for all trials (one-tailed test as we predicted the correlations to be positive). Listed in the table are the number of participants (out of 13) that showed a significant positive TVI-recall error correlation ($p < .05$).

Acknowledgments

We thank Dr. Yuri Saalmann and Jiangang Shan for their critical feedback. Simulations were partially performed using the CloudLab computing facilities (Duplyakin et al., 2019).

Corresponding author: Quan Wan, Department of Psychology, University of Wisconsin–Madison, Madison, Wisconsin, or via e-mail: qwan22@wisc.edu.

Data Availability Statement

All processed data, code and trained network are available at https://osf.io/tnh9x/ on Open Science Framework.

Author Contributions

Quan Wan: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Visualization; Writing—Original draft. Adel Ardalan: Formal analysis; Investigation; Methodology; Software; Visualization; Writing—Review & editing. Jacqueline Fulvio: Data curation; Formal analysis; Investigation; Software; Visualization. Bradley R. Postle: Conceptualization; Funding acquisition; Investigation; Methodology; Supervision; Writing—Review & editing.

Funding Information

National Institutes of Health (https://dx.doi.org/10.13039/100000002), grant numbers: MH064498 and MH131678.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this paper report its proportions of citations by gender category to be: M/M = .615; W/M = .308; M/W = .077; W/W = 0.

REFERENCES

Cai, Y., Fulvio, J. M., Yu, Q., Sheldon, A. D., & Postle, B. R. (2020). The role of location-context binding in nonspatial visual working memory. *eNeuro*, *7*, ENEURO.0430-20.2020. https://doi.org/10.1523/ENEURO.0430-20.2020, PubMed: 33257529

Cueva, C. J., Ardalan, A., Tsodyks, M., & Qian, N. (2021). Recurrent neural network models for working memory of continuous variables: Activity manifolds, connectivity patterns, and dynamic codes. *arXiv*. https://doi.org/10.48550/arXiv.2111.01275

Del Giudice, M. (2021). Effective dimensionality: A tutorial. *Multivariate Behavioral Research*, *56*, 527–542. https://doi.org/10.1080/00273171.2020.1743631, PubMed: 32223436

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*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009, PubMed: 15102499

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205, PubMed: 7605061

Duplyakin, D., Ricci, R., Maricq, A., Wong, G., Duerig, J., Eide, E., et al. (2019). The design and operation of cloudlab. In *Proceedings of the USENIX Annual Technical Conference (ATC)* (pp. 1–14).

Fulvio, J. M., & Postle, B. R. (2020). Cognitive control, not time, determines the status of items in working memory. *Journal of Cognition*, *3*, 8. https://doi.org/10.5334/joc.98, PubMed: 32292872

Fulvio, J. M., Yu, Q., & Postle, B. R. (2023). Strategic control of location and ordinal context in visual working memory.

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024

Cerebral Cortex, *33*, 8821–8834. https://doi.org/10.1093/cercor/bhad164, PubMed: 37164767

Gosseries, O., Yu, Q., LaRocque, J. J., Starrett, M. J., Rose, N. S., Cowan, N., et al. (2018). Parietal-occipital interactions underlying control- and representation-related processes in working memory for nonspatial visual features. *Journal of Neuroscience*, *38*, 4357–4366. https://doi.org/10.1523/JNEUROSCI.2747-17.2018, PubMed: 29636395

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv*. arXiv:1412.6980 [Cs]. https://doi.org/10.48550/arXiv.1412.6980

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., et al. (2016). Demixed principal component analysis of neural population data. *eLife*, *5*, e10989. https://doi.org/10.7554/eLife.10989, PubMed: 27067378

Larocque, J. J., Lewis-Peacock, J. A., & Postle, B. R. (2014). Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in Human Neuroscience*, *8*, 5. https://doi.org/10.3389/fnhum.2014.00005, PubMed: 24478671

LaRocque, J. J., Riggall, A. C., Emrich, S. M., & Postle, B. R. (2017). Within-category decoding of information in different attentional states in short-term memory. *Cerebral Cortex*, *27*, 4881–4890. https://doi.org/10.1093/cercor/bhw283, PubMed: 27702811

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience*, *24*, 61–79. https://doi.org/10.1162/jocn_a_00140, PubMed: 21955164

Libby, A., & Buschman, T. J. (2021). Rotational dynamics reduce interference between sensory and memory representations. *Nature Neuroscience*, *24*, 715–726. https://doi.org/10.1038/s41593-021-00821-9, PubMed: 33821001

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*, 78–84. https://doi.org/10.1038/nature12742, PubMed: 24201281

Mutanen, T. P., Biabani, M., Sarvas, J., Ilmoniemi, R. J., & Rogasch, N. C. (2020). Source-based artifact-rejection techniques available in TESA, an open-source TMS-EEG toolbox. *Brain Stimulation*, *13*, 1349–1351. https://doi.org/10.1016/j.brs.2020.06.079, PubMed: 32659484

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*, 21–59. https://doi.org/10.1037/rev0000044, PubMed: 27869455

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*, e156869. https://doi.org/10.1155/2011/156869, PubMed: 21253357

Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, *592*, 601–605. https://doi.org/10.1038/s41586-021-03390-w, PubMed: 33790467

Piwek, E. P., Stokes, M. G., & Summerfield, C. (2023). A recurrent neural network model of prefrontal brain activity during a working memory task. *PLoS Computational Biology*, *19*, e1011555. https://doi.org/10.1371/journal.pcbi.1011555, PubMed: 37851670

Rogasch, N. C., Sullivan, C., Thomson, R. H., Rose, N. S., Bailey, N. W., Fitzgerald, P. B., et al. (2017). Analysing concurrent transcranial magnetic stimulation and electroencephalographic data: A review and introduction to the open-source TESA software. *Neuroimage*, *147*, 934–951. https://doi.org/10.1016/j.neuroimage.2016.10.031, PubMed: 27771347

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., et al. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, *354*, 1136–1139. https://doi.org/10.1126/science.aah7011, PubMed: 27934762

Rumelhart, D. E., Hinton, G. E., & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533–536. https://doi.org/10.1038/323533a0

Saxe, A. M., McClelland, J. L., & Ganguli, S. (2013). Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *arXiv*. https://doi.org/10.48550/arXiv.1312.6120

Teich, A. F., & Qian, N. (2003). Learning and adaptation in a recurrent model of V1 orientation selectivity. *Journal of Neurophysiology*, *89*, 2086–2100. https://doi.org/10.1152/jn.00970.2002, PubMed: 12611961

Teng, C., & Postle, B. R. (2024). Investigating the roles of the visual and parietal cortex in representing content versus context in visual working memory. *eNeuro*, *11*, ENEURO.0270-20.2024. https://doi.org/10.1523/ENEURO.0270-20.2024, PubMed: 38336475

van Loon, A. M., Olmos-Solis, K., Fahrenfort, J. J., & Olivers, C. N. (2018). Current and future goals are represented in opposite patterns in object-selective cortex. *eLife*, *7*, e38677. https://doi.org/10.7554/eLife.38677, PubMed: 30394873

Wan, Q., Cai, Y., Samaha, J., & Postle, B. R. (2020). Tracking stimulus representation across a 2-back visual working memory task. *Royal Society Open Science*, *7*, 190228. https://doi.org/10.1098/rsos.190228, PubMed: 32968489

Wan, Q., Menendez, J. A., & Postle, B. R. (2022). Priority-based transformations of stimulus representation in visual working memory. *PLoS Computational Biology*, *18*, e1009062. https://doi.org/10.1371/journal.pcbi.1009062, PubMed: 35653404

Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, *25*, 3911–3931. https://doi.org/10.1093/cercor/bhu277, PubMed: 25452571

Yu, Q., Teng, C., & Postle, B. R. (2020). Different states of priority recruit different neural representations in visual working memory. *PLoS Biology*, *18*, e3000769. https://doi.org/10.1371/journal.pbio.3000769, PubMed: 32598358

Downloaded from http://direct.mit.edu/jocn/article-pdf/36/7/1374/2384144/jocn_a_02166.pdf by University of Wisconsin, Madison user on 23 July 2024