Rotational remapping between differently prioritized representations in visual working memory

Quan Wan[1*], Jorge A. Menendez[2], Bradley R. Postle[1,3]

[1]Department of Psychology, University of Wisconsin–Madison, Madison, Wisconsin, United States of America

[2]Gatsby Computational Neuroscience Unit, University College London, London, United Kingdom

[3]Department of Psychiatry, University of Wisconsin–Madison, Madison, Wisconsin, United States of America

[*]Corresponding author:

E-mail: qwan22@wisc.edu

Abstract

How does the brain prioritize among the contents of working memory to appropriately guide behavior? Using inverted encoding modeling (IEM), previous work (Wan et al., 2020) showed that unprioritized memory items (UMI) are actively represented in the brain but in a "flipped", or opposite, format compared to prioritized memory items (PMI). To gain insight into the mechanisms underlying the UMI-to-PMI representational transformation, we trained recurrent neural networks (RNNs) with an LSTM architecture to perform a 2-back working memory task. Visualization of the LSTM hidden layer activity using Principle Component Analysis (PCA) revealed that the UMI representation is rotationally remapped to that of PMI, and this was quantified and confirmed via demixed PCA. The application of the same analyses to the EEG dataset of Wan et al. (2020) revealed similar rotational remapping between the UMI and PMI representations. These results identify rotational remapping as a candidate neural computation employed in the dynamic prioritization within contents of working memory.

Author Summary

Our flexible behavior relies on the successful prioritization among the many pieces of information that may be held in working memory (WM). However, the computations underlying prioritization remain poorly understood. The current work investigated computational bases of prioritization by training recurrent neural networks (RNNs) to perform a 2-back WM task. Dimensionality reduction

techniques revealed that a remembered item undergoes a rotational transformation in multi-dimensional representational space as its priority status changes. Prompted by this finding, we next asked if a similar mechanism might be applied by the human brain. Applying analogous analyses to an empirical EEG dataset of 2-back WM performance, we found a similar pattern of priority-based rotational remapping of working memory representations. This work highlights the power of artificial neural networks in generating mechanistic hypotheses for empirical work in understanding brain and cognition.

Introduction

The ability to flexibly select and prioritize among information held in working memory (WM) is critical for guiding behavior and thought. For this reason, the neural mechanisms of attentional prioritization in WM have been extensively studied in recent years. Many studies of WM for visual material have reported that the prioritization of one item held in WM leads to a decrease in the activity level of the "unprioritized memory item" (UMI; Myers et al., 2017), sometimes to baseline levels (LaRocque et al., 2012; Lewis-Peacock et al., 2011; Rose et al., 2016). Some have interpreted these results as consistent with the idea that whereas prioritized memory items (PMI) are held in an active state, UMIs may be maintained as "activity-silent" traces encoded in synaptic weights (Barak & Tsodyks, 2014; Stokes, 2015). (Note that these two modes of representation need not be mutually exclusive, with recent computational accounts assuming that PMIs are encoded in

both elevated activity and activity-silent patterns of synaptic weights (Manohar et al., 2019; Masse et al., 2019; see review by Stokes et al., 2020).) Although much of this literature on attentional prioritization in WM has emphasized the loss of an active trace of the UMI, in this report we address a fundamentally different consequence of selection in WM: the representational transformation of the UMI.

Experimental tasks used to study prioritization in WM necessarily include multiple steps, such that the information not needed for the impending response (i.e., the UMI) might be needed to guide a subsequent response. This is often done with retrocues. For example, van Loon and colleagues (2018) acquired functional magnetic resonance imaging (fMRI) data while first presenting subjects two target images sequentially (e.g., first a flower then a cow), then indicating with a cue whether memory for the first or second presented image would be tested first. Had the cue been a "1", subjects would next see a test array of six flowers and indicate whether the target flower appeared in the test array, and finally a test array of six cows. On this trial, the target cow spent time as UMI, because the cue indicated that memory for the flower would be tested first. When van Loon et al. (2018) applied multivariate pattern analysis (MVPA) to fMRI data from posterior ventral temporal lobe, they found that a decoder trained on trials when an item was a PMI performed statistically below chance when that item was a UMI. Furthermore, a representational dissimilarity analysis indicated that, within their set of 12 stimuli (four cows, four skates, four dressers), each item's high-dimensional representation in one state (e.g., as a PMI) was maximally different from its representation in the other state (i.e., as a UMI). Using a similar retrocuing procedure, Yu, Teng and Postle

(2020) found, with multivariate inverted encoding modeling (IEM) of fMRI data from early visual cortex, that the reconstructed orientation of a grating "flipped" when it was a UMI relative to a PMI (e.g., a 30º orientation reconstructed as 120º while a UMI). Furthermore, for data from the intraparietal sulcus (IPS), they observed that the IEM reconstruction of the location where an item had been presented also flipped when an item's priority status transitioned to UMI. (Notably, these "flipped" (or "opposite", van Loon et al., 2018) states are not less *active* relative to the PMI representations, rather, they are *different*.)

Shifts of priority are also characteristic of continuous-performance tasks, for which shifts of priority are dictated by task rules rather than by explicit cues. One example, which provided the impetus for the work presented here, is the 2-back WM task from Wan and colleagues (2020; Figure 1). Electroencephalography (EEG) signals were recorded while subjects viewed the serial presentation of oriented gratings and judged for each one whether it was a match or a non-match to the item that had appeared two positions previously in the series. This task entails a predictable transition through priority states for each item: When an item n is initially presented, it serves as probe to compare against the memory of item $n - 2$; after the n-to-$n - 2$ decision is made, item n becomes a UMI while item $n - 1$ is prioritized for the upcoming comparison with $n + 1$. Next, once the $n + 1$-to-$n - 1$ comparison is completed, item n becomes a PMI for its impending comparison with item $n + 2$. To analyze the EEG data, we trained an IEM on the raw EEG voltages from a separate 1-item delayed-recognition task, and tested it on the delay periods separating n and $n + 1$ and separating $n + 1$ and $n + 2$ (i.e., when item n assumed the

status of UMI, then PMI). The results, reminiscent of van Loon et al. (2018) and Yu, Teng and Postle (2020), indicated that the IEM reconstruction of the UMI was "flipped" relative to the training data, then "flipped back" when its status transitioned to PMI (Figure 2). We referred to the transition from PMI to UMI, and back, as "priority-based remapping" (rather than "recoding" or "code morphing"; c.f. Parthasarathy et al., 2017), reasoning that the IEM reconstruction of the UMI would fail if it were represented in a neural code different from the trained model. To gain mechanistic insight into this phenomenon, formal modeling is needed.

Two computational models offer some insight into priority-based remapping. One model, by Lorenc and colleagues (2020), was designed to account for a similar flipped IEM reconstruction observed in an fMRI study using a retrocuing task. This approach was inspired by evidence from monkeys performing WM tasks, in which top-down signals from FEF were shown to alter several receptive field properties of neurons in extrastriate visual areas V4 and MT (Merrikhi et al., 2017). They created simulated data for training IEMs using the basis set that was employed for IEM reconstructions on empirical data, and subsequently created a test dataset where the basis function parameters for memory strength (φ), gain (γ), receptive field width (μ), and receptive field centers (δ) were varied. These parameters were then fitted to experimental data. Although this model reasonably reproduced the flipping of IEM reconstructions, its mechanistic interpretation was equivocal because multiple solutions fit the data similarly well (i.e., width modulation versus memory strength + gain modulation). Additionally, its implications for the PMI-to-UMI transition are unclear, because the simulated "flipped" IEM reconstruction was

obtained from the time period when the stimulus was no longer required for the task. A second model, from Manohar and colleagues (2019), simulated working memory performance in a network comprised of hard-coded feature-selective units and a pool of freely conjunctive units that can form a plastic attractor to keep one item, a PMI, in a state of elevated activity. When attention shifted away from an item (making it a UMI), it remained briefly encoded in a residual pattern of strengthened connections, and, under some conditions, inhibition from activity in other parts of the network produced an "inverted" representation of UMI. Although this model successfully reproduced other empirical findings using simulated data, such as the temporary reactivation of the UMI by a nonspecific pulse of excitation, it was not used to account for empirical neural data. From the perspective of the framework of Marr and Poggio (1976), both of the models reviewed above were intended to address the phenomenon of opposite representations between UMI and PMI at the implementational level. Our interest in this report is at the algorithmic level of analysis: What are the neural computations that allow UMI and PMI representations to differentially drive behavior?

To answer this question, we turned to artificial neural networks (ANNs), which have been playing an increasingly prominent role in providing mechanistic insights into, and generating novel hypotheses of, phenomena in cognition and neuroscience (Kell & McDermott, 2019; Mante et al., 2013; Richards et al., 2019; Sussillo et al., 2015; Yang et al., 2019). In the current work, we use recurrent neural networks (RNNs) with a long short-term memory (LSTM) architecture (Hochreiter & Schmidhuber, 1997) to perform a 2-back WM task modeled on Wan et al., (2020).

LSTMs can generate flexible behavior guided by long range temporal dependencies, and can solve complex tasks such as speech recognition (Graves et al., 2013) and machine translation (Sutskever et al., 2014). Moreover, LSTM might be a good model for WM tasks due to its gating-based architecture, reminiscent of the cortico-striatal mechanisms believed to gate information into and out of WM (Chatham & Badre, 2015; O'Reilly & Frank, 2006).

To summarize our procedure and findings, first we used Principal Component Analysis (PCA) of activity in the RNN hidden layer to visualize its representational dynamics. This revealed a rotational transformation between the UMI and PMI representations. To characterize and quantify this rotational behavior, we projected the population activity onto stimulus-relevant dimensions estimated by demixed Principal Component Analysis (dPCA; Kobak et al., 2016) and fit a rotational transformation to the UMI-to-PMI transition within this low-dimensional subspace. This allowed us to precisely measure rotational structure in the stimulus representations of the RNN, confirming that the UMI and PMI representations were rotated versions of each other. Having observed this with the LSTM model, we then applied the same set of analyses to the EEG data from Wan et al. (2020), which revealed similar rotational representational remapping. The results suggest that the human brain may prioritize items in memory using similar computations to those of LSTMs trained to solve priority-based WM tasks.

Methods

Behavioral task

In each experimental block of the 2-back working memory task, both human subjects (N = 42) and RNNs (N = 10) were serially presented a sequence of stimuli drawn from a closed set of six different identities (128-stimulus blocks for humans, 20-stimulus blocks for RNNs). The task was to indicate, for each stimulus, whether or not it matched the identity of the stimulus that had been presented 2 positions earlier in the series. Each EEG subject performed 4 blocks and each RNN performed 200 blocks.

Recurrent neural network (RNN) model

RNN architecture

Ten RNNs with an LSTM architecture were trained and simulated using the Python-based machine learning package PyTorch. All networks consisted of 6 "orientation"-selective input neurons and 7 LSTM hidden units, which were linearly rectified and linearly read out to a single output neuron. We employed the linear rectification to emulate the nonlinear relationship between subjects' memory representations (as measured by EEG) and their decision outputs. Networks with other numbers of hidden units gave qualitatively similar results.

Stimuli

The identity of each stimulus presented to the network was denoted by an integer randomly generated between 1 and 6. The stimulus input took the form of a one-hot vector, with only the unit corresponding to the stimulus identity activated (e.g., [0, 0, 1, 0, 0, 0] for stimulus #3; we also explored RNNs trained on metrically

varying input vectors following the basis function used to build IEMs in Wan et al. (2020), and these yielded similar results, see Supplementary Materials S1). To simulate the delay period in the human task, we installed 2 "delay" timesteps following the presentation of each stimulus (with an input of [0, 0, 0, 0, 0, 0]; no delay timesteps after the last stimulus in the sequence). A "stimulus event" consisted of the presentation of stimulus *n* and its following two delay timesteps. To evaluate the UMI-to-PMI representational transition of stimulus *n*, we refer to the concatenation of each two consecutive "stimulus events" as a "trial". The target output was 0 during each delay timestep and 1 at the stimulus presentation timestep if the presented stimulus matched the stimulus presented two stimuli before ("match" response); otherwise the target output at stimulus presentation was also 0 ("non-match" response). Each block comprised 18 trials (because no delay period followed stimulus #20; the last trial contained stimulus #18 and #19), and only 16 trials were analyzed (because the first two stimulus events had no target outputs: not enough stimuli preceded them to have a match/non-match decision). We generated 200 random stimulus sequences for training the RNNs and 200 random sequences for testing the trained networks. Because the human 2-back task had a ratio of 1:2 between match and non-match trials, we generated random sequences that satisfied the criterion that each sequence had to contain at least 5 match trials. The outcome was that training sequences had an average of 5.55 match trials (*SD* = 0.78) and testing sequences an average of 5.46 match trials (*SD* = 0.70).

RNN training and testing

Unit activity of the RNNs was initialized with 0, and weights and biases were initialized with random values. The RNNs were trained using the Adam stochastic gradient descent (SGD) algorithm for 5000 iterations (Kingma & Ba, 2017; learning rate = 10^{-3}). In each iteration, a batch of 20 sequences was randomly selected (with replacement) from the 200 training sequences. The loss function minimized was the mean squared error between output activity and target output across all timesteps and sequences.

After 5000 iterations of training, RNNs were tested on an independently sampled set of 200 stimulus sequences to assess generalization. The network's performance accuracy was calculated as the percentage of trials (across all 200 sequences in the test set) on which the network made a correct response, where a response was deemed correct if the absolute difference between the activation of the output neuron and the target output was smaller than 0.5. Only networks with a performance accuracy of 99.5% or above were kept until 10 such networks were obtained. As a result, a total of 12 networks were trained altogether (i.e., 2 networks discarded). The activity timeseries of the LSTM hidden layer units from all 3200 trials (16 trials x 200 sequences) in the training data set were extracted for subsequent analyses.

PCA visualization of the LSTM layer activity

We extracted from each network the activity of the 7 hidden units in the LSTM layer from all 200 training sequences and used Principal Components Analysis (PCA) to project these 7-dimensional activity patterns onto the top two

dimensions accounting for the most variance across all training sequences and timesteps. We then visualized each stimulus *n*'s transition from UMI to PMI within this subspace by plotting the dimensionality-reduced activity across a time course with 9 timesteps. These 9 timesteps comprised the presentations of stimulus *n, n + 1, n + 2* and their subsequent delay periods (*delay 1:1-2, delay 2:1-2, delay 3:1*-2; Figure 3A). Namely, we added an extra stimulus event following the stimulus *n* trial. We only plotted trials with no overlapping stimulus events (e.g., the *n*-plus-*n + 1* trial and the *n + 2*-plus-*n + 3* trial were plotted but not the *n + 1*-plus-*n + 2* trial; consequently, a total of 1600 trials were plotted). To see how the representation of stimulus *n* evolves as it transitions from being a UMI to a PMI, we colored the activity patterns according to the identity of stimulus *n*. As explained in the Introduction, the memory of stimulus *n* is therefore a UMI during the delay period after the presentation of stimulus *n* (*delay 1:1-2*; because it is not needed for the upcoming *n – 1*-to- *n + 1* comparison) and becomes a PMI during the delay period after the presentation of stimulus *n + 1* (*delay 2:1-2*; to prepare for the comparison with *n + 2* next). We focused on the *delay 1:2* and *delay 2:2* timesteps to characterize the UMI-to-PMI representational transformation. To visualize the representation of decision, we re-plotted the same activity patterns but colored them according to the correct response to the *n*-to-*n + 2* comparison when *n + 2* was presented (Figure 3B).

EEG dataset

The experimental protocol for the Wan et al. (2020) EEG study (the data from which was analyzed in this paper), along with the informed consent form, was approved by the University of Wisconsin–Madison Health Institutional Review Board (protocol no. 2016-0500). Prior to each experimental session, informed consent was obtained by lab personnel listed on the IRB-approved protocol.

60-channel EEG data were acquired and preprocessed as per procedures described in Wan et al. (2020). Raw EEG voltages were used for all analyses. Because data from the pilot and replication experiments from Wan et al. (2020) yielded very similar IEM reconstruction results, they were combined to yield a dataset of 42 subjects. As with the RNN data, after excluding the first two stimuli from each block there were 126 stimulus events and hence 125 trials per block (Figure 2). Each stimulus event (stimulus presentation followed by a delay) lasted 3550 ms. A third of the trials in each block were 'match' trials and the other two thirds were 'non-match' trials. EEG data from all trials (both correct and incorrect) were included in the analyses. To characterize stimulus representations in WM, we focused on activity during the delay period succeeding each stimulus (specifically, 1150-3150 ms from stimulus onset). Thus, for stimulus n, during this delay period after its onset, stimulus $n - 1$ had the status of PMI and n had the status of UMI.

WM-specific dimensionality reduction via dPCA

Demixed principal component analysis (Kobak et al., 2016) was employed to identify dimensions of RNN and EEG activity relevant to the stimulus representation in WM. Unlike PCA, which seeks dimensions that maximize the total variance of the

data regardless of task variables, dPCA allows one to parse out dimensions of variability specific to certain task variables (e.g., stimulus identity, decision). Given a task variable of interest, the dPCA algorithm does this by grouping neural activity patterns according to this variable and identifying dimensions that maximize between-group variance and minimize within-group variance. Here, we used this method to identify dimensions of activity that were strongly modulated by the identity of the UMI or PMI during the delay period.

To obtain dimensions of activity strongly modulated by the UMI stimulus identity, we performed the following optimization:

$$V^{UMI}, W^{UMI} = \arg\min_{V,W} \mathbb{E}_{s,t}\left[\left\|(\bar{x}^s - \bar{x}) - VW^T\,(x_t^s - \bar{x})\right\|^2\right]$$

where $x_t^s \in \mathbb{R}^{60}$ is the EEG response at time t within the delay period averaged over all trials in which stimulus s was the UMI (trial averaging was necessary to average away noise), $\bar{x}^s$ is its temporal mean over the delay period, and $\bar{x}$ is the global mean over all trials and delay period timepoints. The $60 \times D$ matrix V^{UMI} is the so-called "encoding" weight matrix and the $D \times 60$ matrix W^{UMI} is the "decoding" weight matrix, used for dimensionality reduction. The rows of W^{UMI} are termed demixed Principal Components (dPC). This optimization problem is called a reduced-rank regression problem, and admits a unique closed-form solution (Kobak et al., 2016). Because there are only 6 different UMI stimuli (and thus 6 different $\bar{x}^s$ vectors), only up to D=5 dPC's can be computed in this way (since the ordinary least-squares solution has rank 5). As in PCA, these dPC's can be ordered in terms of the amount of variance they explain in the data. Because the task variable used for dPCA was the

UMI stimulus, we call these dPC's the UMI dPC's, and call the subspace spanned by them the UMI subspace. We also extracted PMI dPC's, W^{PMI}, and a PMI subspace by exactly repeating the above operation but with the index s now indexing the PMI stimulus identity. Dimensionality-reduced representations were then obtained by reading out activity through the dPC's, Wx_t^s.

In the main text we used only D=2 dPC's for dimensionality reduction, as this permitted straight-forward visualization of the dimensionality-reduced representations and allowed us to characterize the rotational transformation with a single angle. We also explored different numbers of dPCs to construct the UMI- and PMI-dependent subspaces, which yielded similar results (see Supplementary Materials S2).

Percent variance explained calculations were performed as follows. Percent global variance explained by the ith dPC, w_i (i.e. the ith row of the decoder matrix W), was calculated using the corresponding column v_i from the encoder matrix V by

$$1 - \frac{\mathbb{E}_{s,t}\left[\left\|(x_t^s - \bar{x}) - v_i w_i^T (x_t^s - \bar{x})\right\|^2\right]}{\mathbb{E}_{s,t}\left[\left\|(x_t^s - \bar{x})\right\|^2\right]}$$

The percent stimulus variance explained was defined as

$$1 - \frac{\mathbb{E}_{s,t}\left[\left\|(\bar{x}^s - \bar{x}) - v_i w_i^T (\bar{x}^s - \bar{x})\right\|^2\right]}{\mathbb{E}_{s,t}\left[\left\|(\bar{x}^s - \bar{x})\right\|^2\right]}$$

dPCA rotational transformation analysis (RNN and EEG)

Previous IEM findings (Figure 2) suggested that the UMI- and PMI-dependent subspaces might be overlapping. Therefore, to fully characterize the UMI-to-PMI rotational transformation in stimulus-relevant dimensions, we conducted all analyses in both subspaces. We anticipated that UMI and PMI subspaces would yield similar results.

To quantify the rotational transformations between the UMI and PMI representations in both RNN and EEG data, we projected the trial- and time-averaged UMI and PMI delay period activity onto the given subspace,

$$y_{UMI}^{s} = W\bar{x}_{UMI}^{s}, \qquad y_{PMI}^{s} = W\bar{x}_{PMI}^{s}$$

where $W = W^{UMI}$ or W^{PMI} is the $D \times 60$ decoding matrix containing the top D dPC's, and $\bar{x}_{UMI}^{s}$ and $\bar{x}_{PMI}^{s}$ are the mean delay period EEG activity averaged over trials in which the UMI or PMI stimulus was s, respectively (corresponding to the vector $\bar{x}^{s}$ above with the index s denoting the UMI or PMI stimulus identity, respectively). We henceforth refer to their projections onto the stimulus-specific dPC subspace, y_{UMI}^{s}, y_{PMI}^{s}, as the UMI and PMI means, respectively. We then fit a rotation matrix to the transformation between the UMI and PMI means,

$$\hat{R} = \arg\min_{R} \sum_{s=1}^{6} \left\| y_{PMI}^{s} - R y_{UMI}^{s} \right\|^{2}$$
$$\text{subject to } R^{T}R = I, |R| = 1$$

where $|R|$ denotes the determinant of the matrix R, constrained to be positive to ensure the matrix is a proper rotation. This is called a Procrustes problem, and has a closed-form solution. Once R was computed, we calculated the rotation angle of this transformation and the squared reconstruction error between the PMI and the optimally rotated UMI means. We then normalized this by the squared difference between the PMI and UMI means to obtain a *rotation index* (RI) which measured the extent to which the UMI-to-PMI transformation was rotation-like,

$$\mathrm{RI} = \frac{\sum_{s=1}^{6} \left\| y_{PMI}^{s} - \hat{R} y_{UMI}^{s} \right\|^2}{\sum_{s=1}^{6} \left\| y_{PMI}^{s} - y_{UMI}^{s} \right\|^2}$$

The rotation index ranges from 0 and 1, where a smaller value indicates the UMI-PMI representational transformation to be more rotational (i.e. a rotation matrix does a good job of fitting the empirical transformation).

To confirm that the UMI-PMI transformation was more rotation-like than would be expected by chance, we constructed a null distribution of RIs by exhaustively permuting the labels of the 6 PMI means (after excluding the observed PMI configuration: 6! – 1 = 719 iterations) and repeating the RI calculation for each permutation. The p-value was calculated as the proportion of permutations with an RI smaller than the observed UMI-PMI transformation. This analysis was done for each individual subject. A $p < .05$ would indicate that the UMI-PMI transformation could be characterized as a pure rotation with >95% confidence.

Note that it was critical to perform this analysis within the stimulus-specific subspace identified through dPCA. When performing this analysis with PCA instead

of dPCA (by constructing the dimensionality-reduced UMI and PMI means y_{UMI}, y_{PMI} by projecting onto top 2 PC's instead of dPC's), the rotational structure was not observed (Supplementary Materials S3).

Results

Behavioral results of EEG study

Mean accuracy was 86.1% (*SD* = 5.6%), mean *d'* was 2.40 (*SD* = 0.65), and mean response time was 0.82 s (*SD* = 0.18 s).

PCA of LSTM activity patterns

As shown in Figure 3A, upon the presentation of stimulus *n*, the stimulus representations from all trials formed stimulus-specific clusters in the 2D PCA space (*delay 1:1* and *delay 1:2*), revealing a stimulus-dependent axis. During the presentation of stimulus *n + 2* (Figure 3B), when *n* is compared with *n + 2*, match trials congregated in a "band" in the center of PC space, whereas non-match trials clustered in the flanking "clouds", thus revealing a decision-based representational structure. Over the course of a trial, *n*'s stimulus-specific axis appeared to rotate counterclockwise (in the PCA plane) as it transitioned from UMI to PMI (*delay 1:1-2* to *delay 2:1-2*). Visual inspection of this sequence suggested a rotational trajectory of the stimulus-specific axis, beginning with alignment with the decision-based structure (as revealed at timestep *n + 2*) during the *n – 2* versus *n* comparison (timestep *n*), then a rotational trajectory into an orientation "perpendicular" to the

decision structure for the *n + 1* timestep, then further rotation into a new configuration that was again aligned with the decision structure for the *n + 2* timestep (bottom-left panel in Figures 3A and 3B). Note that the configuration of individual values is necessarily different for timestep *n + 2* than for timestep *n,* because item *n* serves different functions at these two timesteps – *probe* at timestep *n* and *memorandum* at timestep *n + 2.* We speculate that the function of this rotational transformation might be to prevent the remembered representation of *n* from influencing the *n – 1* versus *n + 1* decision (at timestep *n + 1*).

dPCA rotation transformation analyses (RNN and EEG)

RNN

To quantify the transformations observed in the PCA (Figure 3A), we applied dPCA to identify the top two UMI-selective dPCs and the top two PMI-selective dPCs, then projected trial-averaged activity during the second delay timesteps of the trial (*delay 1:2* for the UMI and *delay 2:2* for the PMI) onto these dPC subspaces. (See Figure 4A for the 2D projections of UMI and PMI means for an example network, and Supplementary Materials S4 for the analogous projections of all trials). The first and second dPCs of the UMI subspace accounted for 95.1% and 4.1% of the total stimulus variance of the trial-averaged data, respectively. The first and second dPCs of the PMI subspace accounted for 97.4% and 2.4% of the total stimulus variance, respectively (see Supplementary Materials S5 for additional information). The rotational index (RI) between the rotated UMI and the target PMI was 0.12 (*SD* = 0.08) in the UMI subspace and 0.38 (*SD* = 0.11) in the PMI subspace (Figure 5A).

Permutation testing indicated significant rotational behavior in the UMI subspaces and PMI subspaces of all 10 networks (*p*s < .05; Figure 6A). The angle obtained from the best-fitting rotation matrix, averaged across networks, was 146.76° (*SD* = 15.59°) in the UMI subspace and 137.85° (*SD* = 15.42°) in the PMI subspace (Figure 7A).

EEG

Having found evidence for rotational transformations between UMI and PMI in the RNNs, we applied comparable rotation analyses on the EEG data to assess whether the brain might carry out similar rotational transformations while performing the 2-back task. Figure 4B illustrates dPCA projections of the UMI and PMI for an example subject. The first and second dPCs of the UMI subspace accounted for 47.4% and 23.4% of the total stimulus variance of the trial-averaged data, respectively. The first and second dPCs of the PMI subspace accounted for 47.9% and 24.1% of the total stimulus variance, respectively (see Supplementary Materials S5 for additional information). The RI between the rotated UMI and the target PMI was 0.23 (*SD* = 0.14) in the UMI subspace (significant subjects: 0.16 ± 0.08; non-significant subjects: 0.41 ± 0.13; Figure 5B) and 0.25 (*SD* = 0.17) in the PMI subspace (significant subjects: 0.19 ± 0.10; non-significant subjects: 0.45 ± 0.17). Permutation testing indicated that the UMI-to-PMI transformation was significantly rotational in the UMI subspace of 31 subjects and in the PMI subspace of 32 subjects (27 subjects showed significant rotations in both UMI and PMI subspaces; *p*s < .05, Figure 6B). Across all 42 subjects, the average rotation angle

was 164.62° (*SD* = 18.63°) in the UMI subspace (significant subjects: 170.59° ± 7.08°; non-significant subjects: 147.79° ± 28.29°; Figure 7B) and 161.28° (*SD* = 18.01°) in the PMI subspace (significant subjects: 166.41° ± 11.05°; non-significant subjects: 144.85° ± 24.83°).

Discussion

Results from previous neuroimaging studies have given rise to the idea that representations in working memory undergo a "priority-based remapping" when they obtain the status of UMI (van Loon et al., 2018; Wan et al., 2020; Yu, Teng & Postle, 2020), but the mechanism underlying this transformation was unknown. Here, using neural network modeling and dimensionality reduction techniques, we have identified rotation through representational space as a candidate computation underlying this phenomenon. After confirming that our findings were robust in RNNs trained to perform the 2-back working memory task, we applied similar analyses to a sample of human EEG data, and determined that rotation accounted for representational transformations in the majority of these subjects as well.

Our rotational analyses yielded similar results for RNN and EEG data, indicating shared representational mechanisms (Figure 4-7). As expected, comparable rotations were characterized in both UMI and PMI subspaces, confirming that these two subspaces overlap. Strikingly, the rotational angles from the EEG data tightly clustered around 180° across subjects in both subspaces. Note that such a rotation by 180° is consistent with the "flipping" of the IEM reconstruction of UMI relative to PMI representations reported by Wan et al. (2020).

At the present time, we don't have a basis for conjecture as to whether the difference and direction of difference of the angles of rotation in the RNN data (UMI subspace: 146.76°, PMI subspace: 137.85°) is meaningful.

In this report we have considered RNN simulations, which reveal internal dynamics of an abstract artificial system, and EEG recordings, which capture activity and dynamics of the human brain at a macroscopic level. A limitation of this work, however, is that neither of these types of data can directly address how the mechanism of rotational remapping might arise from activity in a population of individual neurons. In one recent study that does assess this critical level of neural functioning, Libby and Buschman (2021) recorded from primary auditory cortex of mice while they were trained to associate sequences of auditory stimuli with varying statistical dependencies. Over the course of 4 days, the researchers found that at the population level, the 'memory axis' (i.e., the axis coding for the learned associate of a stimulus), rotated away from the 'sensory axis', which coded for the actually presented sensory stimulus. At the level of individual neurons, they observed that this rotation resulted when a subpopulation of "switching" neurons changed their selectivity, while "stable" neurons maintained their original selectivity. Notably, in this account rotation does not entail the recruitment of additional neurons that were not involved in the representation of stimulus sequences prior to the transformation. Thus, it is possible that changes in neuronal selectivity based on behavioral priority might also have underlain the rotations that we observed in the EEG data. Importantly, however, we cannot rule out the alternative possibility that the representational rotations observed in the EEG data

were supported by the recruitment of new populations of neurons, including from regions that had not been recruited for the initial representation.

Prior to undertaking this work, we were aware of previous accounts of rotational transformations of neural representations. Importantly, however, our results did not have to turn out as they did. Rather, the representational dynamics that we observed in our RNNs arose from the "behavioral demand" of having to learn to perform our WM task. (The same could not be said had we, for example, carried out a simulation of previously observed neural activity patterns.) These rotations thus constitute a *normative* hypothesis of the neural computations required to solve this task, revealed through the optimization of our RNN model. Having observed (with PCA), then quantified (with dPCA), rotational dynamics in our RNN simulations, we subsequently found evidence that they are also characteristic of stimulus processing in humans performing this visual WM task, as measured with EEG.

Because our results represent an independent "discovery" of a phenomenon that has also been described for another very different domain of neural processing, in a different neural system and a different species (auditory perception in mouse auditory cortex (Libby & Buschman, 2021)), they raise the possibility that rotational remapping may be a canonical neural computation. At an intuitive level, it may be that rotational remapping is a computation that is engaged when a single network needs to simultaneously perform multiple functions. This possibility might be assessed by developing ANN architectures that can perform tasks like that from

Libby and Buschman (2021) to determine if they, too, demonstrate rotational dynamics.

If "switching" is a neuron-level implementation of priority-based remapping, and rotation the algorithm that guides this implementation, what is the computational problem that these operations address? We would argue that it is not prioritization per se, because remapping happens to the representation that is *not* prioritized by a cue in a working memory task (or, in the Libby and Buschman (2021) experiment, to the one that is not currently being perceived). Rather, the goal of rotation might be to enable the active representation of multiple pieces of information, while shielding the temporarily irrelevant ones from interfering with ongoing perception and/or behavior.

Our work complements extant models of attentional prioritization in WM. First, it sheds light on the prioritization mechanisms of a continuous-performance WM task (2-back), a design that has recently received less attention than tasks employing retrocuing. Second, compared with the aforementioned computational accounts (Lorenc et al., 2020; Manohar et al., 2019), our use of dPCA provides a data-driven dimensionality reduction approach that does not make assumptions about the representational structure of stimuli. This allows one to examine the unmodeled structure of stimuli in the representational space. Third, our dPCA analyses were applied on a subject-by-subject basis, without assuming that the same representational and/or computational scheme is employed across individuals. Indeed, recent research has showed that representational biases of stimulus

features vary among individuals in higher-order brain areas (Gong & Liu, 2020). An important question for future research is whether the algorithm that dictates the rotational remapping of stimulus information might be implemented by selective suppression (Manohar et al., 2019) and/or gain modulation of population tuning profiles of the stimulus (Lorenc et al., 2020).

The RNNs we simulated have a simple architecture, with a homogeneous LSTM layer, which is, of course, very different from the brain with its heterogeneous patterns of connectivity between neurons with varied functional and structural properties. Although we are encouraged by the fact that rotational transformation has been observed in mammalian brains, understanding the principles and limitations of its implementation will require simulation in more biologically plausible network models. The RNN simulations of Masse et al. (2019), employing different cell types and explicitly simulating factors like receptor time constants and presynaptic depletion of neurotransmitter, offer one promising example. Also missing from our RNN architecture is an explicit source of control, such as that exerted by prefrontal and posterior parietal circuits in the mammalian brain. Through extensive training, our RNNs gradually learned to adjust their connection weights so as to achieve a high level of performance, but this was only possible because each item presented to the network always followed the same representational trajectory. A hallmark of WM in the real world is the ability to flexibly respond to unpredictable changes in environmental exigencies. Thus, an important future goal will be to extend the present work to a network with separate

modules with different connectivity patterns and governed by different learning rules (e.g., Kruijne et al., 2020; O'Reilly & Frank, 2006).

The way that representational rotation might guide flexible behavior remains to be further elucidated. As illustrated by the 2-back task, multiple pieces of information are kept in WM, but behavior is selectively driven by the prioritized memorandum and not by unprioritized ones. Why is rotation an optimal strategy to carry out this function compared to other alternatives? In the current work, we focused on static representations of the UMI and PMI during the delay period. Further insight will be gained by tracing representational trajectories at a higher temporal resolution, and over the entirety of the trial. In addition, it will likely be fruitful to examine how rotations in stimulus-specific subspaces interface with dimensions relevant for decision-making. For example, compared to the UMI, which is in an "output-null" state, one might expect the PMI to be in an "output-potent" state that can interact with downstream readout mechanisms (Kaufman et al., 2014). Another intriguing future direction will be to investigate whether the different metrics and parameters of the representational rotation (e.g., rotational index, rotational angle, etc.) may relate to individual differences in behavioral performance. This could lead to insights on how representational transformation is linked to behavioral success or failure. The current data set is not fit to conduct such analyses because our subjects' performance was overall too high to assess errors.

Finally, an important outstanding question is understanding the factors that trigger the representational transformations that we have studied here. Because the

PMI is represented in a format similar to its perceptual representation (e.g., Wan et al., 2020), and the retrocuing of one item leads to the priority-based remapping of the uncued item (van Loon et al., 2018; Yu, Teng & Postle, 2020), we speculate that the rotation away from the format of the perceived item may not be an active, controlled process. By analogy to thermodynamics, it's as though being in the focus of attention entails being in a high-energy state, and the withdrawal of attention allows a representation to "relax" into a low-energy state. From an information theoretic perspective, such "relaxation" might correspond to the recoding that minimizes distortion/information loss during storage in the noisy medium that is the brain (Koyluoglu et al., 2017). Nonetheless, even if such a "relaxation" account is true, it only accounts for the transformation of an item's representation from its perceived/attended state to its unprioritized state. The ensuing operation of reinstating the UMI into a PMI format must necessarily be an active, controlled process. This process corresponds to what psychologists refer to as attentional selection or retrieval into the focus of attention.

To conclude, using neural network models to generate hypotheses, then testing these hypotheses on empirical neural data, we successfully identified rotational remapping as a candidate neural computation that governs the dynamic prioritization of a subset of the contents of working memory.

Data Availability. All processed human EEG data, code, network training sets and trained networks are available at https://osf.io/sgqvn/ (DOI 10.17605/OSF.IO/SGQVN) on Open Science Framework.

Competing interests. The authors have no competing interests.

Author Contributions.

Conceptualization: Q.W., B.R.P.

Data Curation: Q.W.

Formal Analysis: Q.W., J.A.M.

Funding Acquisition: B.R.P.

Investigation: Q.W.

Methodology: Q.W., J.A.M.

Project Administration: Q.W.

Resources: *Not Applicable*

Software: Q.W., J.A.M.

Supervision: B.R.P.

Validation: Q.W., J.A.M.

Visualization: Q.W., J.A.M.

Writing – Original Draft Preparation: Q.W.

Writing – Review & Editing: Q.W., J.A.M., B.R.P.

Funding. This work was supported by the National Institutes of Health (grant no. R01-MH064498).

Acknowledgements. We thank Drs. Yuri Saalmann, Joseph Austerweil, Timothy Rogers, Jacqueline Fulvio, Qing Yu and Mohsen Afrasiabi for helpful discussion and critical feedback.

References

Barak, O., & Tsodyks, M. (2014). Working models of working memory. *Current Opinion in Neurobiology*, *25*, 20–24. https://doi.org/10.1016/j.conb.2013.10.008

Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Current Opinion in Behavioral Sciences*, *1*, 23–31. https://doi.org/10.1016/j.cobeha.2014.08.001

Gong, M., & Liu, T. (2020). Biased Neural Representation of Feature-Based Attention in the Human Frontoparietal Network. *The Journal of Neuroscience*, *40*(43), 8386–8395. https://doi.org/10.1523/JNEUROSCI.0690-20.2020

Graves, A., Mohamed, A., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 6645–6649. https://doi.org/10.1109/ICASSP.2013.6638947

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Kaufman, M. T., Churchland, M. M., Ryu, S. I., & Shenoy, K. V. (2014). Cortical activity in the null space: Permitting preparation without movement. *Nature Neuroscience*, *17*(3), 440–448. https://doi.org/10.1038/nn.3643

Kell, A. J., & McDermott, J. H. (2019). Deep neural network models of sensory systems: Windows onto the role of task constraints. *Current Opinion in Neurobiology*, *55*, 121–132. https://doi.org/10.1016/j.conb.2019.02.003

Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *ArXiv:1412.6980 [Cs]*. http://arxiv.org/abs/1412.6980

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., Qi, X.-L., Romo, R., Uchida, N., & Machens, C. K. (2016). Demixed principal component analysis of neural population data. *ELife*, *5*, e10989. https://doi.org/10.7554/eLife.10989

Koyluoglu, O. O., Pertzov, Y., Manohar, S., Husain, M., & Fiete, I. R. (2017). Fundamental bound on the persistence and capacity of short-term memory stored as graded persistent activity. *ELife*, *6*, e22225. https://doi.org/10.7554/eLife.22225

Kruijne, W., Bohte, S. M., Roelfsema, P. R., & Olivers, C. N. L. (2020). Flexible Working Memory Through Selective Gating and Attentional Tagging. *Neural Computation*, *33*(1), 1–40. https://doi.org/10.1162/neco_a_01339

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Decoding Attended Information in Short-term Memory: An EEG Study. *Journal of Cognitive Neuroscience*, *25*(1), 127–142. https://doi.org/10.1162/jocn_a_00305

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2011). Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience*, *24*(1), 61–79. https://doi.org/10.1162/jocn_a_00140

Libby, A., & Buschman, T. J. (2021). Rotational dynamics reduce interference between sensory and memory representations. *Nature Neuroscience*, 1–12. https://doi.org/10.1038/s41593-021-00821-9

Lorenc, E. S., Vandenbroucke, A. R. E., Nee, D. E., de Lange, F. P., & D'Esposito, M. (2020). Dissociable neural mechanisms underlie currently-relevant, future-relevant, and discarded working memory representations. *Scientific Reports*, *10*(1), 11195. https://doi.org/10.1038/s41598-020-67634-x

Manohar, S. G., Zokaei, N., Fallon, S. J., Vogels, T. P., & Husain, M. (2019). Neural mechanisms of attending to items in working memory. *Neuroscience & Biobehavioral Reviews*, *101*, 1–12. https://doi.org/10.1016/j.neubiorev.2019.03.017

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84. https://doi.org/10.1038/nature12742

Masse, N. Y., Yang, G. R., Song, H. F., Wang, X.-J., & Freedman, D. J. (2019). Circuit mechanisms for the maintenance and manipulation of information in working memory. *Nature Neuroscience*, *22*(7), 1159. https://doi.org/10.1038/s41593-019-0414-3

Merrikhi, Y., Clark, K., Albarran, E., Parsa, M., Zirnsak, M., Moore, T., & Noudoost, B. (2017). Spatial working memory alters the efficacy of input to visual cortex. *Nature Communications*, *8*(1), 15041. https://doi.org/10.1038/ncomms15041

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends in Cognitive Sciences*, *21*(6), 449–461. https://doi.org/10.1016/j.tics.2017.03.010

O'Reilly, R. C., & Frank, M. J. (2006). Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. *Neural Computation*, *18*(2), 283–328. https://doi.org/10.1162/089976606775093909

Parthasarathy, A., Herikstad, R., Bong, J. H., Medina, F. S., Libedinsky, C., & Yen, S.-C. (2017). Mixed selectivity morphs population codes in prefrontal cortex. *Nature Neuroscience*, *20*(12), 1770. https://doi.org/10.1038/s41593-017-0003-2

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., Berker, A. de, Ganguli, S., Gillon, C. J., Hafner, D., Kepecs, A., Kriegeskorte, N., Latham, P., Lindsay, G. W., Miller, K. D., Naud, R., Pack, C. C., … Kording, K. P. (2019). A deep learning framework for neuroscience. *Nature Neuroscience*, *22*(11), 1761–1770. https://doi.org/10.1038/s41593-019-0520-2

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, *354*(6316), 1136–1139. https://doi.org/10.1126/science.aah7011

Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences*, *19*(7), 394–405. https://doi.org/10.1016/j.tics.2015.05.004

Stokes, M. G., Muhle-Karbe, P. S., & Myers, N. E. (2020). Theoretical distinction between functional states in working memory and their corresponding neural states. *Visual Cognition*, *28*(5–8), 420–432. https://doi.org/10.1080/13506285.2020.1825141

Sussillo, D., Churchland, M. M., Kaufman, M. T., & Shenoy, K. V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. *Nature Neuroscience*, *18*(7), 1025–1033. https://doi.org/10.1038/nn.4042

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. *ArXiv:1409.3215 [Cs]*. http://arxiv.org/abs/1409.3215

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

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*(8), 190228. https://doi.org/10.1098/rsos.190228

Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T., & Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, *22*(2), 297–306. https://doi.org/10.1038/s41593-018-0310-2

Yu, Q., Teng, C., & Postle, B. R. (2020). Different states of priority recruit different neural representations in visual working memory. *PLOS Biology*, *18*(6), e3000769. https://doi.org/10.1371/journal.pbio.3000769

Figure captions

Figure 1. 2-back task structure in the Wan et al. (2020) EEG study. The presentation of each stimulus is followed by a 50 ms blank screen, a 200 ms radial checkerboard mask, a variable delay from 2.8 to 3.2 s (only the first 2.8 s, which is common to all stimulus events, was used for analysis), and then the next stimulus was presented, upon which the match vs. non-match response is to be made.

Figure 2. IEM reconstruction of the 2-back task. Averaged concatenation of the item *n* and item *n + 1* stimulus events to form a "trial" across which *n* transitions from probe to UMI to PMI, each stimulus event running from 200 ms before stimulus onset to 2.8 s after mask offset. The dash-dot line in the center of the figure marks the discontinuity between the two stimulus events (due to variable-length ISIs). White squares in the middle of the 90° column mark timepoints with significant stimulus reconstruction (cluster-based permutation test: $p < .05$, two-tailed). On the right are IEM reconstructions corresponding to the two 2 s windows centered in the post-mask ISIs after item *n* and item *n + 1*, respectively (denoted by dashed lines). "*" indicates $p < .05$ (two-tailed *t* test), FDR-corrected for multiple comparisons. As the figure shows, IEM reconstruction of stimulus *n* is "flipped" relative to the training data when it is a UMI and "flipped back" when its status becomes PMI, demonstrating priority-based remapping.

Figure 3. PCA visualization of LSTM hidden layer activity of an example network (#7). (A) shows a 9-timestep time course of the 2-back task, which includes a "trial" (from stimulus *n*'s presentation to *delay 2:2*), plus the subsequent stimulus event (*delay 3:1* and *delay 3:2* intended to show representation of item when it is no longer relevant for the task). Plotted trials contain no overlapping stimuli (1600 in total). Each dot in the figure indicates the representation of stimulus *n*. Each color corresponds to one of the six stimulus types, and the black dashed line illustrates the stimulus coding axis. (B) Same as (A) except that colors now correspond to an item's status for the *n*-to-*n + 2* comparison that occurs at timestep *n + 2* (red: match trials, black: non-match trials). Blue dashed line with shadows in (A) and (B) at timestep *n + 2* illustrates the decision-based structure. As can be seen in (A), the stimulus coding axis rotates counterclockwise (in the image plane) over time such that it becomes "perpendicular" to the decision structure at timestep *n + 1* and aligns with it at timestep *n + 2*.

Figure 4. Rotation analyses for one representative RNN and for the EEG data from one representative network/subject. Both sets of plots are projections of UMI and PMI data onto the UMI- and PMI-dependent dPCA subspaces for that RNN/subject

(*x*-axis: 1st dPC; *y*-axis: 2nd dPC). (A) Results for RNN #7. Each dot denotes RNN hidden layer activity averaged over the 2nd delay timestep of the epoch, colors same as Fig. 3A. In each subspace, UMI means can be rotationally mapped to the PMI means. The *p*-values from permutation testing, rotational index (RI) and rotational angle are noted below. (B) Results for subject #20. Each dot denotes EEG data averaged over the delay time window (1150-3150 ms after stimulus onset), for the orientation indicated in the legend. (Note that legend only applies to panel B, because orientation was not a feature in the RNN simulations.) Dots of adjacent orientations are connected by a gray line to facilitate visualization of the representational structure.

Figure 5. Distributions of rotational indexes (RI) for the rotational transformations between UMI and PMI means, in UMI-dependent (top) and PMI-dependent (bottom) subspaces, for RNN networks (A) and EEG subjects (B). Smaller RI values indicate more rotational structure in the transformation between the UMI and PMI means. Networks/subjects with significant rotations were plotted in blue bars and those with non-significant rotations in red bars.

Figure 6. *P*-values from the permutation test on RI of the rotational transformations between UMI and PMI means, in both UMI-dependent (top) and PMI-dependent (bottom) subspaces, for RNN networks (A) and EEG subjects (B). The red horizontal line marks the significance threshold ($p = 0.05$).

Figure 7. Distribution of angles-of-rotation for the rotational transformations between UMI and PMI means in both UMI-dependent (top) and PMI-dependent (bottom) subspaces, for RNN networks (A) and EEG subjects (B). Networks/subjects with significant rotations are plotted in blue bars and those with non-significant rotations in red bars.

Figure 1.

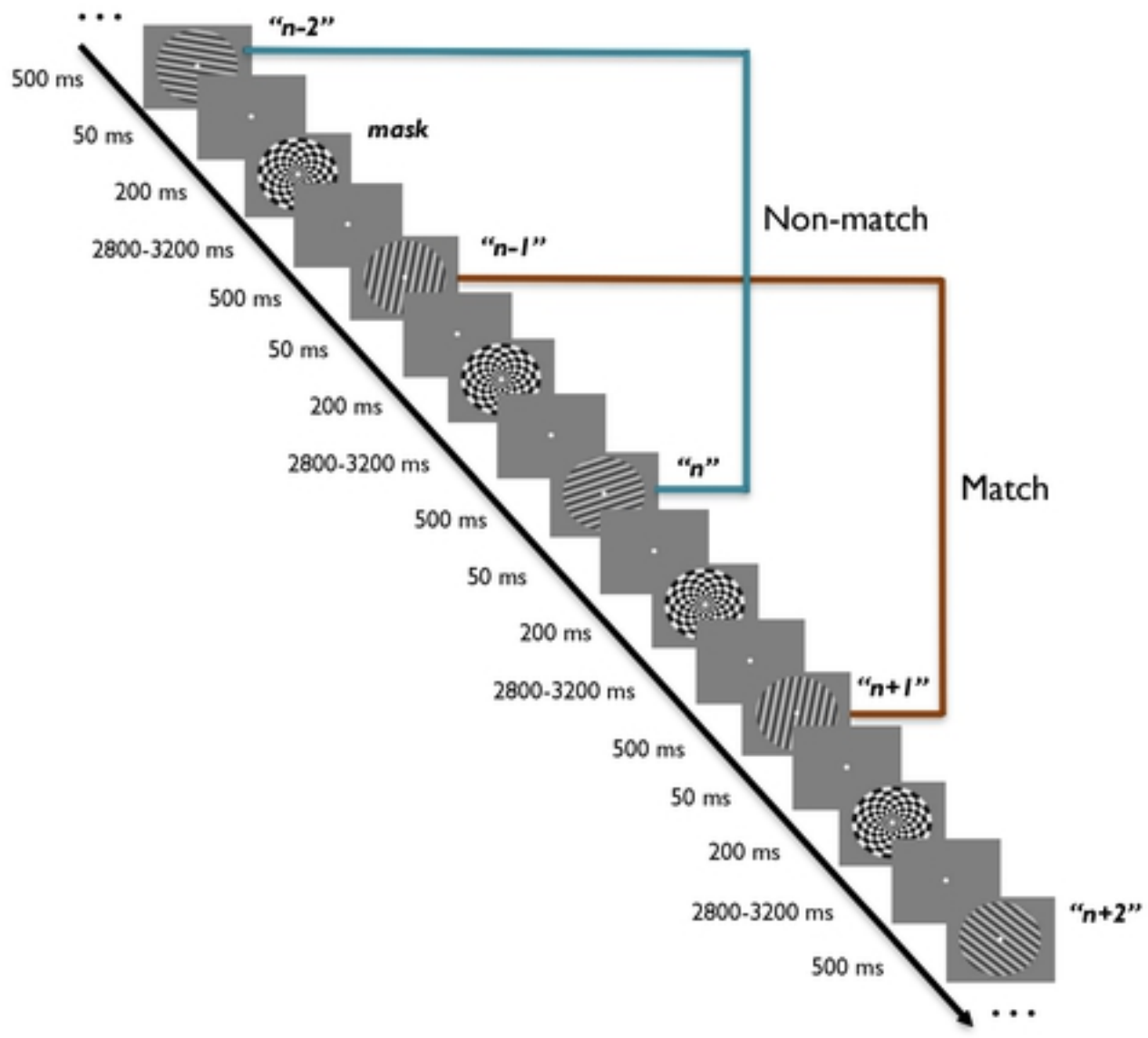
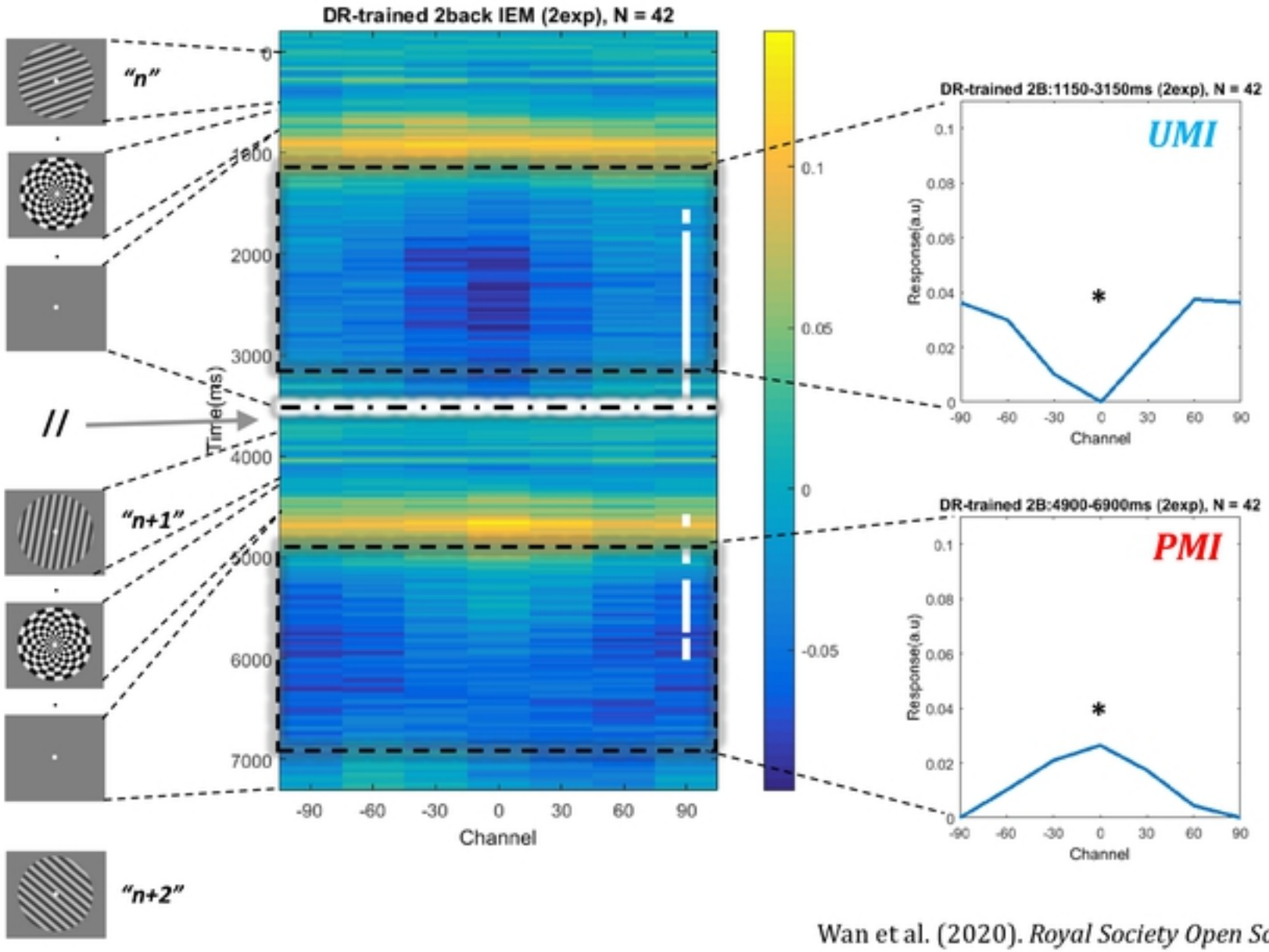

. . .
"n-2"
500 ms
50 ms
mask
200 ms
2800-3200 ms
"n-1"
500 ms
50 ms
200 ms
2800-3200 ms
"n"
500 ms
50 ms
200 ms
2800-3200 ms
"n+1"
500 ms
50 ms
200 ms
2800-3200 ms
"n+2"
500 ms
. . .
Non-match
Match

Figure 2.



Wan et al. (2020). *Royal Society Open Science*

Figure 3.

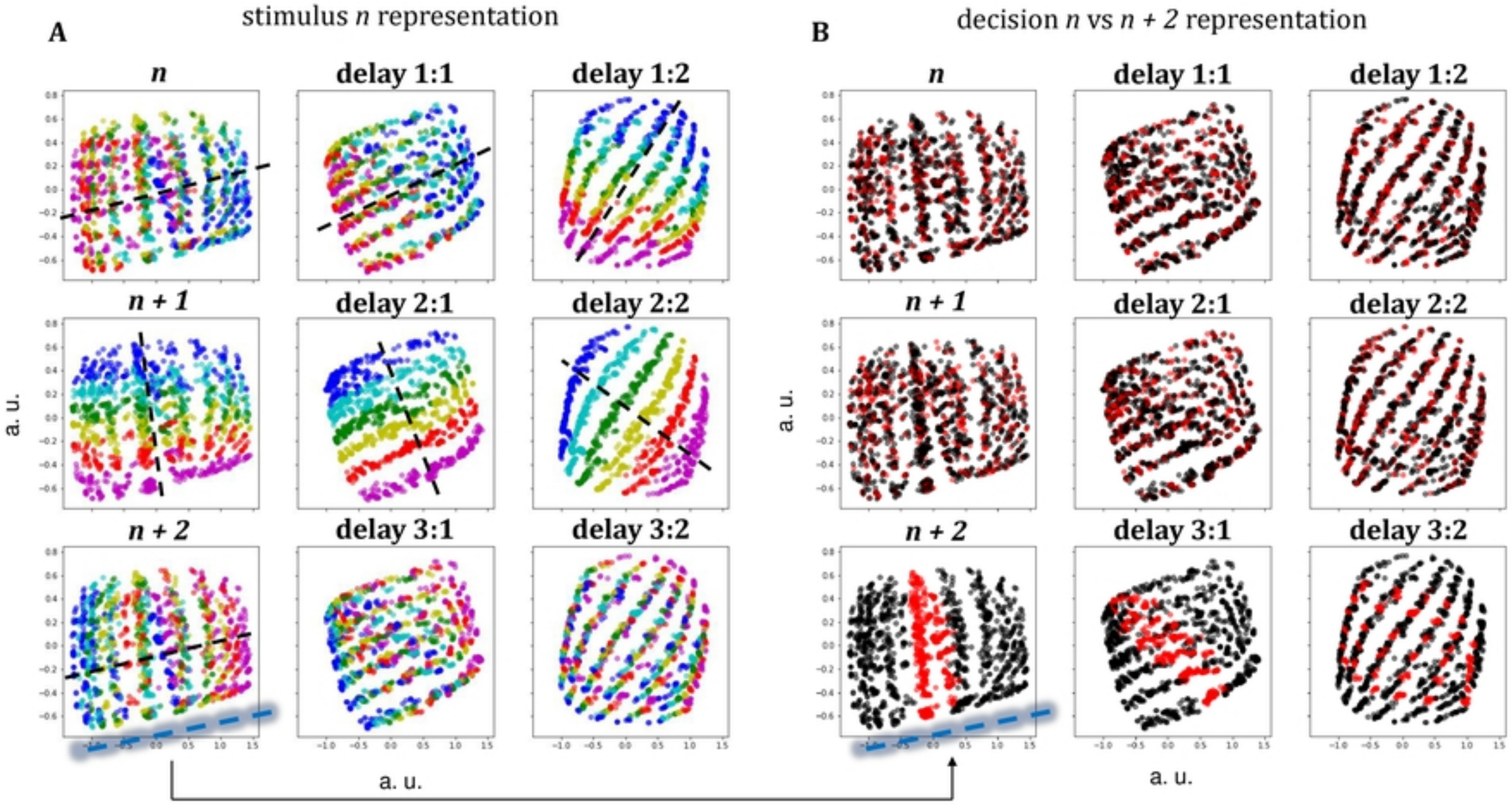

A
stimulus n representation
n
delay 1:1
delay 1:2
n + 1
delay 2:1
delay 2:2
n + 2
delay 3:1
delay 3:2
a. u.
a. u.
B
decision n vs n + 2 representation
n
delay 1:1
delay 1:2
n + 1
delay 2:1
delay 2:2
n + 2
delay 3:1
delay 3:2
a. u.
a. u.

Figure 4.

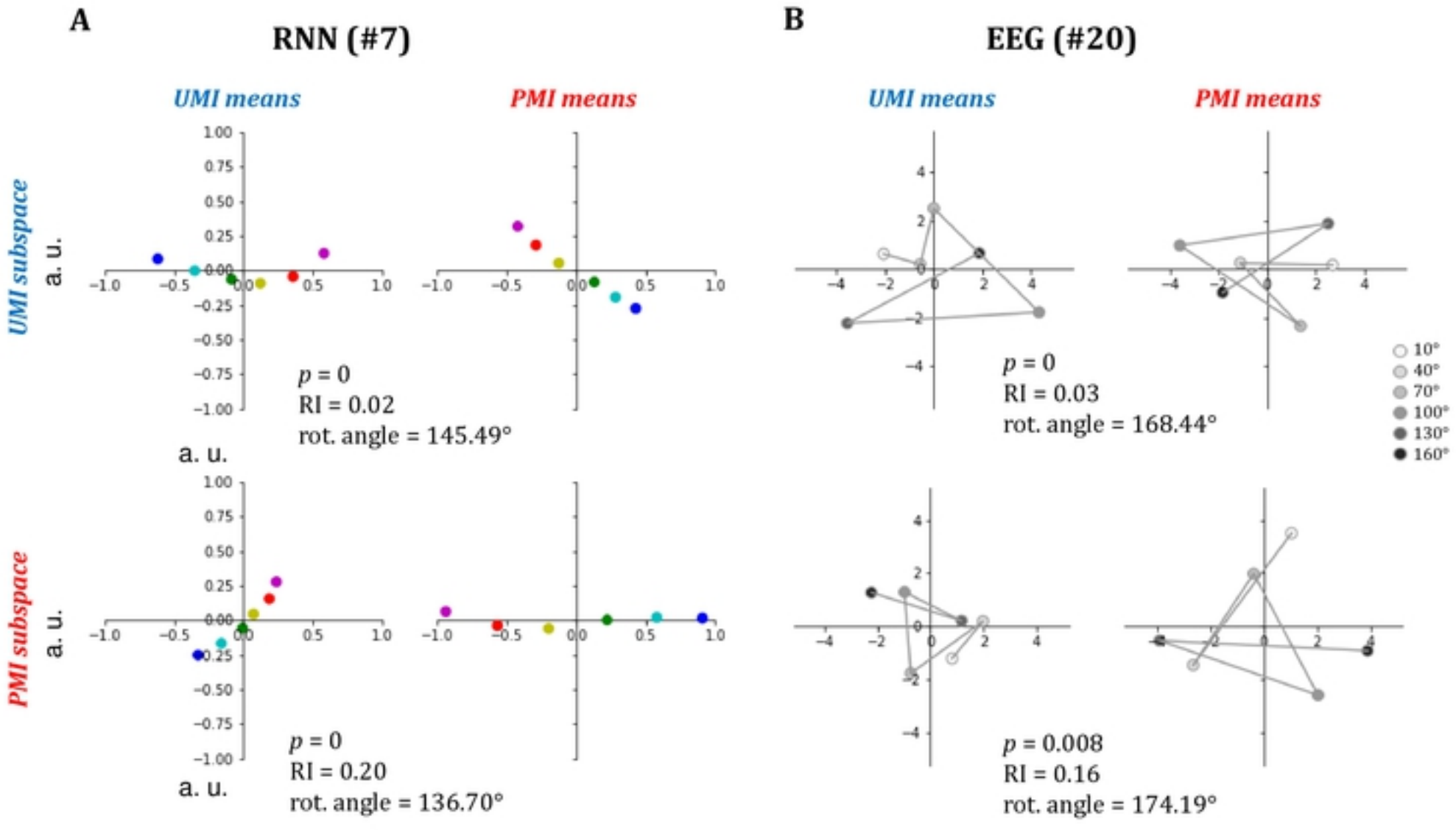

A
RNN (#7)
UMI means
PMI means
UMI subspace
PMI subspace
a. u.
p = 0
RI = 0.02
rot. angle = 145.49°
p = 0
RI = 0.20
rot. angle = 136.70°
B
EEG (#20)
UMI means
PMI means
p = 0
RI = 0.03
rot. angle = 168.44°
p = 0.008
RI = 0.16
rot. angle = 174.19°
10°
40°
70°
100°
130°
160°

Figure 5.

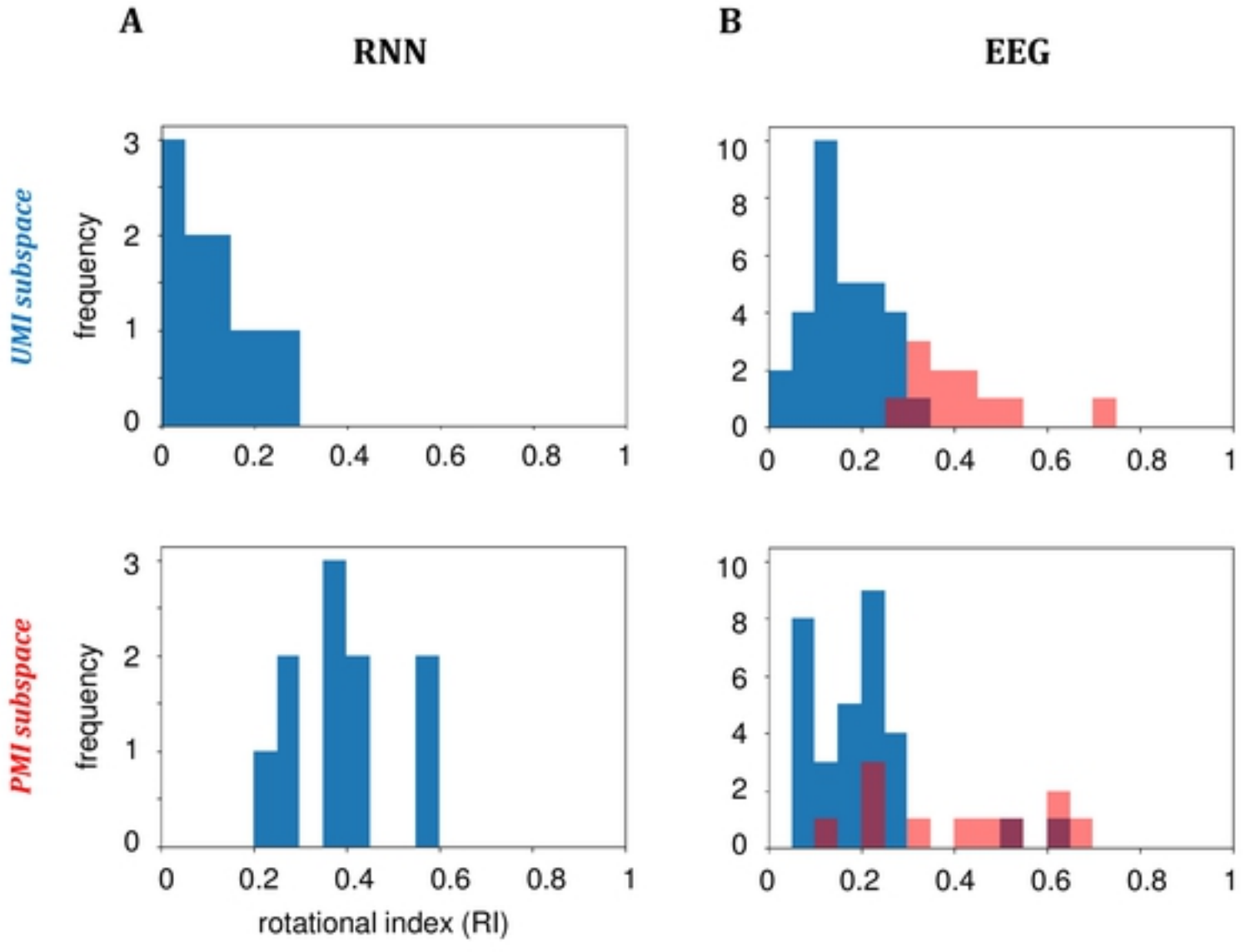

A
RNN
B
EEG
UMI subspace
PMI subspace
frequency
rotational index (RI)
0
0.2
0.4
0.6
0.8
1
0
1
2
3
0
2
4
6
8
10

Figure 6.

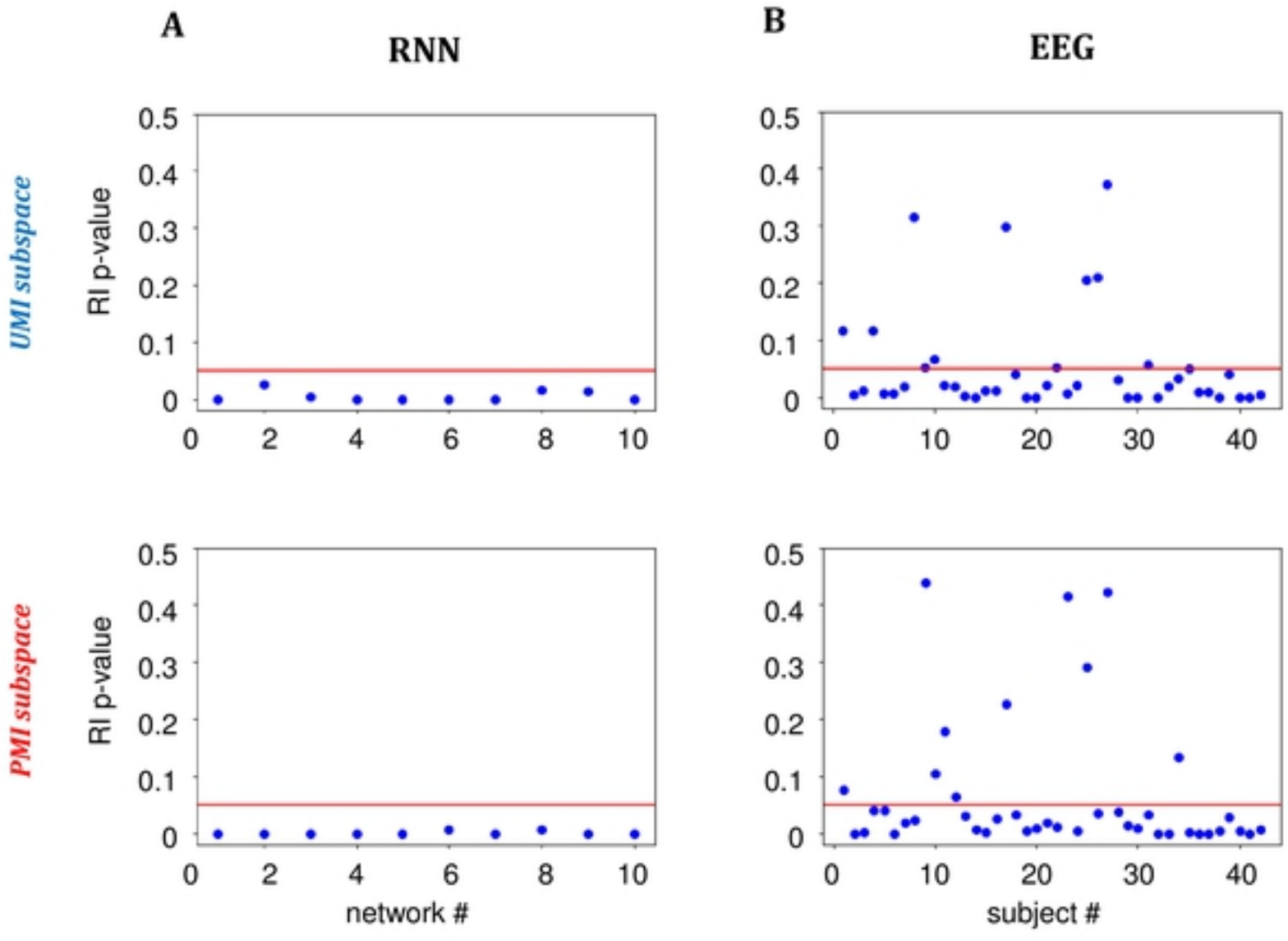
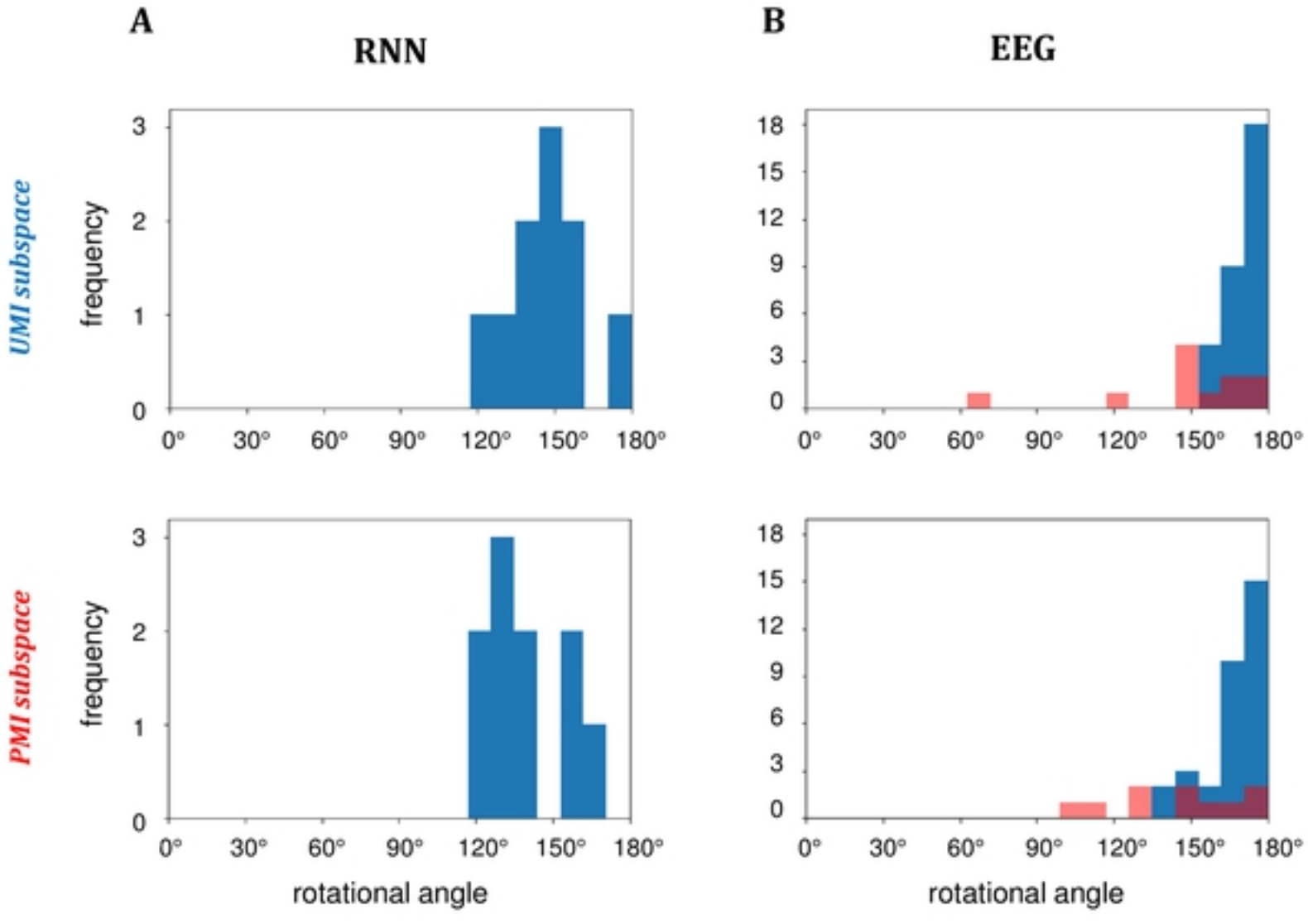

A
RNN
B
EEG
UMI subspace
PMI subspace
RI p-value
RI p-value
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
network #
0
10
20
30
40
subject #

Figure 7.


A
RNN
B
EEG
UMI subspace
PMI subspace
frequency
rotational angle
0°
30°
60°
90°
120°
150°
180°
0
1
2
3
6
9
12
15
18