Neuron

Volume 109, Issue 22, 17 November 2021, Pages 3699-3712.e6

Article

Joint representation of working memory and uncertainty in human cortex

Hsin-Hung Li [1, 4], Thomas C. Sprague [1, 3, 4], Aspen H. Yoo [1], Wei Ji Ma [1, 2, 5, 6], Clayton E. Curtis [1, 2, 5, 6, 7]

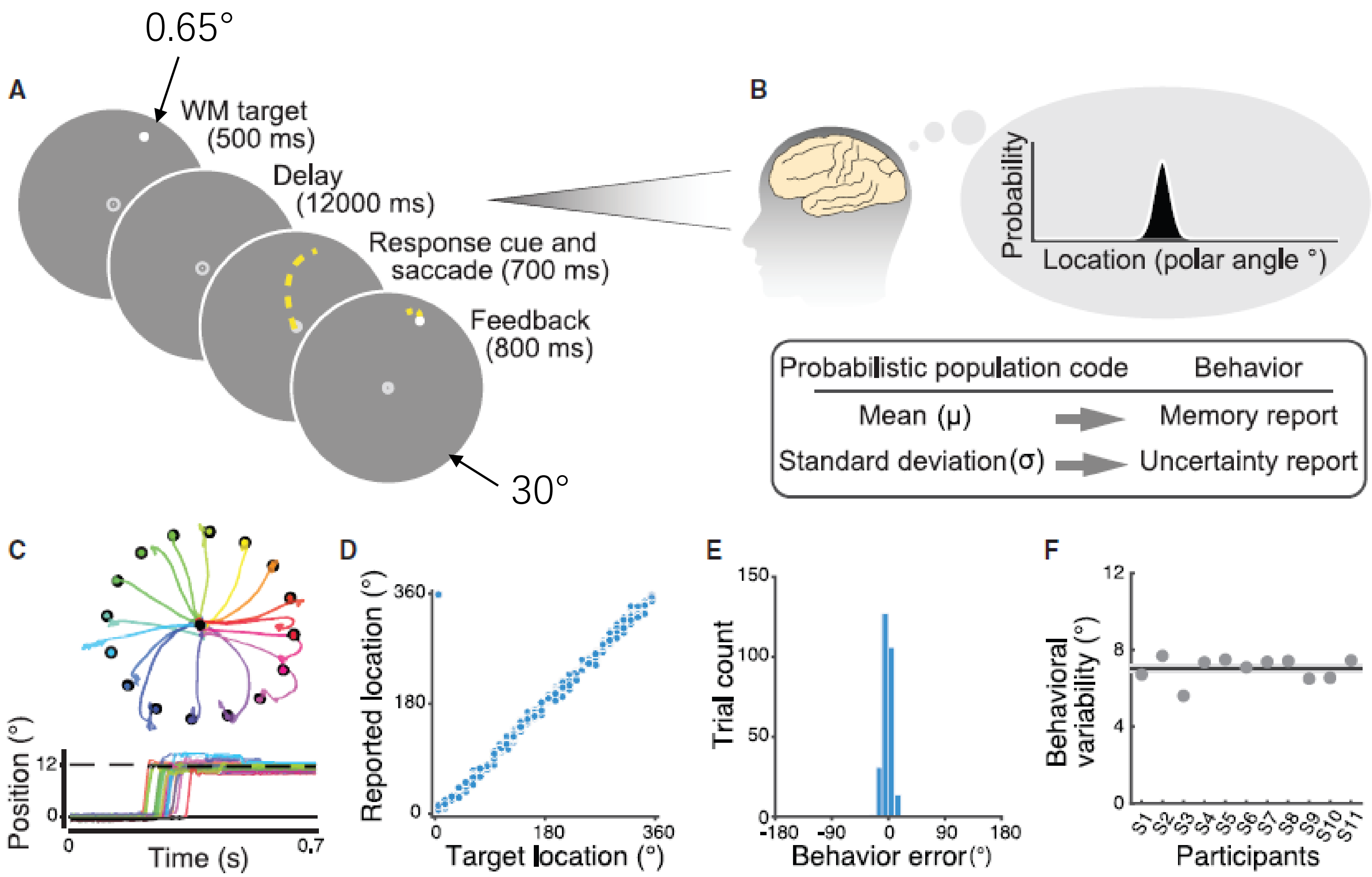
1 Department of Psychology, New York University, New York, NY 10003, USA

2 Center for Neural Science, New York University, New York, NY 10003, USA

3 Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA 93106, USA

Reporter: Chenmiao Liu

Procedures and working memory performance in experiment 1



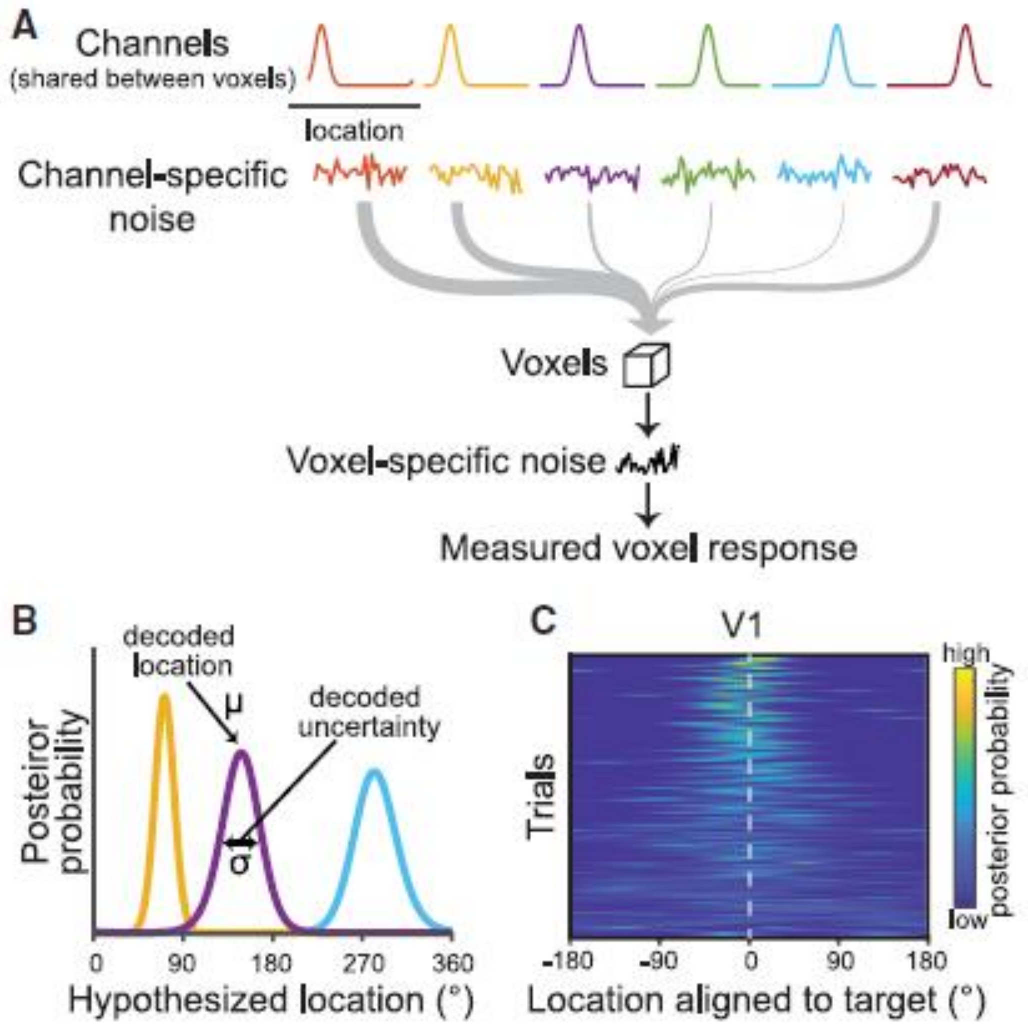
- 11 participants took part in a memory-guided saccade task
- The intertrial interval was pseudo-randomly chosen to be 6, 9, or 12 s
- Each participant completed 304 to 496 trials (346 trials per participant on average)

16 trials × 9 to 10 runs × 2 to 3 sessions

- The working memory target had an eccentricity at 12° from the central fixation point and its polar angle was pseudo-randomly chosen from 1 of 32 locations that evenly tiled the full circle within each run

The behavioral variability quantified by the SD of the memory error distribution

Generative model used to estimate and decode working memory representations



- Use the method named TAFKAP model the multivariate voxel response given the stimulus location (polar angle) as a multivariate normal distribution

$$f(s)_k = \lfloor \cos(s - \phi_k) \rfloor^8$$

[ ] represents half-wave rectification and f_k is the center of the kth channel. The response of ith voxel b_i given a stimulus s is then modeled as

$$b_i(s) = \sum_{k=1}^{8} W_{ik}(f_k(s) + \eta_k) + \upsilon_i$$

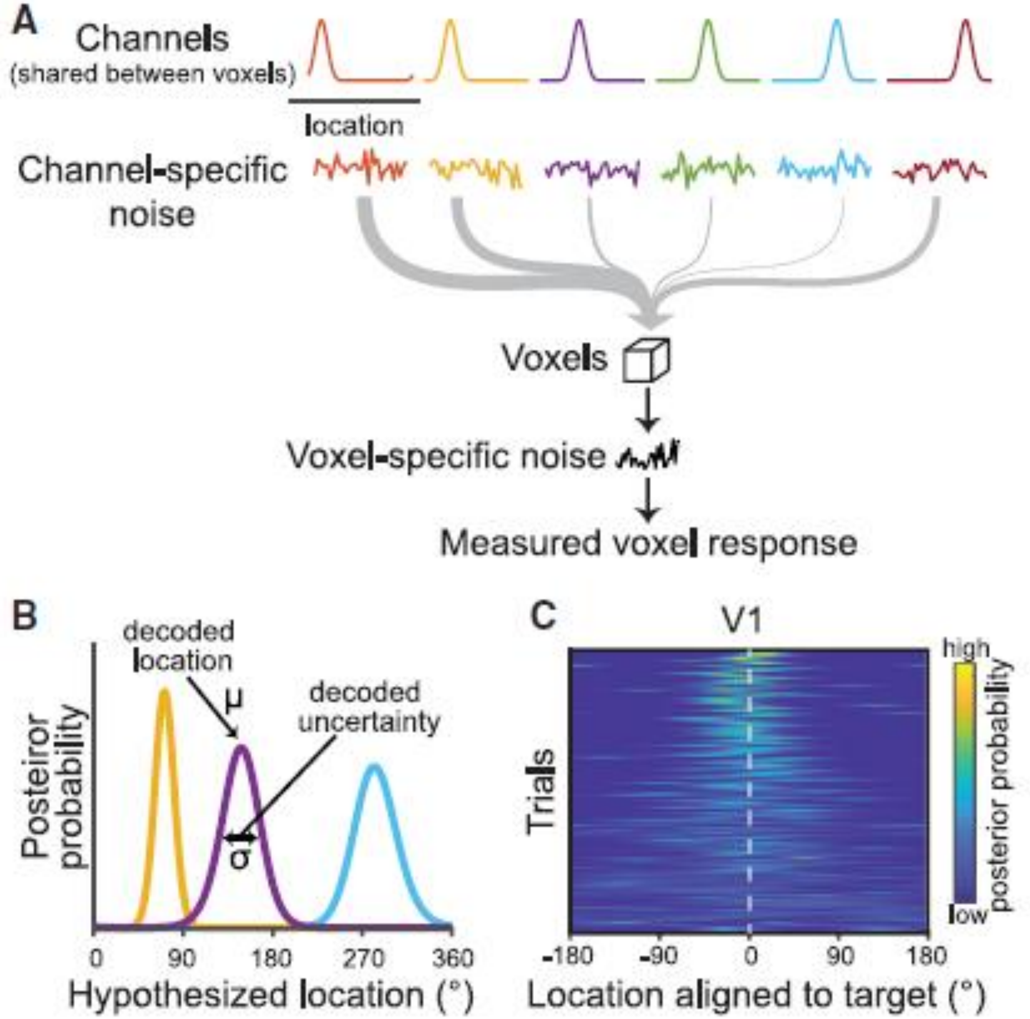
where **W** is a weighting matrix that determines the weights of each basis function for each voxel.

$$\eta \sim N(0, \sigma^2 \mathbf{I}) \qquad \upsilon \sim N(0, \Sigma). \qquad \Sigma = \rho\tau\tau^T + (1-\rho)\mathbf{I} \circ \tau\tau^T$$

where ○ represents Hadamard product, element-wise product between two matrices. Thus, the theoretical covariance matrix of the multivariate response of the voxels given a stimulus s is

$$\Omega_0 = \rho\tau\tau^T + (1-\rho)\mathbf{I} \circ \tau\tau^T + \sigma^2 \mathbf{W}\mathbf{W}^T$$

Generative model used to estimate and decode working memory representations



$$\Omega_0 = \rho\tau\tau^T + (1-\rho)\mathbf{I}\circ\tau\tau^T + \sigma^2\mathbf{W}\mathbf{W}^T$$

τ is a vector representing the standard deviation of the noise of each voxel

In addition to the theoretical covariance matrix, the model also considered the empirical sample covariance

$$\Omega_{sample} = \frac{1}{N_{train}}\left(\mathbf{B} - \widehat{\mathbf{W}}\mathbf{G}\right)\left(\mathbf{B} - \widehat{\mathbf{W}}\mathbf{G}\right)^T$$

B is the training data and **G** is the response of the basis functions given the training set stimuli. Thus, for each training dataset, we assumed that the voxel activity pattern followed a multivariate normal distribution.

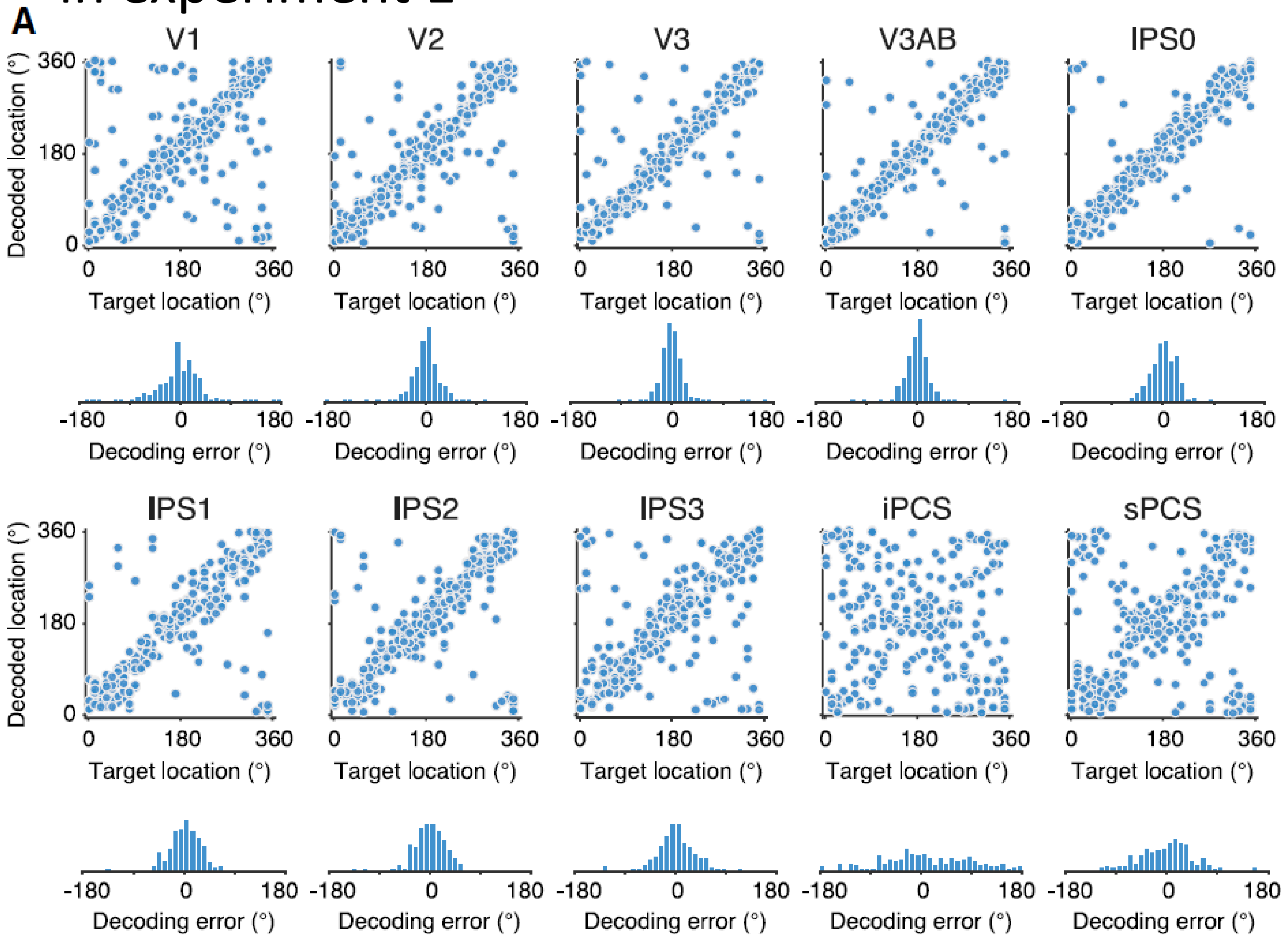
$$p(\mathbf{b}|s) \sim N(\mathbf{W}f(s), \Omega)$$

$$\Omega = \lambda\Omega_0 + (1-\lambda)\Omega_{sample}$$

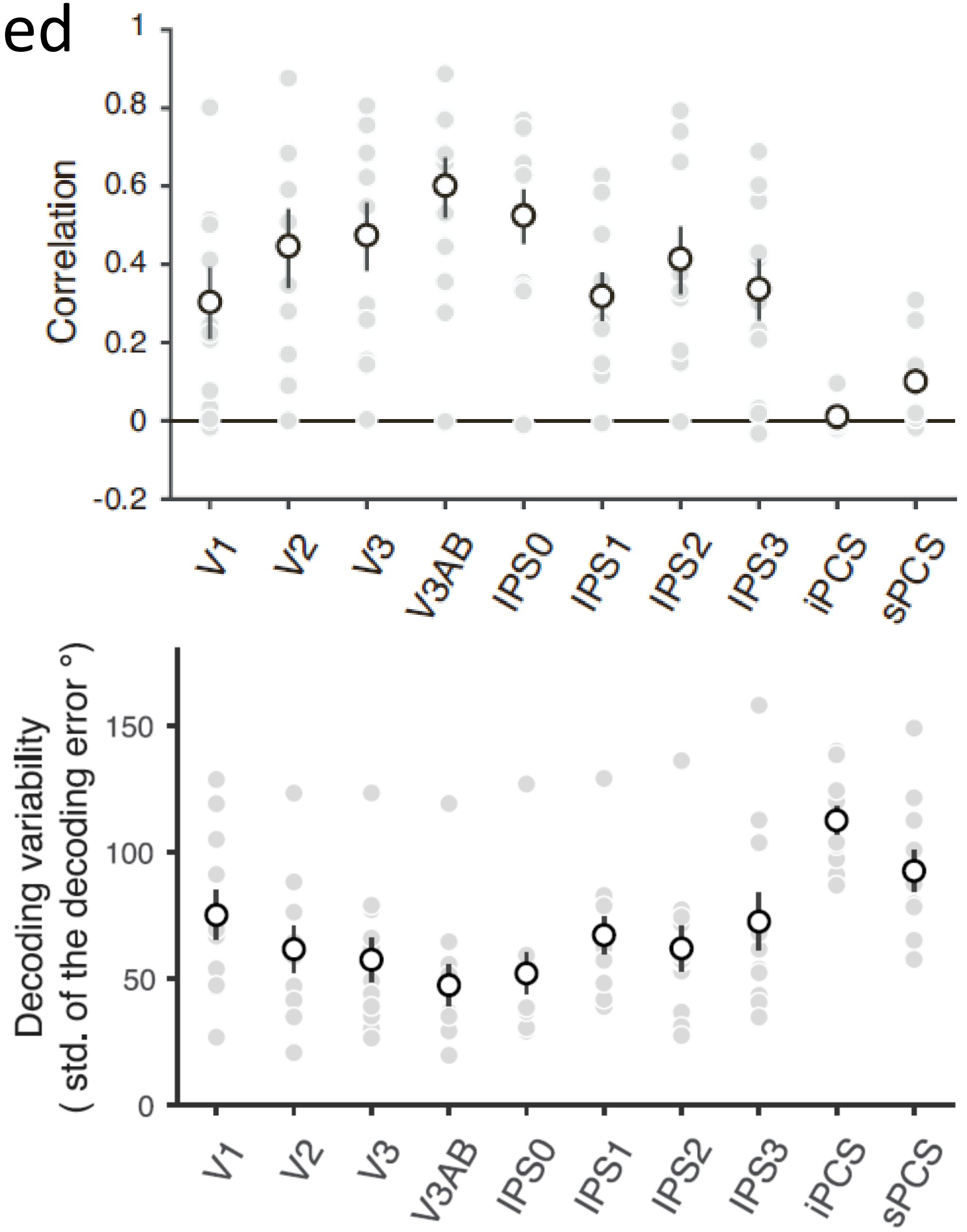
Put voxel response into the model and using a leave-one-run-out cross-validation and bootstrap procedure. the posterior probability of the stimulus given the multivariate voxel response **b** was computed as

$$p(s|\mathbf{b};\theta_j) = \frac{p(\mathbf{b}|s;\theta_j)p(s)}{\int p(\mathbf{b}|s;\theta_j)p(s)ds}$$

Working memory content can be precisely decoded in experiment 1

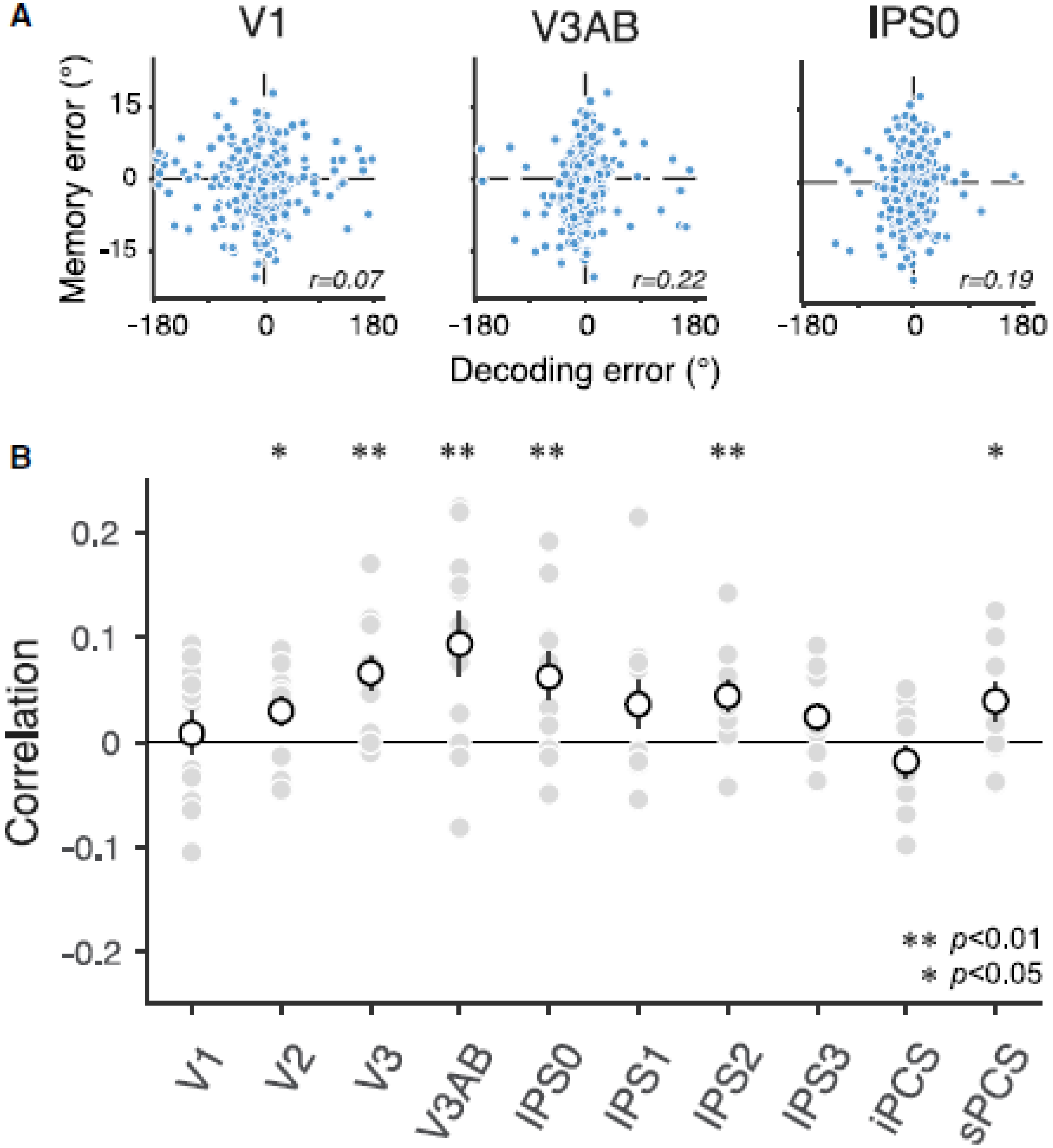


Decoding performance of an example participant

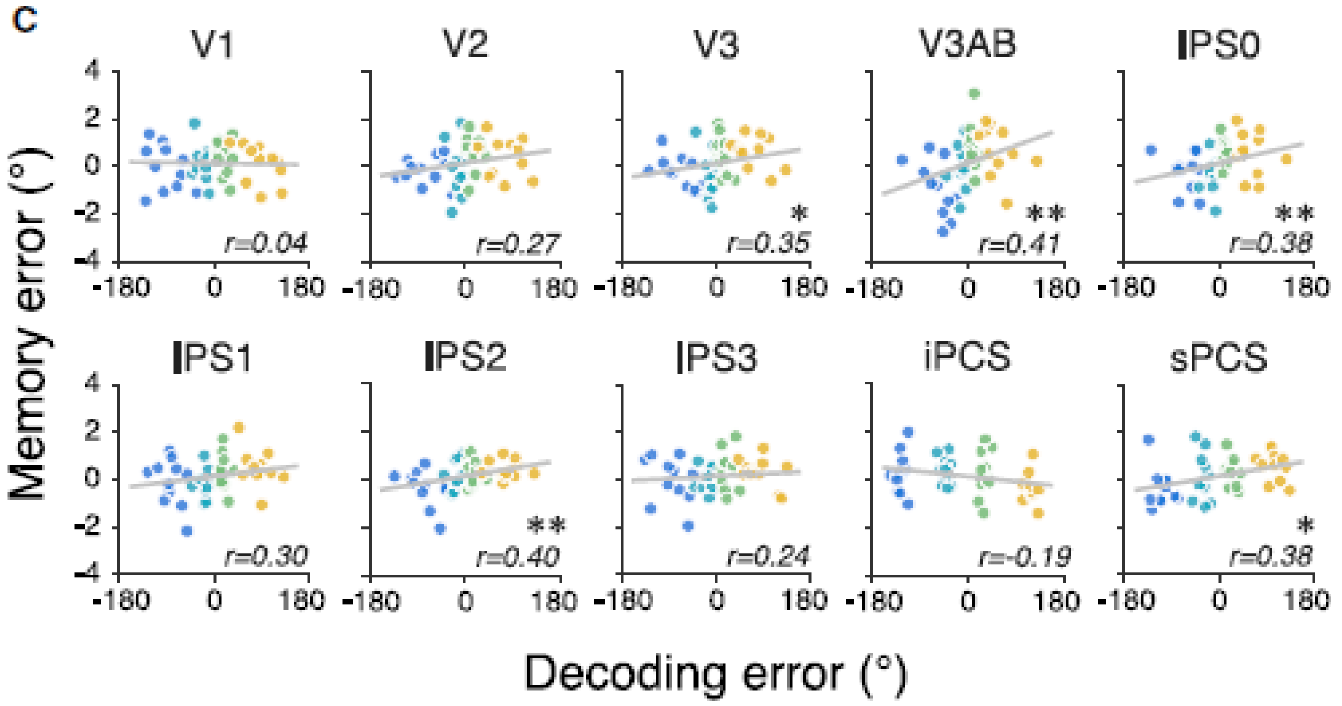


Decoding performance at the individual level

Errors in neural decoding of working memory predict behavioral memory errors in experiment 1

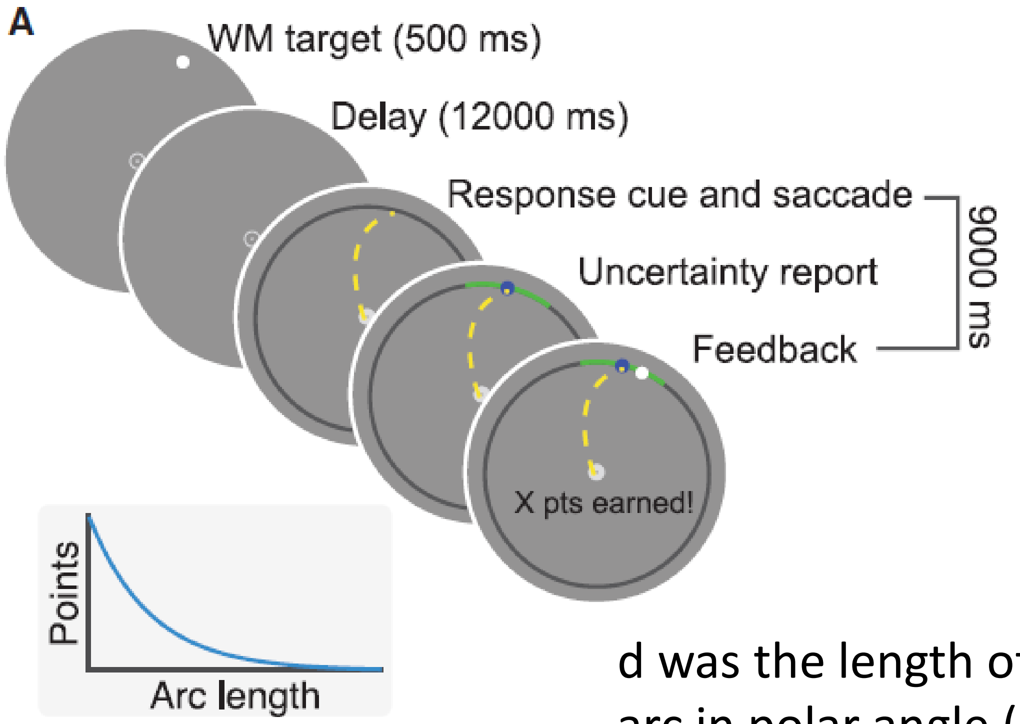


The trial-wise circular correlation between memory error and decoding error for each participant and ROI



we quantified the correlations by binning the trials on the basis of their decoding errors, computing the memory error of each bin, and pooling the data across participants.

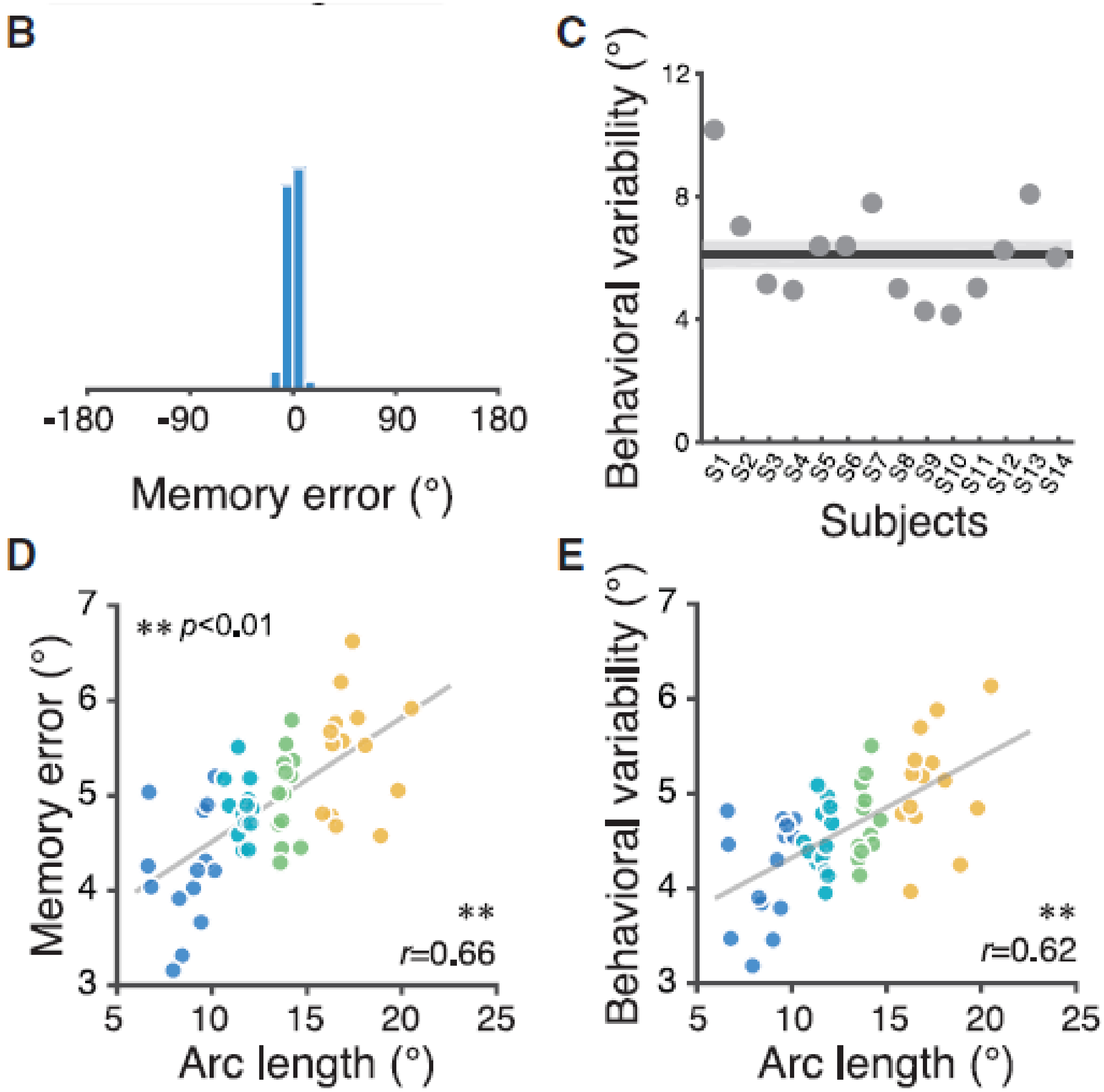
Procedures and working memory performance in experiment 2



$100e^{-0.08d}$

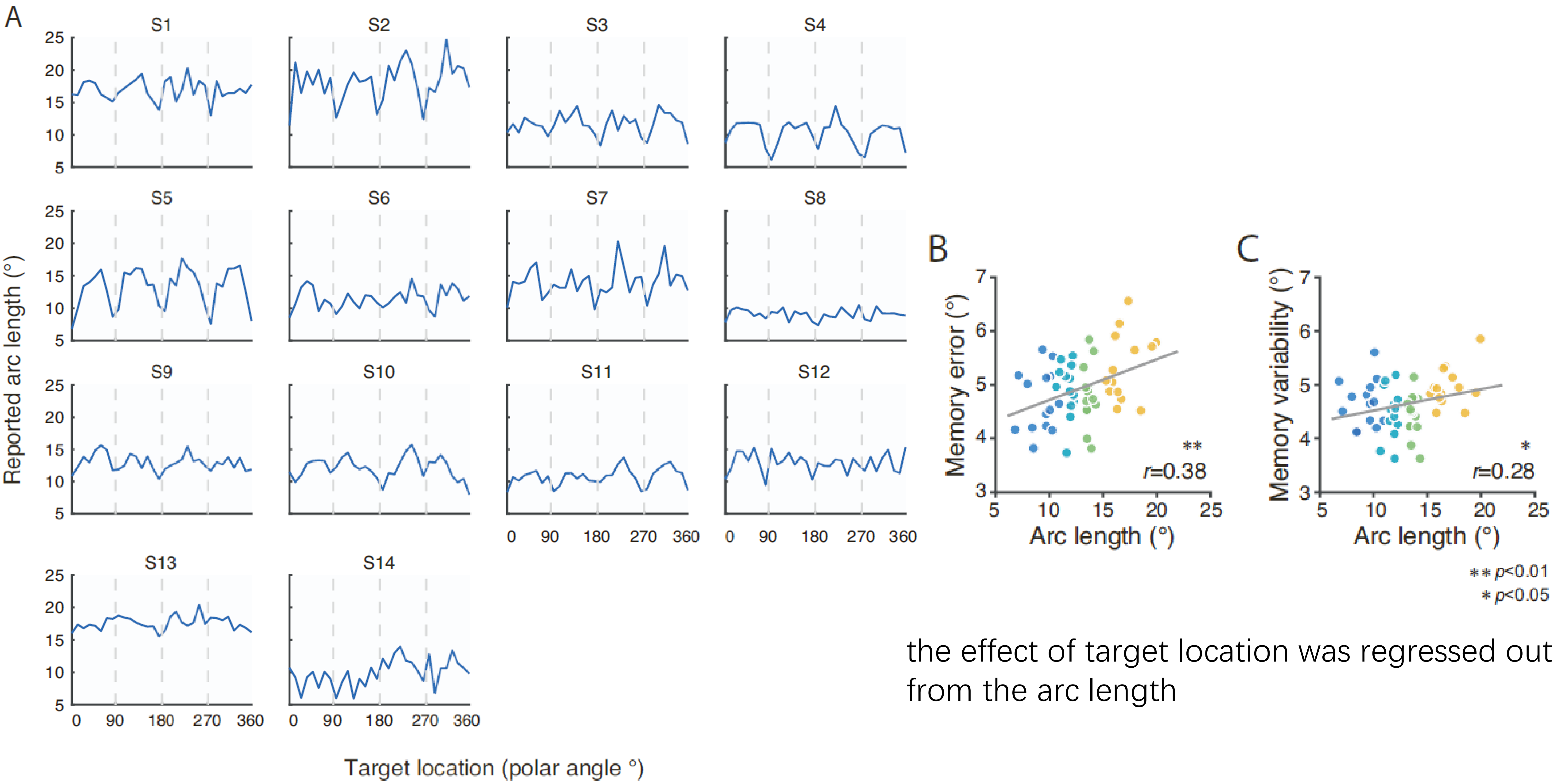
d was the length of the arc in polar angle (°)

- 14 participants reported their memory by making a saccade to the position on the annulus
- Participants were instructed to use the length of the arc to reflect the uncertainty of their memory
- In order to obtain the highest points, an optimal observer would increase the length of the arc with higher VWM uncertainty



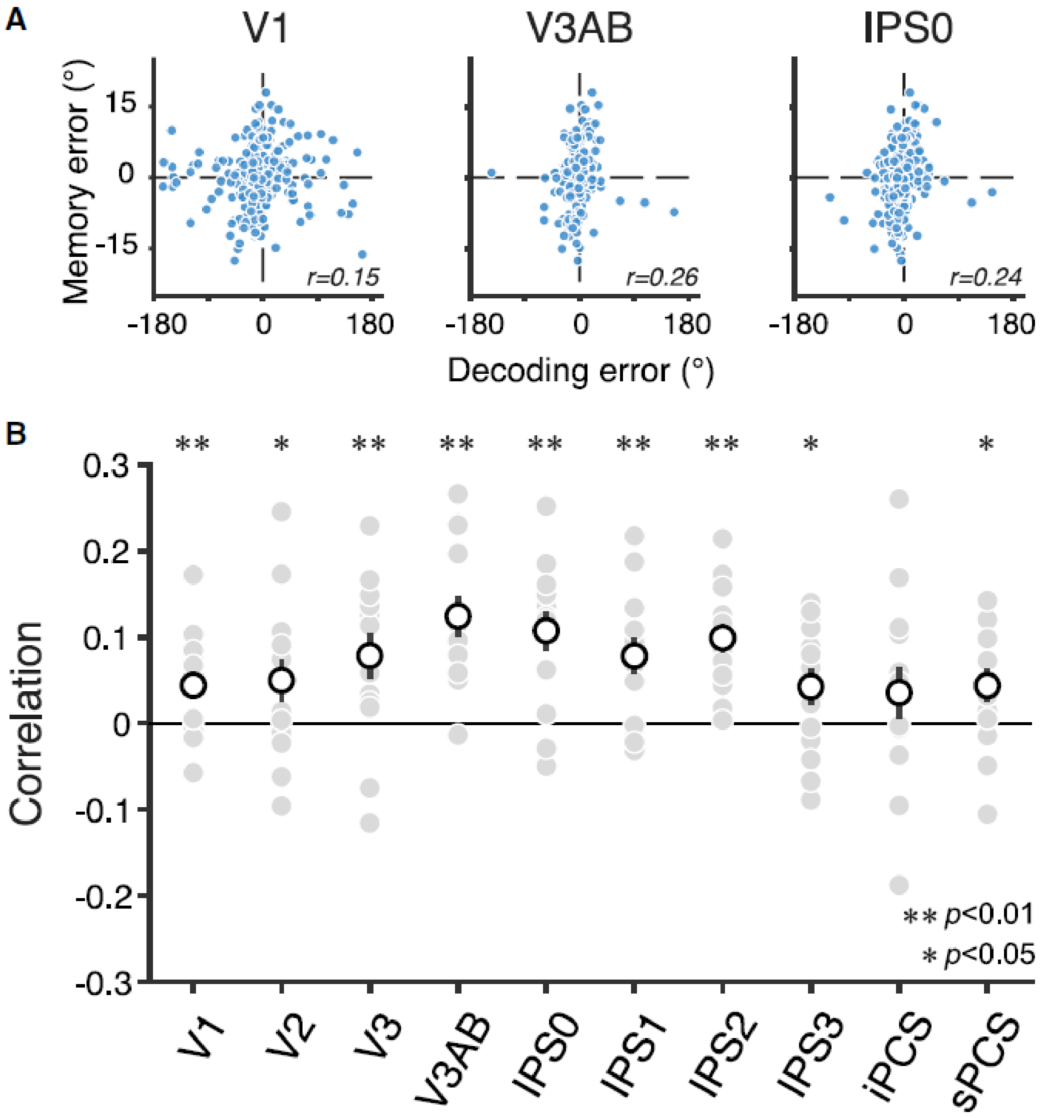
Both the magnitude of memory errors and the variability of memory reports increased with the reported arc length

Arc length as a function of target location for individual participants

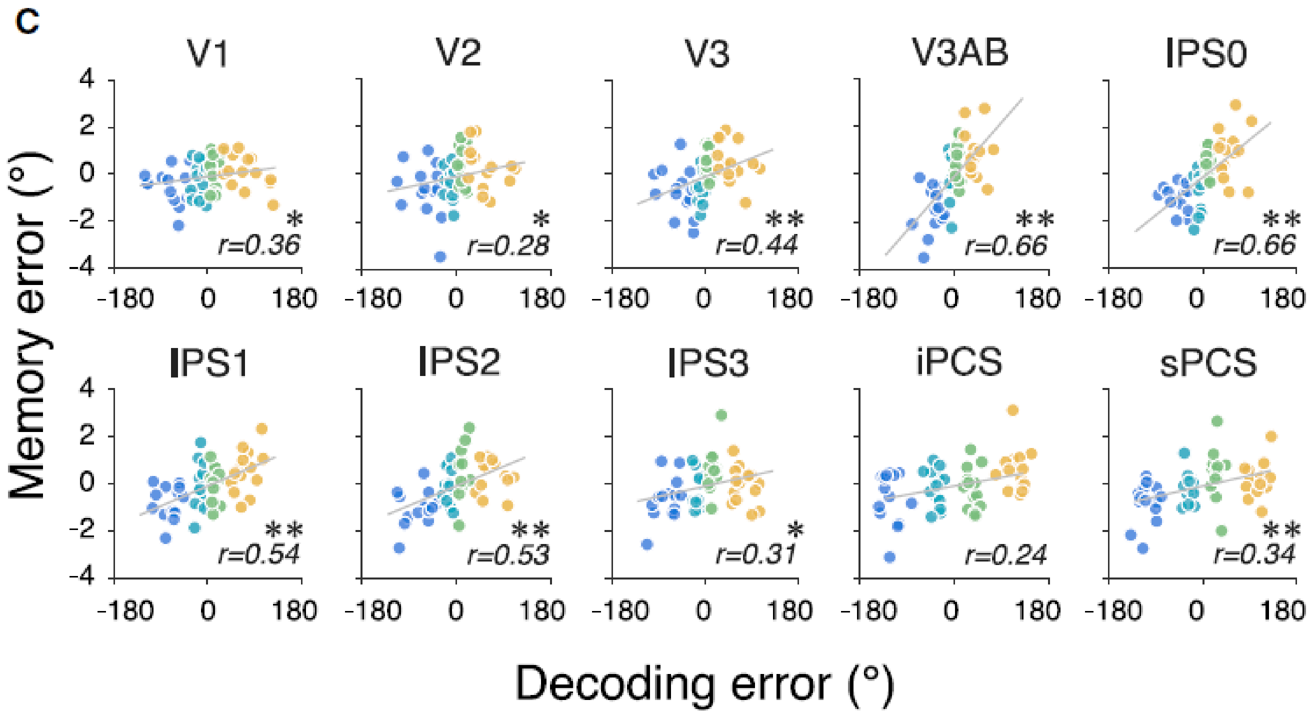


the effect of target location was regressed out from the arc length

Errors in neural decoding of working memory predict behavioral memory errors in experiment 2

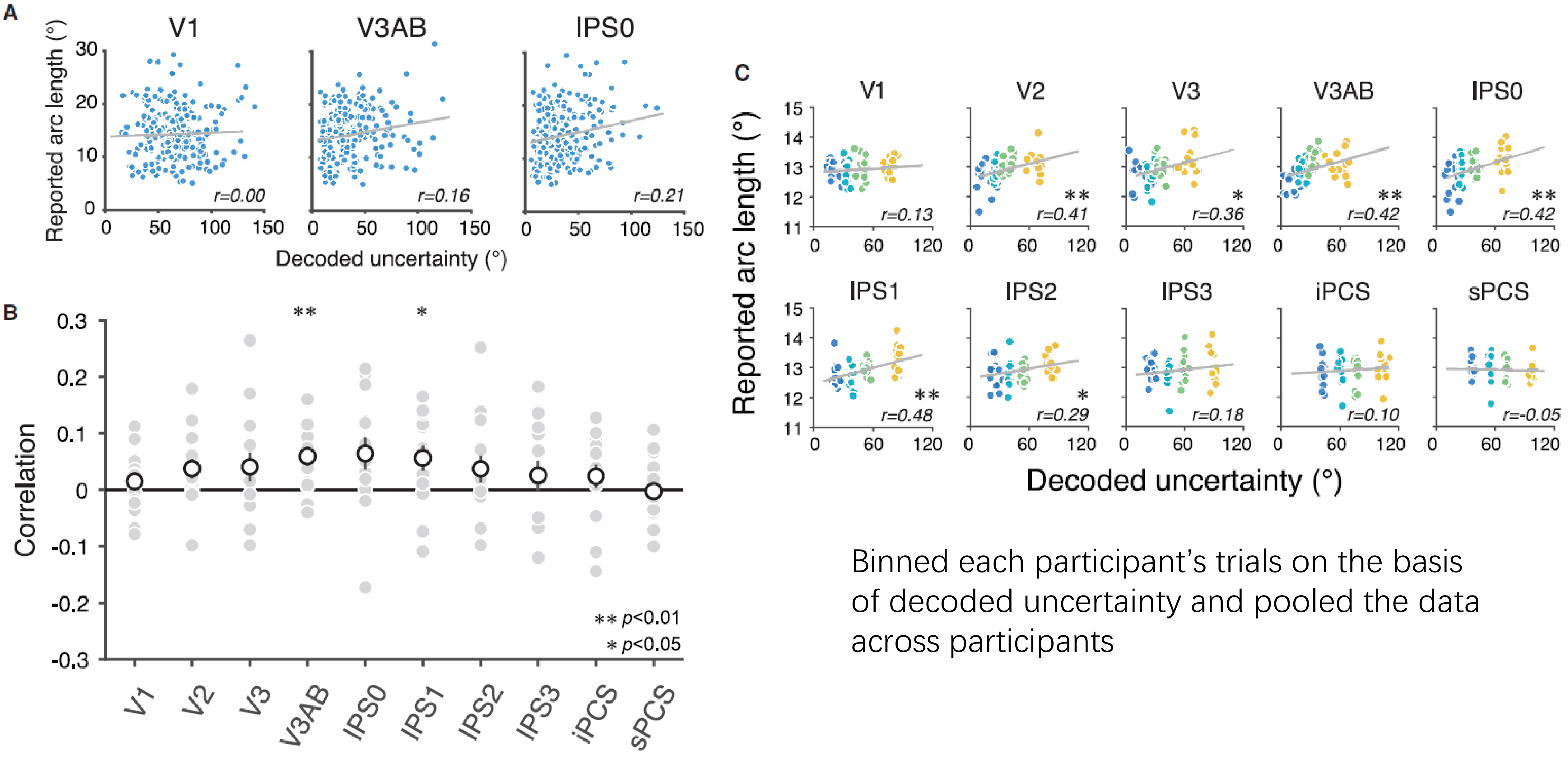


The trial-wise circular correlation between memory error and decoding error for each participant and ROI



binning the trials on the basis of their decoding errors, computing the memory error of each bin, and pooling the data across participants.

Decoded memory uncertainty predicts subjective memory uncertainty

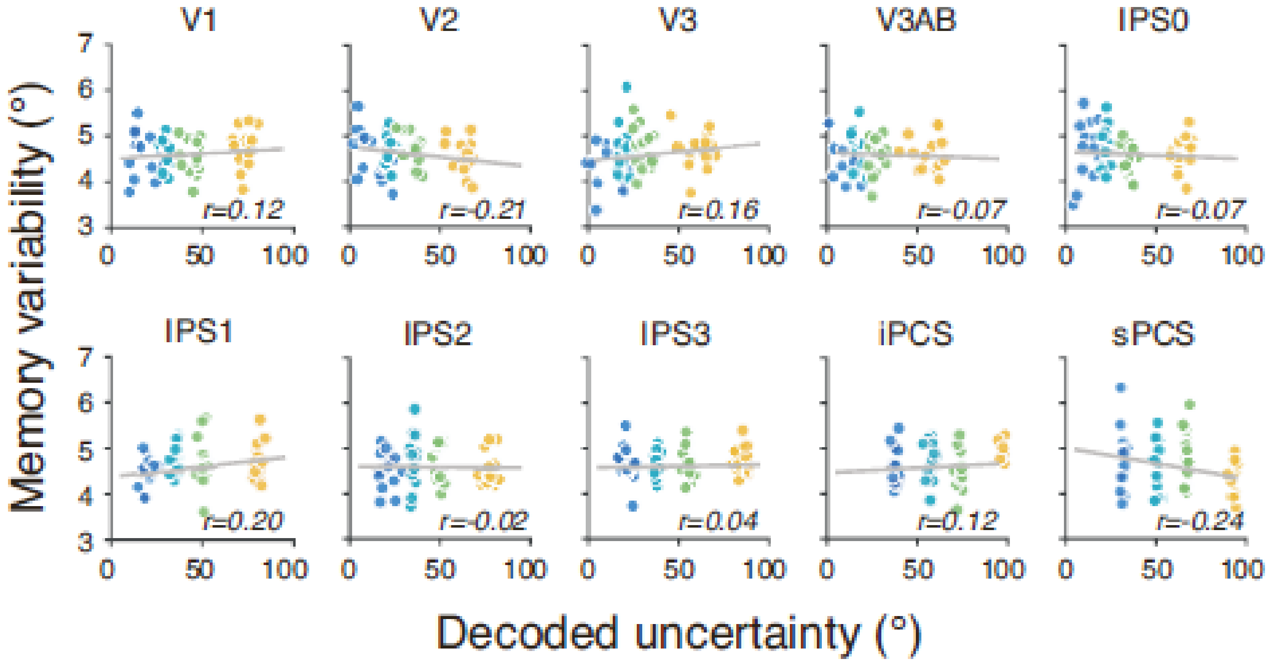


Binned each participant's trials on the basis of decoded uncertainty and pooled the data across participants

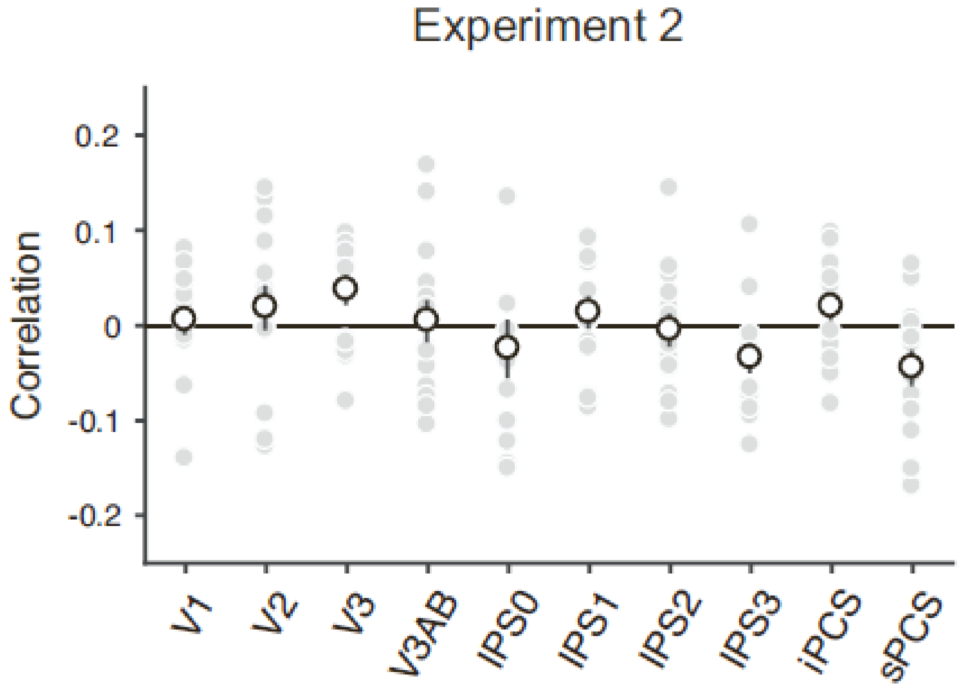
At a single trial level for each participant and each ROI

Correlations between the decoded uncertainty, memory error and memory variability

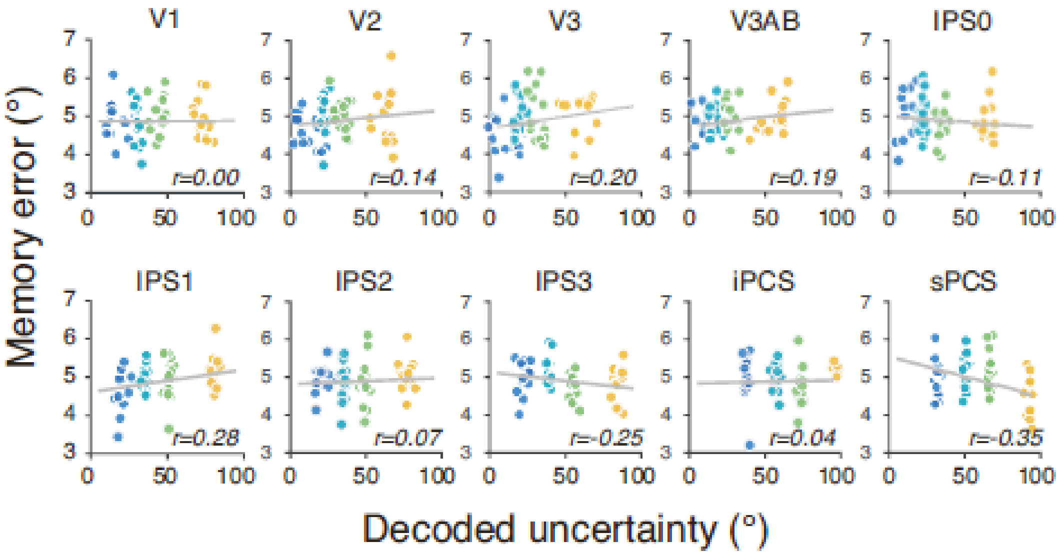
F



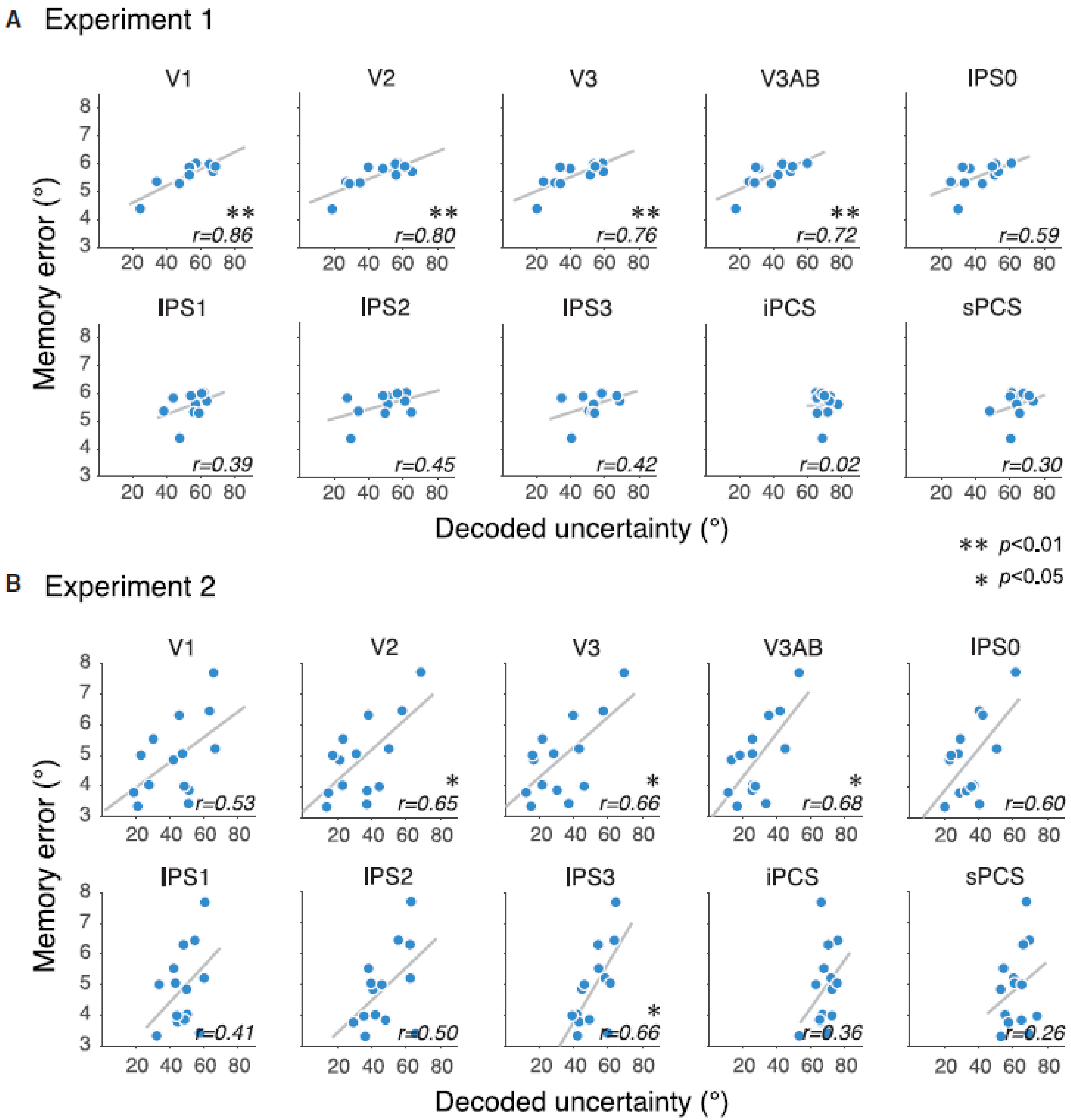
D



E

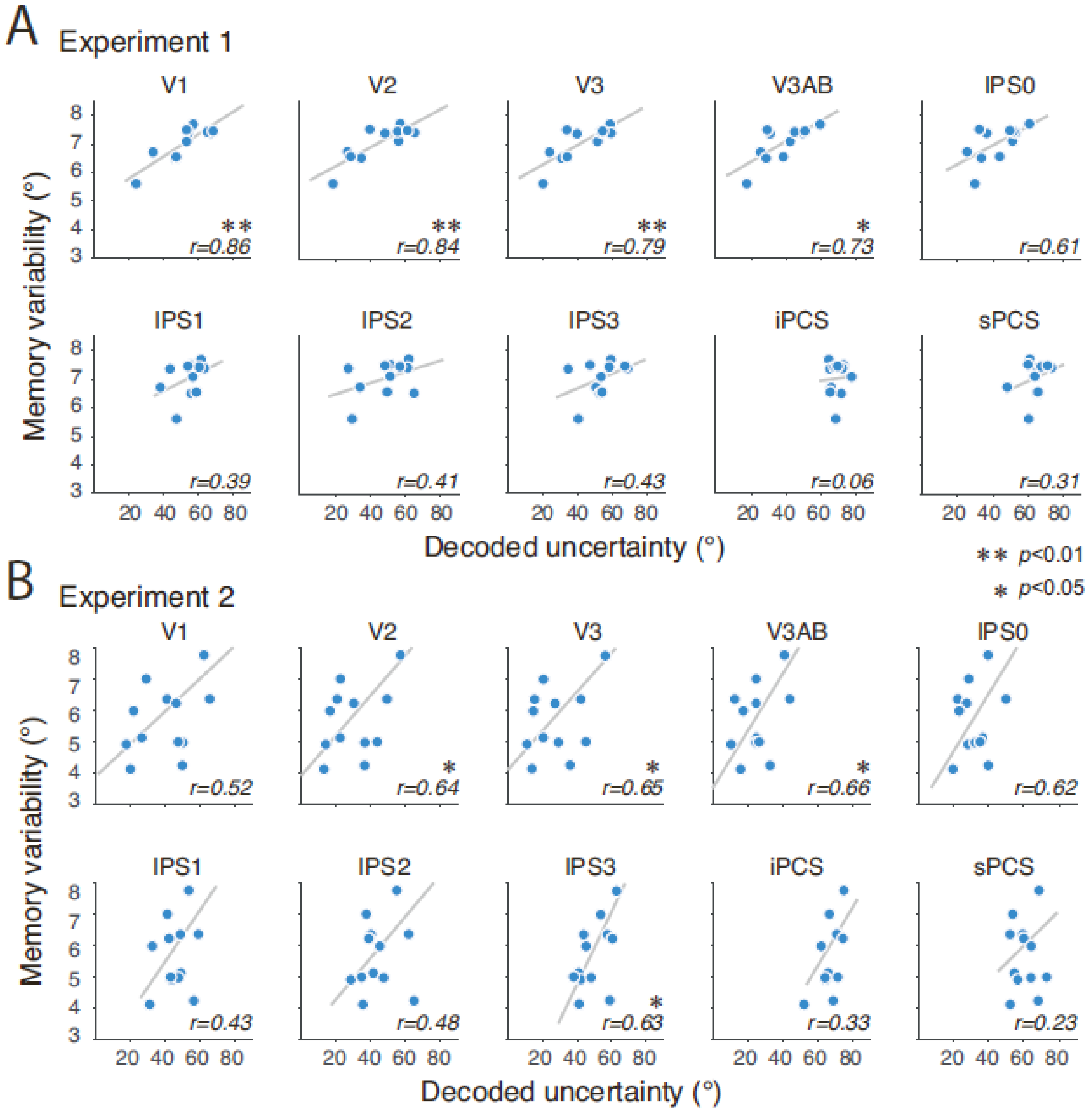


Within the context of spatial VWM, decoded uncertainty did not correlate with the magnitude of memory error or with the variability of memory reports



To investigate whether such relationships between decoded uncertainty and memory errors exist at a cross-subject level, for each participant, we averaged the decoded uncertainty across trials.

Participants with overall greater decoded uncertainty have less precise working memory



To investigate whether such relationships between decoded uncertainty and memory errors exist at a cross-subject level, for each participant, we averaged the decoded uncertainty across trials.

Participants with overall greater decoded uncertainty have less precise working memory

Discussion

- First, we discovered that errors in our neural decoder predicted the direction and amplitude of memory errors made later in the trial.
- Second, we discovered that the uncertainty in our neural decoder predicted the memory uncertainty explicitly reported by participants.
- The content of our WM is a readout of a noisy probability distribution encoded in the population activity of neurons whose distribution width conveys information about memory uncertainty.
- Theoretically, an estimate of an item and the uncertainty of the estimate can be jointly encoded as a single probability distribution by the same population of neurons.