Supplemental Materials for "On testing for spatial correspondence between maps of human brain structure and function"

Aaron Alexander-Bloch[1], Haochang Shou[2], Siyuan Liu[4], Theodore D. Satterthwaite[3], David Glahn[1], Russell T. Shinohara[2], Simon N. Vandekar[2*], and Armin Raznahan[4*]

[1]Yale University School of Medicine, Department of Psychiatry
[2]University of Pennsylvania School of Medicine, Department of Biostatistics, Epidemiology, and Informatics
[3]University of Pennsylvania School of Medicine, Department of Psychiatry
[4]Developmental Neurogenomics Unit, Human Genetics Branch, National Institute of Mental Health, Intramural Program

[*]These authors contributed equally

1 Technical description of spatial permutation test

We use the following notation throughout the supplement. Let $\mathbb{S}^2$ denote the unit sphere in $\mathbb{R}^3$. Let $X(v), Y(w) \in \mathcal{B}(\mathbb{S}^2)$, where $\mathcal{B}(\mathbb{S}^2)$ denotes the space of continuous bounded functions from $\mathbb{S}^2$ to $\mathbb{R}$. $X(v)$ and $Y(w)$ are functions[1] that represent the brain surface maps without information about their alignment. We assume that the anatomical alignment of X and Y is unknown, so that there exists an unknown rotation matrix $R \in \mathrm{SO}(3)$ such that $Y \circ R = Y(Rw) = Y(v)$ is anatomically aligned with $X(v)$. SO(3) denotes the group of all rotations about the origin in 3-dimensions. We also assume the investigator has some measure of association, Ψ, between the two measures X and Y,

$$\Psi(X, Y) : \mathcal{B}(\mathbb{S}^2) \times \mathcal{B}(\mathbb{S}^2) \mapsto \mathbb{R}. \tag{1}$$

The spatial permutation test (SPT) attempts to assess whether two surfaces are associated by producing a p-value that represents the evidence against the null hypothesis that the value of the surface maps do not provide information about their alignment:

$$H_0 : R \mid X(v), Y(w) \sim \mathrm{Unif}\{\mathrm{SO}(3)\}. \tag{2}$$

where Unif{SO(3)} denotes the uniform distribution on SO(3). This hypothesis states that, given the two surfaces without their true alignment, we are no more confident about any particular alignment of the surfaces over any other.

Procedure 1.1 (The SPT)**.** Let P_0 denote the uniform measure on SO(3). For given surface maps X, Y, and association function Ψ in (1), we define $\psi : \mathrm{SO}(3) \to \mathbb{R}$, by $\psi(R; X, Y) = \Psi(X, Y \circ R)$

[1]A goal for future work is to broaden the interpretation of the test to include stochastic processes (as opposed to nonrandom functions), in which case inference about dataset to dataset variability would be possible.

and let R_{obs} denote the nonrandom rotation for the data in their observed alignment. Then the SPT procedure computes the p-value

$$p = \int_{\mathrm{SO}(3)} I(\psi(R) \geq \psi(R_{obs}))dP_0(R), \tag{3}$$

and rejects the null hypothesis that the two surfaces are not associated if $p < \alpha$ for some predetermined rejection threshold $\alpha \in [0, 1]$.

In practice the integral (3) is computed with Monte Carlo simulations. For example, the SPT rotates one of the 3D spherical functions using L random rotations, computes $\psi(R_\ell)$ for $\ell = 1, \ldots, L$, and stores the L values as an estimate the null distribution. For both hemispheres the rotations are symmetric with respect to the y-axis. The following theorem states that the SPT controls the type 1 error rate at the nominal level under the null hypothesis (2).

Theorem 1.1. *Let $P_R(R \in B) = P(R \in B \mid X, Y)$ for B in the Borel σ-algebra on $SO(3)$. Then under the null* (2)*, for all p,* (3) *implies*

$$P_R\{\psi(R) \geq \psi(R_{obs})\} = p.$$

The proof is straightforward because the null hypothesis (2) implies that $P_R\{\psi(R) \geq \psi(R_{obs})\}$ is equal to the integral (3).

The interpretation of the test is critical; because we are comparing the null distribution of ψ to the the value of ψ for the anatomically aligned surfaces. The hypothesis (2) in words is,

H_0 : The observed association of X and Y provides no information about the anatomical alignment of the surfaces.

In other words, the association, measured by ψ, could plausibly have resulted from a random alignment (an alignment determined by a random rotation drawn from uniform distribution defined above). To reject this null hypothesis is to say that it is improbable that the extent to which the surfaces are associated when they are in anatomical alignment (as measured by $\psi(R_{obs})$) could have resulted from a random alignment of the surfaces. In which case, we reject the null hypothesis that there is no correspondence between the surfaces.

2 Simulated data

2.1 Methods

We performed simulations to assess whether the SPT procedure controls the type 1 error rate for maps generated according to the null hypothesis (2) with the actual functional maps generated by Neurosynth (`http://neurosynth.org`). Because we condition on the value of the surfaces, but the orientation is random, these maps were simulated by randomly rotating one surface of each surface pair. We used a subset of 11 unrotated meta-analytic maps (mental imagery, multisensory, naming, object recognition, remembering, sentence comprehension, spatial attention, speech production, stress, visual attention, and word recognition) and a partially overlapping subset of 8 rotated maps (cognitive control, episodic memory, executive function, face recognition, facial expression, mental imagery, multisensory and naming) for the simulated null data. In each simulation, we rotated the subset of 8 maps and computed Pearson's and Spearman's correlations between the rotated and unrotated maps as the correspondence statistic. For each pairs of variables, we performed 1,000 simulations, with the SPT performed using 1,000 permutations within each simulation to compute p-values for the correspondence statistics. We computed the type 1 error rate for a given threshold α for each pair of variables (X, Y), as the proportion of simulations where $p_{X,Y} < \alpha$.

2.2 Results

As expected, the type 1 error rate for Pearson and Spearman SPT of the simulated data was controlled at the nominal α level (see Table S1). At $\alpha = 0.05$, approximately 5% of the randomly generated data resulted in false positives, while at $\alpha = 0.01$, approximately 1% of the randomly generated data resulted in false positives.

Neurosynth Task	**PE05 ± STD**	**SE05 ± STD**	**PE01 ± STD**	**SE01 ± STD**
mental imagery	0.047 ± 0.011	0.049 ± 0.010	0.008 ± 0.002	0.008 ± 0.003
multisensory	0.053 ± 0.011	0.054 ± 0.011	0.010 ± 0.003	0.011 ± 0.005
naming	0.054 ± 0.010	0.055 ± 0.010	0.009 ± 0.003	0.010 ± 0.005
object recognition	0.050 ± 0.009	0.049 ± 0.009	0.010 ± 0.004	0.010 ± 0.003
remembering	0.062 ± 0.006	0.057 ± 0.008	0.011 ± 0.004	0.011 ± 0.004
sentence comprehension	0.051 ± 0.012	0.049 ± 0.013	0.009 ± 0.004	0.011 ± 0.005
spatial attention	0.062 ± 0.014	0.054 ± 0.013	0.011 ± 0.003	0.011 ± 0.004
speech production	0.057 ± 0.009	0.064 ± 0.007	0.009 ± 0.003	0.012 ± 0.004
stress	0.058 ± 0.011	0.059 ± 0.012	0.013 ± 0.004	0.013 ± 0.003
visual attention	0.051 ± 0.016	0.047 ± 0.015	0.009 ± 0.002	0.008 ± 0.003
word recognition	0.048 ± 0.010	0.049 ± 0.010	0.010 ± 0.004	0.011 ± 0.003

Table S1: Spatial permutation test (SPT) type 1 error rates for based on Pearson ("PE") and Spearman ("SE") correlations between 11 maps derived from Neurosynth Tasks and simulated null data, where the simulated data was comprised of randomly aligned maps derived from a partially overlapping supset of Neurosynth Tasks (see simulated data methods above). For nominal α levels of $\alpha = 0.05$ ("PE05" and "SE05") and $\alpha = 0.01$ ("PE01" and "SE01"), tables show the mean and standard deviation (STD) of the false positive rates across the simulated datasets (1,000 simulations, with 1,000 rotational permutations per simulation, within each of the 8 datasets).

3 Comparison with naive parametric approach

To demonstrate the importance of controlling for false positives with a method that takes into consideration the spatial dependencies of the cortical surface, we re-analyzed the correlations between the 120 cognitive terms (i.e., Neurosynth maps) using a commonly-used parametric approach. Here, the test statistic was Pearson's product moment correlation coefficient, assumed to follow a t distribution with n-2 degrees of freedom (in this case n is the number of vertices). In contrast to the 35 correlations that were statistically significant based on the family-wise correction for multiple comparisons based on the SPT (see Figure 2b in main text), even after using a conservative Bonferroni correction for 7,140 multiple comparisons, 5,331 correlations were "significant" at the $p < 0.05$ threshold. Figure S1 illustrates this result, showing a subset of 200 of these false positive connections. This approach does not effectively control for false positives because it assumes that the samples follow independent normal distributions.

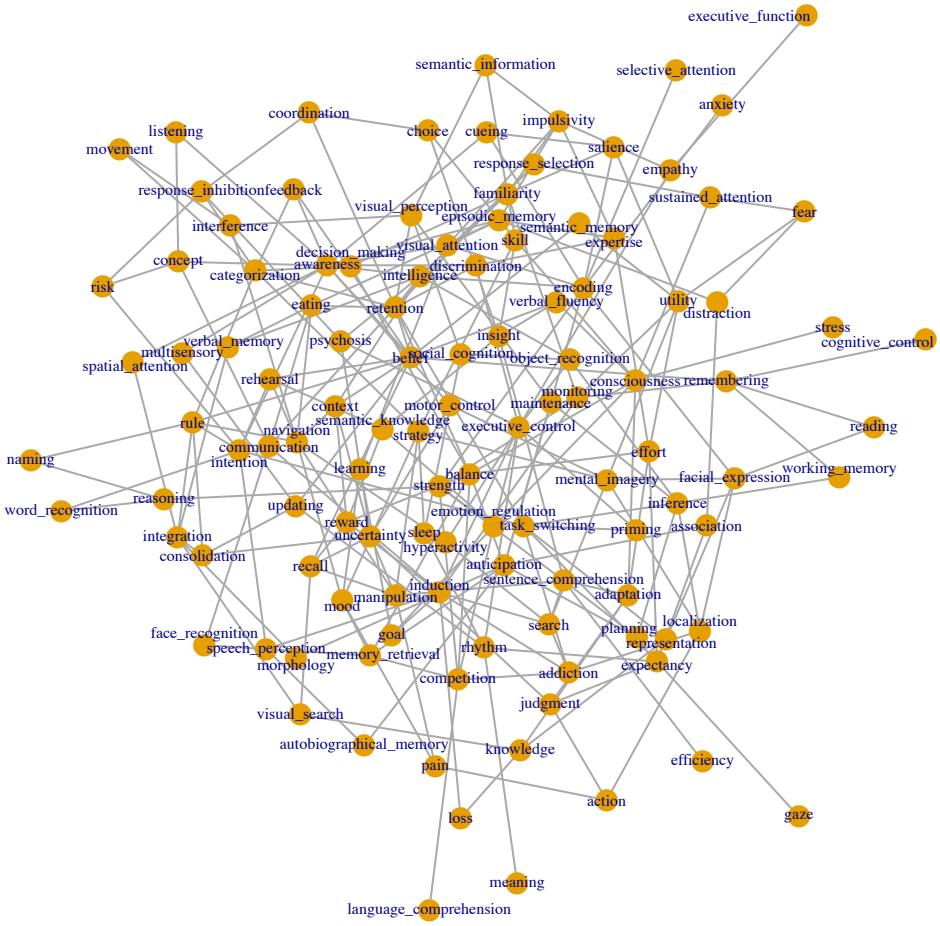

Figure S1: Network illustration where the edges between the terms are illustrated as nodes. In comparison to Figure 2b in the main text, which shows the correlations that are labelled as significant based on the SPT, this plot shows 200 of 5,331 false positive connections which would result if p values are calculated based on a parametric approach which does not incorporate the spatial non-dependence of the cortical maps.

4 Visualization of statistically significant correlations

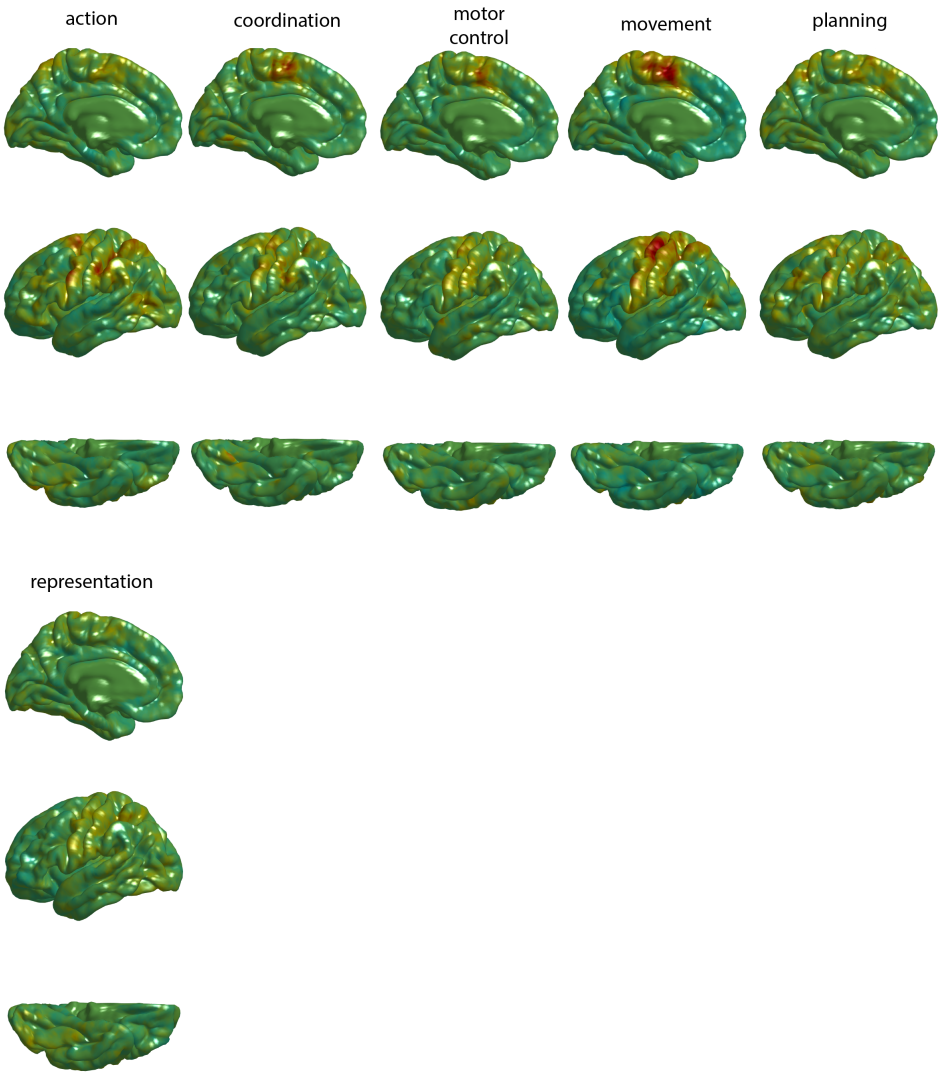

Figure S2: Illustrations of cognitive terms with significant anatomical correspondence with other cognitive terms, as determined by SPT. This plot shows the component of the network illustration (*cf* Figure 2b in the main text) with terms related to the motor system.

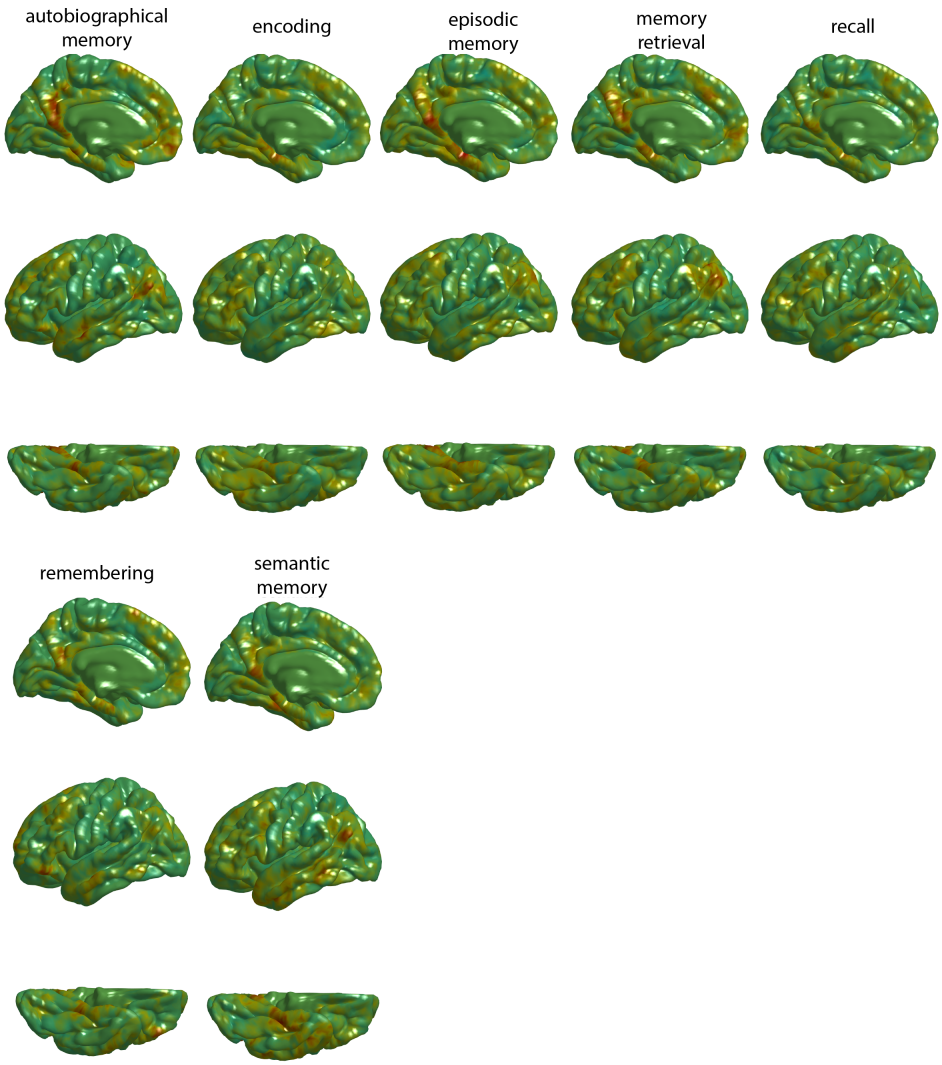

Figure S3: Illustrations of cognitive terms with significant anatomical correspondence with other cognitive terms, as determined by SPT. This plot shows the component of the network illustration (*cf* Figure 2b in the main text) with terms related to memory.

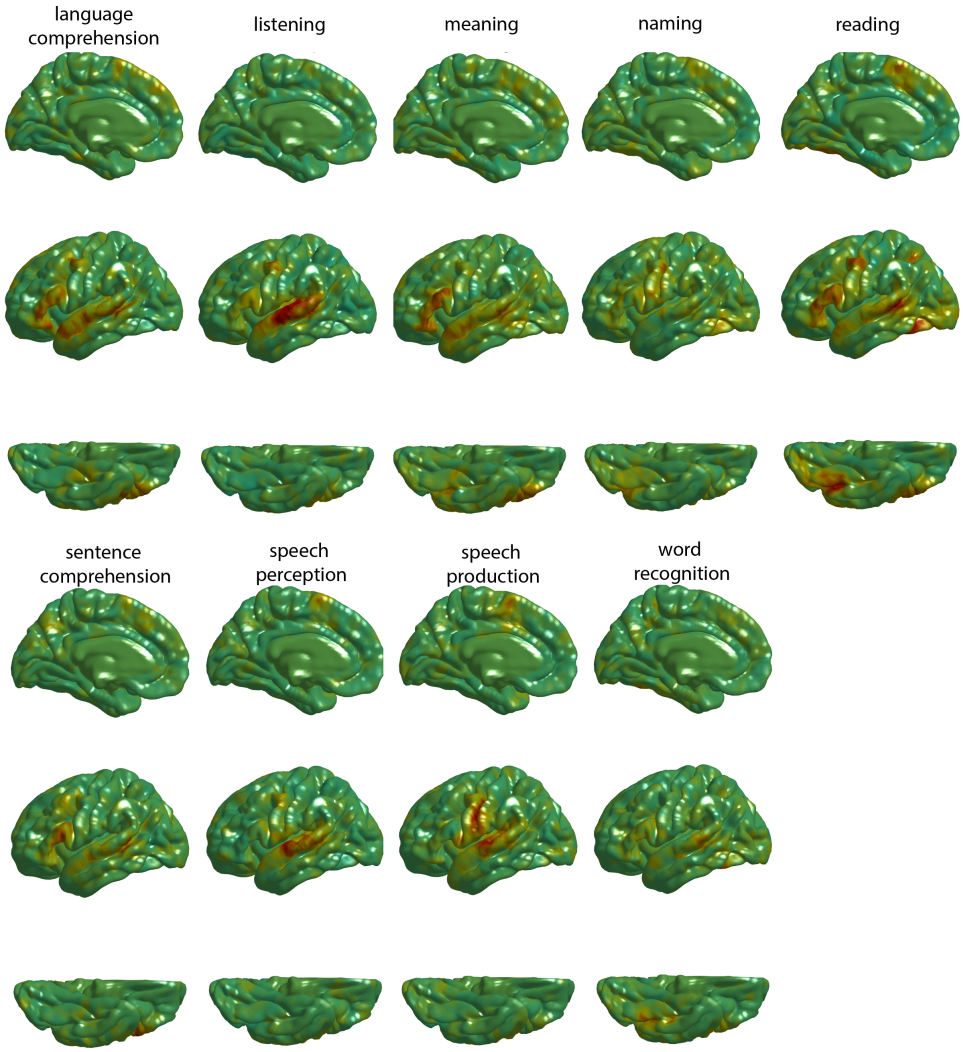

Figure S4: Illustrations of cognitive terms with significant anatomical correspondence with other cognitive terms, as determined by SPT. This plot shows the component of the network illustration (*cf* Figure 2b in the main text) with terms related to language.