

<sup>1</sup> High-level cognition during story listening is reflected in  
<sup>2</sup> high-order dynamic correlations in neural activity patterns

<sup>3</sup> Lucy L. W. Owen<sup>1</sup>, Thomas H. Chang<sup>1,2</sup>, and Jeremy R. Manning<sup>1,†</sup>

<sup>1</sup>Department of Psychological and Brain Sciences,  
Dartmouth College, Hanover, NH

<sup>3</sup>Amazon.com, Seattle, WA

<sup>†</sup>Address correspondence to jeremy.r.manning@dartmouth.edu

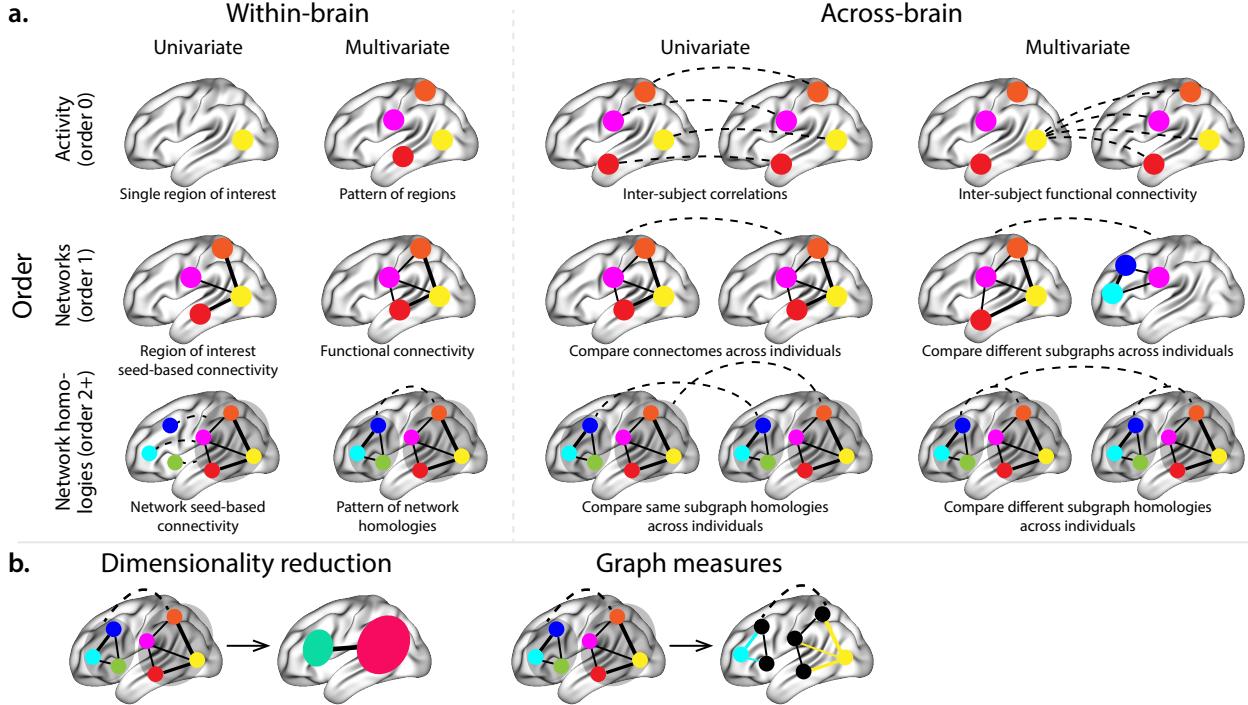
<sup>4</sup> September 5, 2019

<sup>5</sup> **Abstract**

Our thoughts arise from coordinated patterns of interactions between brain structures that change with our ongoing experiences. High-order dynamic correlations in brain activity patterns reflect different subgraphs of the brain’s connectome that display homologous lower-level dynamic correlations. We tested the hypothesis that high-level cognition is supported by high-order dynamic correlations in brain activity patterns. We developed an approach to estimating high-order dynamic correlations in timeseries data, and we applied the approach to neuroimaging data collected as human participants either listened to a ten-minute story, listened to a temporally scrambled version of the story, or underwent a resting state scan. We trained across-participants pattern classifiers to decode (in held-out data) when in the session each neural activity snapshot was collected. We found that classifiers trained to decode from high-order dynamic correlations yielded the best performance on data collected as participants listened to the (unscrambled) story. By contrast, classifiers trained to decode data from scrambled versions of the story or during the resting state scan yielded the best performance when they were trained using first-order dynamic correlations or non-correlational activity patterns. We suggest that as our thoughts become more complex, they are supported by higher-order patterns of dynamic network interactions throughout the brain.

<sup>20</sup> **Introduction**

<sup>21</sup> A central goal in cognitive neuroscience is to elucidate the *neural code*: the mapping between (a) mental  
<sup>22</sup> states or cognitive representations and (b) neural activity patterns. One means of testing models of the  
<sup>23</sup> neural code is to ask how accurately that model is able to “translate” neural activity patterns into known  
<sup>24</sup> (or hypothesized) mental states or cognitive representations (e.g., Haxby et al., 2001; Huth et al., 2016, 2012;  
<sup>25</sup> Kamitani & Tong, 2005; Mitchell et al., 2008; Nishimoto et al., 2011; Norman et al., 2006; Pereira et al.,  
<sup>26</sup> 2018; Tong & Pratte, 2012). Training decoding models on different types of neural features (Fig. 1a) can  
<sup>27</sup> also help to elucidate which specific aspects of neural activity patterns are informative about cognition—  
<sup>28</sup> and, by extension, which types of neural activity patterns might comprise the neural code. For example,  
<sup>29</sup> prior work has used region of interest analyses to estimate the anatomical locations of specific neural  
<sup>30</sup> representations (e.g., Etzel et al., 2009), or to compare the relative contributions to the neural code of



**Figure 1: Neural patterns. a. A space of neural analyses** Within-brain analyses are carried out within a single brain, whereas across-brain analyses compare neural patterns across two or more individuals' brains. Univariate analyses characterize the activities of individual units (e.g., nodes, small networks, hierarchies of networks, etc.), whereas multivariate analyses characterize the patterns of activities across units. Order 0 patterns involve individual nodes; order 1 patterns involve node-node interactions; order 2 (and higher) patterns relate to interactions between homologous networks. Each of these patterns may be static (e.g., averaging over time) or dynamic. **b. Summarizing neural patterns.** To efficiently compute with complex neural patterns, it can be useful to characterize the patterns using summary measures. Dimensionality reduction algorithms project the patterns onto lower-dimensional spaces whose dimensions reflect weighted combinations of the dimensions in the original space. Graph measures characterize each unit's participation in its associated network.

31 multivariate activity patterns versus patterns of dynamic correlations between neural activity patterns (e.g.,  
 32 Fong et al., 2019; Manning et al., 2018). An emerging theme in this literature is that cognition is mediated  
 33 by complex dynamic interactions between brain structures (Bassett et al., 2006; Demertzi et al., 2019; Lurie  
 34 et al., 2018; Mack et al., 2017; Preti et al., 2017; Solomon et al., 2019; Sporns & Honey, 2006; Turk-Browne,  
 35 2013; Zou et al., 2019).

36 Studies of the neural code to date have primarily focused on univariate or multivariate neural pat-  
 37 terns (for review see Norman et al., 2006), or (more recently) on patterns of dynamic first-order correla-  
 38 tions (i.e., interactions between pairs of brain structures; Demertzi et al., 2019; Fong et al., 2019; Lurie et al.,  
 39 2018; Manning et al., 2018; Preti et al., 2017; Zou et al., 2019). We wondered what the future of this line  
 40 of work might hold. For example, is the neural code mediated by higher-order interactions between brain

41 structures (e.g., see Reimann et al., 2017)? Second-order correlations reflect *homologous* patterns of corre-  
42 lation. In other words, if the changing patterns of correlations between two regions,  $A$  and  $B$ , are similar  
43 to those between two other regions,  $C$  and  $D$ , this would be reflected in the second-order correlations be-  
44 tween ( $A-B$ ) and ( $C-D$ ). In this way, second-order correlations identify similarities and differences between  
45 subgraphs of the brain’s connectome. Analogously, third-order correlations reflect homologies between  
46 second-order correlations— i.e., homologous patterns of homologous interactions between brain regions.  
47 More generally, higher-order correlations reflect homologies between patterns of lower-order correlations.  
48 We can then ask: which “orders” of interaction are most reflective of high-level cognitive processes?

49 Another central question pertains to the extent to which the neural code is carried by activity patterns  
50 that directly reflect ongoing cognition (e.g., following Haxby et al., 2001; Norman et al., 2006), versus the  
51 dynamic properties of the network structure itself, independent of specific activity patterns in any given  
52 set of regions (e.g., following Bassett et al., 2006). For example, graph measures such as centrality and  
53 degree (Bullmore & Sporns, 2009) may be used to estimate how a given brain structure is “communicating”  
54 with other structures, independently of the specific neural representations carried by those structures.  
55 If one considers a brain region’s position in the network (e.g., its eigenvector centrality) as a dynamic  
56 property, one can compare how the positions of different regions are correlated, and/or how those patterns  
57 of correlations change over time. We can also compute higher-order patterns in these correlations to  
58 characterize homologous subgraphs in the connectome that display similar changes in their constituent  
59 brain structures’ interactions with the rest of the brain.

60 To gain insights into the above aspects of the neural code, we developed a computational framework  
61 for estimating dynamic high-order correlations in timeseries data. This framework provides an important  
62 advance, in that it enables us to examine patterns in higher-order correlations that are computationally  
63 intractable to estimate via conventional methods. Given a multivariate timeseries, our framework provides  
64 timepoint-by-timepoint estimates of the first-order correlations, second-order correlations, and so on. Our  
65 approach combines a kernel-based method for computing dynamic correlations in timeseries data with a di-  
66 mensionality reduction step (Fig. 1b) that projects the resulting dynamic correlations into a low-dimensional  
67 space. We explored two dimensionality reduction approaches: principle components analysis (PCA; Pear-  
68 son, 1901), which preserves an approximately invertible transformation back to the original data (e.g., this  
69 follows related approaches taken by Gonzalez-Castillo et al., 2019; McIntosh & Jirsa, 2019; Toker & Som-  
70 mer, 2019); and a second non-invertible algorithm that explored patterns in eigenvector centrality (Landau,  
71 1895). This latter approach characterizes correlations between each feature dimension’s relative *position* in  
72 the network in favor of the specific activity histories of different features (also see Betzel et al., 2019; Reimann  
73 et al., 2017; Sizemore et al., 2018).

We validated our approach using synthetic data where the underlying correlations were known. We then applied our framework to a neuroimaging dataset collected as participants listened to either an audio recording of a ten-minute story or a temporally scrambled version of the story, or underwent a resting state scan (Simony et al., 2016). We used a subset of the data to train across-participant classifiers to decode listening times using a blend of neural features (comprising neural activity patterns, as well as different orders of dynamic correlations between those patterns that were inferred using our computational framework). We found that both the PCA-based and eigenvector centrality-based approaches yielded neural patterns that could be used to decode accurately (well above chance). Both approaches also yielded the best decoding accuracy for data collected during (intact) story listening when high-order (PCA: second-order; eigenvector centrality: fourth-order) dynamic correlation patterns were included as features. When we trained classifiers on the scrambled stories or resting state data, only lower-order dynamic patterns were informative to the decoders. Taken together, our results indicate that high-level cognition is supported by high-order dynamic patterns of communication between brain structures.

## Methods

Our general approach to comprises four general steps (Fig. 2). First, we derive a kernel-based approach to computing dynamic pairwise correlations in a  $T$  (timepoints) by  $K$  (features) multivariate timeseries,  $\mathbf{X}_0$ . This yields a  $T$  by  $O(K^2)$  matrix of dynamic correlations,  $\mathbf{Y}_1$ , where each row comprises the upper triangle of the correlation matrix at a single timepoint, reshaped into a row vector (this reshaped vector is  $(\frac{K^2-K}{2})$ -dimensional). Second, we apply a dimensionality reduction step to project the matrix of dynamic correlations back onto a  $K$ -dimensional space. This yields a  $T$  by  $K$  matrix,  $\mathbf{X}_1$ , that reflects an approximation of the dynamic correlations reflected in the original data. Third, we use repeated applications of the kernel-based dynamic correlation step to  $\mathbf{X}_n$  and the dimensionality reduction step to the resulting  $\mathbf{Y}_{n+1}$  to estimate high-order dynamic correlations. Each application of these steps to a  $T$  by  $K$  time series  $\mathbf{X}_n$  yields a  $T$  by  $K$  matrix,  $\mathbf{X}_{n+1}$ , that reflects the dynamic correlations between the columns of  $\mathbf{X}_n$ . In this way, we refer to  $n$  as the *order* of the timeseries, where  $\mathbf{X}_0$  (order 0) denotes the original data and  $\mathbf{X}_n$  denotes (approximated)  $n^{\text{th}}$ -order dynamic correlations between the columns of  $\mathbf{X}_0$ . Finally, we use a cross-validation-based decoding approach to evaluate how well information contained in a given order (or weighted mixture of orders) may be used to decode relevant cognitive states. If including a given  $\mathbf{X}_n$  in the feature set yields higher classification accuracy on held-out data, we interpret this as evidence that the given cognitive states are reflected in patterns of  $n^{\text{th}}$ -order correlations. All of the code used to produce the figures and results in this manuscript, along with links to the corresponding datasets, may be found

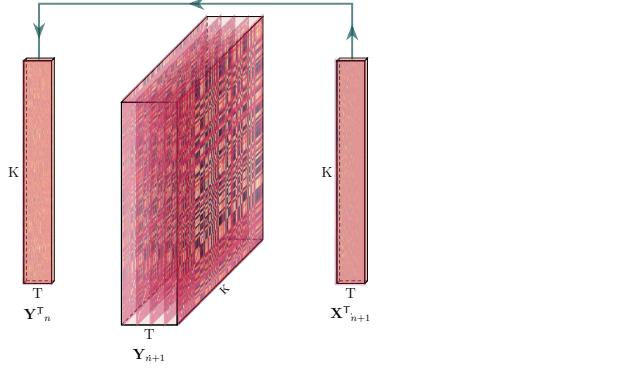


Figure 2: **Estimating dynamic high-order correlations.** Given a  $T$  by  $K$  matrix of multivariate timeseries data,  $\mathbf{X}_n$  (where  $n \in \mathbb{N}, n \geq 0$ ), we use Equation 4 to compute a timeseries of  $K$  by  $K$  correlation matrices,  $\mathbf{Y}_{n+1}$ . We then approximate  $\mathbf{Y}_{n+1}$  with the  $T$  by  $K$  matrix  $\mathbf{X}_{n+1}$ . This process may be repeated to scalably estimate iteratively higher-order correlations in the data. Note that the transposes of  $\mathbf{X}_n$  and  $\mathbf{X}_{n+1}$  are displayed in the figure for compactness.

at [github.com/ContextLab/timecorr-paper](https://github.com/ContextLab/timecorr-paper). In addition, we have released a Python toolbox for computing dynamic high-order correlations in timeseries data; our toolbox may be found at [timecorr.readthedocs.io](https://timecorr.readthedocs.io).

**JRM NOTE: CHECK LINK**

## Kernel-based approach for computing dynamic correlations

Given a  $T$  by  $K$  matrix of observations,  $\mathbf{X}$ , we can compute the (static) Pearson's correlation between any pair of columns,  $\mathbf{X}(\cdot, i)$  and  $\mathbf{X}(\cdot, j)$  using (Pearson, 1901):

$$\text{corr}(\mathbf{X}(\cdot, i), \mathbf{X}(\cdot, j)) = \frac{\sum_{t=1}^T (\bar{\mathbf{X}}(t, i) - \mathbf{X}(t, i)) (\mathbf{X}(t, j) - \bar{\mathbf{X}}(\cdot, j))}{\sqrt{\sum_{t=1}^T \sigma_{\mathbf{X}(\cdot, i)}^2 \sigma_{\mathbf{X}(\cdot, j)}^2}}, \text{ where} \quad (1)$$

$$\bar{\mathbf{X}}(\cdot, k) = \frac{1}{T} \sum_{t=1}^T \mathbf{X}(t, k), \text{ and} \quad (2)$$

$$\sigma_{\mathbf{X}(\cdot, k)}^2 = \frac{1}{T} \sum_{t=1}^T (\mathbf{X}(t, k) - \bar{\mathbf{X}}(\cdot, k))^2 \quad (3)$$

We can generalize this formula to compute time-varying correlations by incorporating a *kernel function* that takes a time  $t$  as input, and returns how much the observed data at each timepoint  $\tau \in [-\infty, \infty]$  contributes to the estimated instantaneous correlation at time  $t$  (Fig. 3; also see Allen et al., 2012, for a similar approach).

Given a kernel function  $\kappa_t(\cdot)$  for timepoint  $t$ , evaluated at timepoints  $\tau \in [1, \dots, T]$ , we can update the

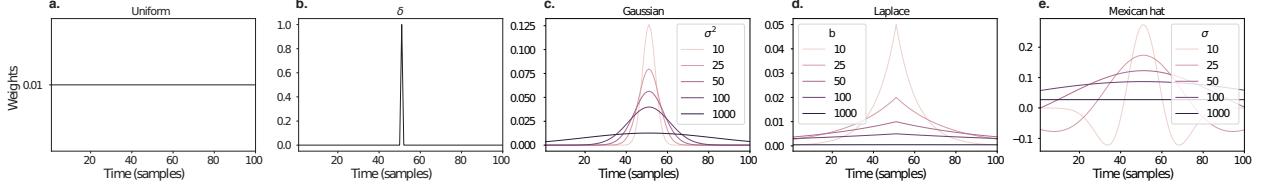


Figure 3: **Examples of kernel functions.** Each panel displays per-timepoint weights at  $t = 50$ , evaluated for 100 timepoints ( $\tau \in [1, \dots, 100]$ ). **a. Uniform kernel.** The weights are timepoint-invariant; observations at all timepoints are weighted equally, and do not change as a function of  $\tau$ . This is a special case of weight function that reduces dynamic correlations to static correlations. **b. Dirac  $\delta$  kernel.** Only the observation at timepoint  $t$  is given a non-zero weight (of 1). **c. Gaussian kernels.** Each kernel's weights fall off in time according to a Gaussian probability density function centered on time  $t$ . Weights derived using several different example width parameters ( $\sigma^2$ ) are displayed. **d. Laplace kernels.** Each kernel's weights fall off in time according to a Laplace probability density function centered on time  $t$ . Weights derived using several different example width parameters ( $b$ ) are displayed. **e. Mexican hat (Ricker wavelet) kernels.** Each kernel's weights fall off in time according to a Ricker wavelet centered on time  $t$ . This function highlights the *contrasts* between local versus surrounding activity patterns in estimating dynamic correlations. Weights derived using several different example width parameters ( $\sigma$ ) are displayed.

static correlation formula in Equation 2 to estimate the *instantaneous correlation* at timepoint  $t$ :

$$\text{timecorr}_{\kappa_t}(\mathbf{X}(\cdot, i), \mathbf{X}(\cdot, j)) = \frac{\sum_{\tau=1}^T (\mathbf{X}(\tau, i) - \tilde{\mathbf{X}}_{\kappa_t}(\cdot, i))(\mathbf{X}(\tau, j) - \tilde{\mathbf{X}}_{\kappa_t}(\cdot, j))}{\sqrt{\sum_{\tau=1}^T \tilde{\sigma}_{\kappa_t}^2(\mathbf{X}(\cdot, i))\tilde{\sigma}_{\kappa_t}^2(\mathbf{X}(\cdot, j))}}, \text{ where} \quad (4)$$

$$\tilde{\mathbf{X}}_{\kappa_t}(t) = \sum_{\tau=1}^T \kappa_t(\tau, k) \mathbf{X}(\tau, k), \quad (5)$$

$$\tilde{\sigma}_{\kappa_t}^2(\mathbf{X}(\cdot, k)) = \sum_{\tau=1}^T (\mathbf{X}(\tau, k) - \tilde{\mathbf{X}}_{\kappa_t}(t, k))^2. \quad (6)$$

113 Here  $\text{timecorr}_{\kappa_t}(\mathbf{X}(\cdot, i), \mathbf{X}(\cdot, j))$  reflects the correlation at time  $t$  between columns  $i$  and  $j$  of  $\mathbf{X}$ , estimated using  
 114 the kernel  $\kappa_t$ . We evaluate Equation 4 in turn each pair of columns in  $\mathbf{X}$  and for kernels centered on each  
 115 timepoint in the timeseries, respectively, to obtain a  $T$  by  $K$  by  $K$  timeseries of dynamic correlations,  $\mathbf{Y}$ . For  
 116 convenience, we then reshape the upper triangles and diagonals of each timepoint's correlation matrix into  
 117 a row vector to obtain an equivalent  $T$  by  $\frac{K^2-K}{2} + K$  matrix.

#### 118 **Dynamic inter-subject functional connectivity (DISFC)**

Equation 4 provides a means of taking a single observation matrix,  $\mathbf{X}_n$  and estimating the dynamic correlations from moment to moment,  $\mathbf{Y}_{n+1}$ . Suppose that one has access to a set of multiple observation matrices that reflect the same phenomenon. For example, one might collect neuroimaging data from several experimental participants, as each participant performs the same task (or sequence of tasks). Let  $\mathbf{X}_n^1, \mathbf{X}_n^2, \dots, \mathbf{X}_n^P$  reflect the  $T$  by  $K$  observation matrices ( $n = 0$ ) or reduced correlation matrices ( $n > 0$ ) for each of

$P$  participants in an experiment. We can use *inter-subject functional connectivity* (ISFC; Simony et al., 2016) to compute the degree of stimulus-driven correlations reflected in the multi-participant dataset at a given timepoint  $t$  using:

$$\bar{\mathbf{C}}(t) = M \left( R \left( \frac{1}{2P} \sum_{p=1}^P Z(Y_n^p(t))^T + Z(Y_n^p(t)) \right) \right), \quad (7)$$

where  $M$  extracts and vectorizes the diagonal and upper triangle of a symmetric matrix,  $Z$  is the Fisher z-transformation (Zar, 2010):

$$Z(r) = \frac{\log(1+r) - \log(1-r)}{2} \quad (8)$$

$R$  is the inverse of  $Z$ :

$$R(z) = \frac{\exp(2z-1)}{\exp(2z+1)}, \quad (9)$$

and  $\mathbf{Y}_n^p(t)$  denotes the correlation matrix (Eqn. 2) between each column of  $\mathbf{X}_n^p$  and each column of the average  $\mathbf{X}_n$  from all *other* participants,  $\bar{\mathbf{X}}_n^{np}$ :

$$\bar{\mathbf{X}}_n^{np} = R \left( \frac{1}{P-1} \sum_{q \in \setminus p} Z(\mathbf{x}_n^q) \right), \quad (10)$$

where  $\setminus p$  denotes the set of all participants other than participant  $p$ . In this way, the  $T$  by  $\frac{K^2-K}{2} + K$  DISFC matrix  $\bar{\mathbf{C}}$  provides a time-varying extension of the ISFC approach developed by Simony et al. (2016).

## 121 Low-dimensional representations of dynamic correlations

Given a  $T$  by  $\frac{K^2-K}{2} + K$  matrix of dynamic correlations,  $\mathbf{Y}_n$ , we propose two general approaches to computing a  $T$  by  $K$  low-dimensional representation of these correlations,  $\mathbf{X}_n$ . The first approach uses dimensionality reduction algorithms to project  $\mathbf{Y}_n$  onto a  $K$ -dimensional space. The second approach uses graph measures to characterize the relative positions of each feature ( $k \in [1, \dots, K]$ ) in the network defined by the correlation matrix at each timepoint.

### 127 Dimensionality reduction-based approaches to computing $\mathbf{X}_n$

The modern toolkit of dimensionality reduction algorithms include Principal Components Analysis (PCA; Pearson, 1901), Probabilistic PCA (PPCA; Tipping & Bishop, 1999), Exploratory Factor Analysis (EFA;

130 Spearman, 1904), Independent Components Analysis (ICA; Comon et al., 1991; Jutten & Herault, 1991),  
131  $t$ -Stochastic Neighbor Embedding ( $t$ -SNE; van der Maaten & Hinton, 2008), Uniform Manifold Approximation and Projection (UMAP; McInnes & Healy, 2018), non-negative matrix factorization (NMF; Lee &  
132 Seung, 1999), Topographic Factor Analysis (TFA) Manning et al. (2014), Hierarchical Topographic Factor  
133 analysis (HTFA) Manning et al. (2018), Topographic Latent Source Analysis (Gershman et al., 2011, TLSA; ),  
134 Dictionary learning (J. Mairal et al., 2009; J. B. Mairal et al., 2009), deep auto-encoders (Hinton & Salakhutdinov, 2006), among others. While complete characterizations of each of these algorithms is beyond the scope  
135 of the present manuscript, the general intuition driving these approaches is to compute the  $T$  by  $M$  matrix,  
136  $\mathbf{X}$ , that is closest to the original  $T$  by  $N$  matrix,  $\mathbf{Y}$ , where (typically)  $M \ll N$ . The different approaches place  
137 different constraints on what properties  $\mathbf{X}$  must satisfy and which aspects of the data are compared (and  
138 how) to characterize the match between  $\mathbf{X}$  and  $\mathbf{Y}$ .

141 Applying dimensionality reduction algorithms to  $\mathbf{Y}$  yields an  $\mathbf{X}$  whose columns reflect weighted combinations  
142 (or nonlinear transformations) of the original columns of  $\mathbf{Y}$ . This has two main consequences. First,  
143 with each repeated dimensionality reduction, the resulting  $\mathbf{X}_n$  has lower and lower fidelity (with respect to  
144 what the “true”  $\mathbf{Y}_n$  might have looked like without using dimensionality reduction to maintain scalability).  
145 In other words, computing  $\mathbf{X}_n$  is a lossy operation. Second, whereas each columns of  $\mathbf{Y}_n$  may always be  
146 mapped directly onto specific pairs of columns of  $\mathbf{Y}_{n-1}$ , the columns of  $\mathbf{X}_n$  reflect weighted combinations  
147 and/or nonlinear transformations of the columns of  $\mathbf{Y}_n$ . Many dimensionality reduction algorithms are  
148 invertible (or approximately invertible). However, attempting to map a given  $\mathbf{X}_n$  back onto the original  
149 feature space of  $\mathbf{Y}_0$  will usually require  $O(TK^{2n})$  space and therefore quickly becomes intractable as  $n$  or  $K$   
150 grow large.

### 151 **Graph measure approaches to computing $\mathbf{X}_n$**

152 The above dimensionality reduction approaches to approximating a given  $\mathbf{Y}_n$  with a lower-dimensional  
153  $\mathbf{X}_n$  preserve a (potentially recombined and transformed) mapping back to the original data in  $\mathbf{Y}_0$ . We also  
154 explore graph measures that instead characterize each feature’s relative *position* in the broader network of  
155 interactions and connections. To illustrate the distinction between the two general approaches we explore,  
156 suppose a network comprises nodes  $A$ ,  $B$ , and  $C$ . If  $A$  and  $B$  exhibit uncorrelated activity patterns, the  
157 functional connection (correlation) between them will be (by definition) close to 0. However, if  $A$  and  $B$   
158 each interact with  $C$  in similar ways, we might attempt to capture those similarities using a measure that  
159 reflects the how  $A$  and  $B$  interact in the network.

160 In general, graph measures take as input a matrix of interactions (e.g., using the above notation, a  $K$

161 by  $K$  correlation matrix or binarized correlation matrix reconstituted from a single timepoint's row of  $\mathbf{Y}$ )  
162 and return as output a set of  $K$  measures describing how each node (feature) sits within that correlation  
163 matrix with respect to the rest of the population. Widely used measures include betweenness centrality (the  
164 proportion of shortest paths between each pair of nodes in the population that involves the given node  
165 in question; e.g., Barthélemy, 2004; Freeman, 1977; Geisberger et al., 2008; Newman, 2005; Opsahl et al.,  
166 2010); diversity and dissimilarity (characterizations of how differently connected a given node is from others  
167 in the population; e.g., Lin, 2009; Rao, 1982; Ricotta & Szeidl, 2006); Eigenvector centrality and pagerank  
168 centrality (measures of how influential a given node is within the broader network; e.g., Bonacich, 2007;  
169 Halu et al., 2013; Lohmann et al., 2010; Newman, 2008); transfer entropy and flow coefficients (a measure  
170 of how much information is flowing from a given node to other nodes in the network; e.g., Honey et  
171 al., 2007; Schreiber, 2000);  $k$ -coreness centrality (a measure of the connectivity of a node within its local  
172 sub-graph; e.g., Alvarez-Hamelin et al., 2005; Christakis & Fowler, 2010); within-module degree (a measure  
173 of how many connections a node has to its close neighbors in the network; e.g., Rubinov & Sporns, 2010);  
174 participation coefficient (a measure of the diversity of a node's connections to different sub-graphs in the  
175 network; e.g., Rubinov & Sporns, 2010); and sub-graph centrality (a measure of a node's participation in all  
176 of the network's sub-graphs; e.g., Estrada & Rodríguez-Velázquez, 2005); among others.

177 For a given graph measure,  $\eta : \mathbb{R}^{K \times K} \rightarrow \mathbb{R}^K$ , we can use  $\eta$  to transform each row of  $\mathbf{Y}_n$  in a way that  
178 characterizes the corresponding graph properties of each column. This results in a new  $T$  by  $K$  matrix,  $\mathbf{X}_n$ ,  
179 that reflects how the features reflected in the columns of  $\mathbf{Y}_n$  participate in the network during each timepoint  
180 (row).

## 181 Dynamic higher-order correlations

182 Because  $\mathbf{X}_n$  has the same shape as the original data  $\mathbf{X}_0$ , approximating  $\mathbf{Y}_n$  with a lower-dimensional  $\mathbf{X}_n$   
183 enables us to estimate high-order dynamic correlations in a scalable way. Given a  $T$  by  $K$  input matrix, the  
184 output of Equation 4 requires  $O(TK^2)$  space to store. Repeated applications of Equation 4 (i.e., computing  
185 dynamic correlations between the columns of the outputted dynamic correlation matrix) each require  
186 exponentially more space; in general the  $n^{\text{th}}$ -order dynamic correlations of a  $T$  by  $K$  timeseries occupies  
187  $O(TK^{2n})$  space. However, when we approximate or summarize the output of Equation 4 with a  $T$  by  $K$  matrix  
188 (as described above), it becomes feasible to compute even very high-order correlations in high-dimensional  
189 data. Specifically, approximating the  $n^{\text{th}}$ -order dynamic correlations of a  $T$  by  $K$  timeseries requires only  
190  $O(TK^2)$  additional space— the same as would be required to compute first-order dynamic correlations. In  
191 other words, the space required to store  $n + 1$  multivariate timeseries reflecting up to  $n^{\text{th}}$  order correlations

192 in the original data scales linearly with  $n$  using our approach (Fig. 2).

## 193 Data

194 We examined two types of data: synthetic data and human functional neuroimaging data. We constructed  
195 and leveraged the synthetic data to evaluate our general approach (for a related approach, see Thompson et  
196 al., 2018). Specifically, we tested how well Equation 4 could be used to recover known dynamic correlations  
197 using different choices of kernel ( $\kappa$ ; Fig. 3), for each of several synthetic datasets that exhibited different tem-  
198 poral properties. We applied our approach to a functional neuroimaging dataset to test the hypothesis that  
199 ongoing cognitive processing is reflected in high-order dynamic correlations. We used an across-participant  
200 classification test to estimate whether dynamic correlations of different orders contain information about  
201 which timepoint in a story participants were listening to.

### 202 Synthetic data

203 We constructed a total of 40 multivariate timeseries, collectively reflecting a total of 4 different patterns of  
204 dynamic correlations (i.e., 10 datasets reflecting each type of dynamic pattern). Each timeseries comprised  
205 50 features (dimensions) that varied over 300 timepoints. The observations at each timepoint were drawn  
206 from a zero-mean multivariate Gaussian distribution with a covariance matrix defined for each timepoint  
207 as described below. We drew the observations at each timepoint independently from the draws at all other  
208 timepoints; in other words, for each observation  $s_t \sim \mathcal{N}(\mathbf{0}, \Sigma_t)$  at timepoint  $t$ ,  $p(s_t) = p(s_t|p_{\setminus t})$ .

**Constant.** We generated data with stable underlying correlations to evaluate how Equation 4 characterized correlation “dynamics” when the ground truth correlations were static. We constructed 10 multivariate timeseries, whose observations were each drawn from a single (stable) Gaussian distribution. For each dataset, we constructed a random covariance matrix,  $\Sigma_m$ :

$$\mathbf{C}(i, j) \sim \mathcal{N}(0, 1) \quad (11)$$

$$\Sigma_m = \mathbf{C}\mathbf{C}^\top, \text{ where} \quad (12)$$

209  $i, j \in [1, 2, \dots, 50]$ . In other words, all of the observations (for each of the 300 timepoints) within each dataset  
210 were drawn from a multivariate Gaussian distribution with the same covariance matrix, and the 10 datasets  
211 each used a different covariance matrix.

212 **Random.** We generated a second set of 10 synthetic datasets whose observations at each timepoint were  
213 drawn from a Gaussian distribution with a new randomly constructed (using Eqn. 12) covariance matrix.  
214 Because each timepoint’s covariance matrix was drawn independently of the covariance matrices for all  
215 other timepoints, these datasets provided a test of reconstruction accuracy in the absence of any meaningful  
216 underlying temporal structure in the dynamic correlations underlying the data.

**Ramping.** We generated a third set of 10 synthetic datasets whose underlying correlations changed gradually over time. For each dataset, we constructed two *anchor* correlation matrices using Equation 12,  $\Sigma_{\text{start}}$  and  $\Sigma_{\text{end}}$ . For each of the 300 timepoints in each dataset, we drew the observations from a multivariate Gaussian distribution whose covariance matrix at each timepoint  $t \in [0, \dots, 299]$  was given by

$$\Sigma_t = \left(1 - \frac{1-t}{299}\right)\Sigma_{\text{start}} + \frac{t}{299}\Sigma_{\text{end}}. \quad (13)$$

217 The gradually changing correlations underlying these datasets allow us to evaluate the recovery of dynamic  
218 correlations when each timepoint’s correlation matrix is unique (as in the random datasets), but where the  
219 correlation dynamics are structured.

220 **Event.** We generated a fourth set of 10 synthetic datasets whose underlying correlation matrices exhibited  
221 prolonged intervals of stability, interspersed with abrupt changes. For each dataset, we used Equation 12  
222 to generate 5 random covariance matrices. We constructed a timeseries where each set of 60 consecutive  
223 samples was drawn from a Gaussian with the same covariance matrix. These datasets were intended to  
224 simulate a system that undergoes occasional abrupt state changes.

## 225 Functional neuroimaging data collected during story listening

226 We examined an fMRI dataset collected by Simony et al. (2016) that the authors have made publicly available  
227 at [arks.princeton.edu/ark:/88435/dsp015d86p269k](http://arks.princeton.edu/ark:/88435/dsp015d86p269k). The dataset comprises neuroimaging data collected as  
228 participants listened to an audio recording of a story (intact condition; 36 participants), listened to time  
229 scrambled recordings of the same story (17 participants in the paragraph-scrambled condition listened to  
230 the paragraphs in a randomized order and 36 in the word-scrambled condition listened to the words in a  
231 randomized order), or lay resting with their eyes open in the scanner (rest condition; 36 participants). Full  
232 neuroimaging details may be found in the original paper for which the data were collected (Simony et al.,  
233 2016).

234 **Hierarchical topographic factor analysis (HTFA).** Following our prior related work, we used HTFA (Manning et al., 2018) to derive a compact representation of the data. In brief, this approach approximates the  
235 timeseries of voxel activations (44,415 voxels) using a much smaller number of radial basis function (RBF)  
236 nodes (in this case 700 nodes). This provides a convenient representation for examining full-brain network  
237 dynamics. All of the analyses we carried out on the neuroimaging dataset were performed in this lower-  
238 dimensional space. In other words, each participant's data matrix,  $\mathbf{Y}_0$ , was a number-of-timepoints by 700  
239 matrix of HTFA-derived factor weights (where the row and column labels were matched across partici-  
240 pants). Code for carrying out HTFA on fMRI data may be found as part of the BrainIAK toolbox (Capota et  
241 al., 2017), which may be downloaded at [brainiak.org](http://brainiak.org).

## 243 **Temporal decoding**

244 We sought to identify neural patterns that reflected participants' ongoing cognitive processing of incoming  
245 stimulus information. As reviewed by Simony et al. (2016), one way of homing in on these stimulus-driven  
246 neural patterns is to compare activity patterns across individuals (e.g., using ISFC analyses). In particular,  
247 neural patterns will be similar across individuals, to the extent that the neural patterns under consideration  
248 are stimulus driven, and to the extent that the corresponding cognitive representations are reflected in similar  
249 spatial patterns across people. Following this logic, we used an across-participants temporal decoding test  
250 developed by Manning et al. (2018) to assess the degree to which different neural patterns reflected ongoing  
251 stimulus-driven cognitive processing across people. The approach entails using a subset of the data to  
252 train a classifier to decode which stimulus timepoint (i.e., moment in the story participants listened to). We  
253 use decoding (forward inference) accuracy on held-out data, from held-out participants, as a proxy for the  
254 extent to which the inputted neural patterns reflected stimulus-driven cognitive processing in a similar way  
255 across individuals.

## 256 **Forward inference and decoding accuracy**

257 We used an across-participants correlation-based classifier to decode which stimulus timepoint matched a  
258 given neural pattern. We first divided the participants into two groups: a template group,  $\mathcal{G}_{\text{template}}$ , and a  
259 to-be-decoded group,  $\mathcal{G}_{\text{decode}}$ . We used Equation 7 to compute a DISFC matrix for each group ( $\bar{\mathbf{C}}_{\text{template}}$  and  
260  $\bar{\mathbf{C}}_{\text{decode}}$ , respectively). We then correlated the rows of  $\bar{\mathbf{C}}_{\text{template}}$  and  $\bar{\mathbf{C}}_{\text{decode}}$  to form a number-of-timepoints by  
261 number-of-timepoints decoding matrix,  $\Lambda$ . In this way, the rows of  $\Lambda$  reflected timepoints from the template  
262 group, while the columns reflected timepoints from the to-be-decoded group. We assigned temporal labels  
263 to each row  $\bar{\mathbf{C}}_{\text{decode}}$  using the row of  $\bar{\mathbf{C}}_{\text{template}}$  to which it was most highly correlated. We then repeated

264 this decoding procedure, but using  $\mathcal{G}_{\text{decode}}$  as the template group and  $\mathcal{G}_{\text{template}}$  as the to-be-decoded group.  
265 Given the true timepoint labels (for each group), we defined the *decoding accuracy* as the proportion of  
266 correctly decoded timepoints, across both groups.

267 **Feature weighting and testing**

268 We sought to examine which types of neural features (i.e., activations, first-order dynamic correlations, and  
269 higher-order dynamic correlations) were informative to the temporal decoders. Using the notation above,  
270 these features correspond to  $\mathbf{Y}_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ , and so on (we examined up to tenth order correlations, or  $\mathbf{X}_{10}$ ).

271 One challenge to fairly evaluating high-order correlations is that if the kernel used in Equation 4 is  
272 wider than a single timepoint, each repeated application of the equation will result in further temporal  
273 blur. Because our primary assessment metric is temporal decoding accuracy, this unfairly biases against  
274 detecting meaningful signal in higher-order correlations (relative to lower-order correlations). We attempted  
275 to mitigate temporal blur in estimating each  $\mathbf{X}_n$  by using a Dirac  $\delta$  function kernel (which places all of its  
276 mass over a single timepoint; Fig. 3b) to compute each lower-order correlation ( $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{n-1}$ ). We then used  
277 a (potentially wider, as described below) kernel to compute  $\mathbf{X}_n$  from  $\mathbf{X}_{n-1}$ . In this way, temporal blurring  
278 was applied only in the last step of computing  $\mathbf{X}_n$ . We note that, because each  $\mathbf{X}_n$  is a low-dimensional  
279 representation of the corresponding  $\mathbf{Y}_n$ , the higher-order correlations we estimated reflect true correlations  
280 in the data with lower-fidelity than estimates of lower-order correlations.

281 After computing each  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{n-1}$  for each participant, we divided participants into two equally sized  
282 groups ( $\pm 1$  for odd numbers of participants):  $\mathcal{G}_{\text{train}}$  and  $\mathcal{G}_{\text{test}}$ . We then further subdivided  $\mathcal{G}_{\text{train}}$  into  $\mathcal{G}_{\text{train}_1}$   
283 and  $\mathcal{G}_{\text{train}_2}$ . We then computed  $\Lambda$  temporal correlation matrices for each type of neural feature, using  $\mathcal{G}_{\text{train}_1}$   
284 and  $\mathcal{G}_{\text{train}_2}$ . This resulted in  $n + 1$   $\Lambda$  matrices (one for the original timeseries of neural activations, and one  
285 for each of  $n$  orders of dynamic correlations). Our objective was to find a set of weights of each of these  $\Lambda$   
286 matrices such that the weighted average of the  $n + 1$  matrices yielded the highest decoding accuracy. We  
287 used quasi-Newton gradient ascent (Nocedal & Wright, 2006), using decoding accuracy as the objective  
288 function to be maximized, to find an optimal set of training data-derived weights,  $\phi_{0,1,\dots,n}$ , where  $\sum_{i=0}^n \phi_i = 1$   
289 and where  $\phi_i \geq 0 \forall i \in [0, 1, \dots, n]$ .

290 After estimating an optimal set of weights, we computed a new set of  $n + 1$   $\Lambda$  matrices correlating the  
291 DISFC patterns from  $\mathcal{G}_{\text{train}}$  and  $\mathcal{G}_{\text{test}}$  at each timepoint. We use the resulting decoding accuracy of  $\mathcal{G}_{\text{test}}$   
292 timepoints to estimate how informative the set up neural features containing up to  $n^{\text{th}}$  order correlations  
293 were.

294 We used a permutation-based procedure to form a stable estimate of decoding accuracy for each set of

295 neural features. In particular, we computed the decoding accuracy for each of 10 random group assignments  
296 of  $\mathcal{G}_{\text{train}}$  and  $\mathcal{G}_{\text{test}}$ . We report the mean accuracy (and 95% confidence intervals) for each set of neural features.

297 **Identifying robust decoding results**

298 The temporal decoding procedure we use to estimate which neural features support ongoing cognitive  
299 processing is governed by many parameters. For example, Equation 4 requires defining a kernel function,  
300 which can take on different shapes and widths. For a fixed set of neural features, each of these parameters  
301 can yield different decoding accuracies. Further, the best decoding accuracy for a given timepoint may be  
302 reliably achieved by one set of parameters, whereas the best decoding accuracy for another timepoint might  
303 be reliably achieved by a different set of parameters, and the best decoding accuracy across *all* timepoints  
304 might be reliably achieved by still another different set of parameters. Rather than attempting to maximize  
305 decoding accuracy, we sought to discover the trends in the data that were robust to specific classifier  
306 parameters choices. Specifically, we sought to characterize how decoding accuracy varied (under different  
307 experimental conditions) as a function of which neural features were considered.

308 To identify decoding results that were robust to specific classifier parameter choices, we repeated our  
309 decoding analyses that substituted in a variety of kernel shapes and widths for Equation 4. We examined  
310 Gaussian (Fig. 3c), Laplace (Fig. 3d), and Mexican Hat (Fig. 3e) kernels, each with widths of 5, 10, 20, and  
311 50 samples. We then report the average decoding accuracies across all of these parameter choices. This  
312 enabled us to (roughly) factor out performance characteristics that were parameter dependent (within the  
313 space of parameters we examined).

314 **Reverse inference**

315 The dynamic patterns we examine comprise high-dimensional correlation patterns at each timepoint. To  
316 help interpret the resulting patterns in the context of other studies, we created summary maps by computing  
317 the across-timepoint average pairwise correlations at each order of analysis (first order, second order, etc.,  
318 up to fifteenth order correlations). We selected the 10 strongest (absolute value) correlations at each order.  
319 Each correlation is between the dynamic activity patterns (or patterns of dynamic high-order correlations)  
320 measured at two RBF nodes (see *Hierarchical Topographic Factor Analysis*). Therefore, the 10 strongest  
321 correlations involved up to 20 RBF nodes. Each RBF defines a spatial function whose activations range  
322 from 0 to 1. We thresholded each RBF at 0.999 to construct a map of spherical components that denoted the  
323 endpoints of the 10 strongest correlations. We then carried out a meta analysis using Neurosynth (Rubin et  
324 al., 2017) to identify the 10 terms most commonly associated with the given map. This resulted in a set of

325 10 terms associated with the average dynamic correlation patterns at each order.

## 326 Results

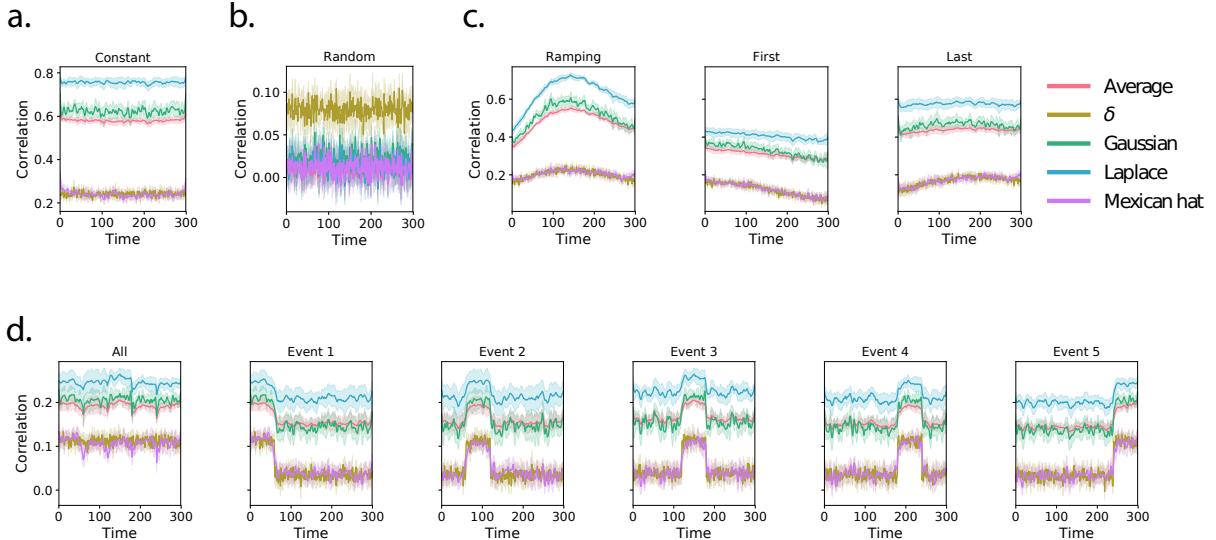
327 We sought to understand whether high-level cognition is supported by dynamic patterns of high-order  
328 correlations. To that end, we developed a computational framework for estimating the dynamics of high-  
329 order correlations in multivariate timeseries data (see *Dynamic inter-subject functional connectivity (DISFC)*  
330 and *Dynamic higher-order correlations*). We evaluated the efficacy of this framework at recovering known  
331 patterns in several synthetic datasets (see *Synthetic data*). We then applied the framework to a public fMRI  
332 dataset collected as participants listened to an auditorily presented story, a temporally scrambled version  
333 of the story, or underwent a resting state scan (see *Functional neuroimaging data collected during story listening*).  
334 We used the relative decoding accuracies of classifiers trained on different sets of neural features to estimate  
335 which types of features reflected ongoing cognitive processing.

### 336 Recovering known dynamic correlations from synthetic data

337 We generated synthetic datasets that differed in how the underlying correlations changed over time. For  
338 each dataset, we applied Equation 4 with a variety of kernel shapes and widths. We assessed how well  
339 the true underlying correlations at each timepoint matched the recovered correlations (Fig. 4). For every  
340 kernel and dataset we tested, our approach recovered the correlation dynamics we embedded into the data.  
341 However, the quality of these recoveries varied across different synthetic datasets in a kernel-dependent  
342 way.

343 In general, wide monotonic kernels (Laplace, Gaussian) performed best when the correlations varied  
344 gradually from moment-to-moment (Figs. 4a, c, and d). **TODO: Say something about kernel widths within**  
345 **a shape.** In the extreme, as the rate of change in correlations approaches 0 (Fig. 4a), an infinitely wide kernel  
346 would exactly recover the Pearson's correlation (e.g., compare Eqns. 2 and 4).

347 When the correlation dynamics were unstructured in time (Fig. 4b), a Dirac  $\delta$  kernel (infinitely narrow)  
348 performed best. This is because, when every timepoint's correlations are independent of the correlations in  
349 every other timepoint, averaging data over time dilutes the available signal. **TODO: Say something about**  
350 **kernel widths within a shape.**



**Figure 4: Recovering known dynamic correlations from synthetic data.** Each panel displays the average correlations between the vectorized upper triangles of the recovered correlation matrix at each timepoint and either the true underlying correlation at each timepoint or a reference correlation matrix. (The averages are taken across 10 different randomly generated synthetic datasets of the given category.) Error ribbons denote 95% confidence intervals (taken across datasets). Different colors denote different kernel shapes, whereas the shading within each color type denotes kernel widths. For a complete description of each synthetic dataset, see *Synthetic data*. **a. Constant correlations.** These datasets have a stable (unchanging) underlying correlation matrix. **b. Random correlations.** These datasets are generated using a new independently drawn correlation matrix at each new timepoint. **c. Ramping correlations.** These datasets are generated by smoothly varying the underlying correlations between the randomly drawn correlation matrices at first and last timepoints. The left panel displays the correlations between the recovered dynamic correlations and the underlying ground truth correlations. The middle panel compares the recovered correlations with the *first* timepoint’s correlation matrix. The right panel compares the recovered correlations with the *last* timepoint’s correlation matrix. **d. Event-based correlations.** These datasets are each generated using five randomly drawn correlation matrices that each remain stable for a fifth of the total timecourse. The left panel displays the correlations between the recovered dynamic correlations and the underlying ground truth correlations. The right panels compare the recovered correlations with the correlation matrices unique to each event.

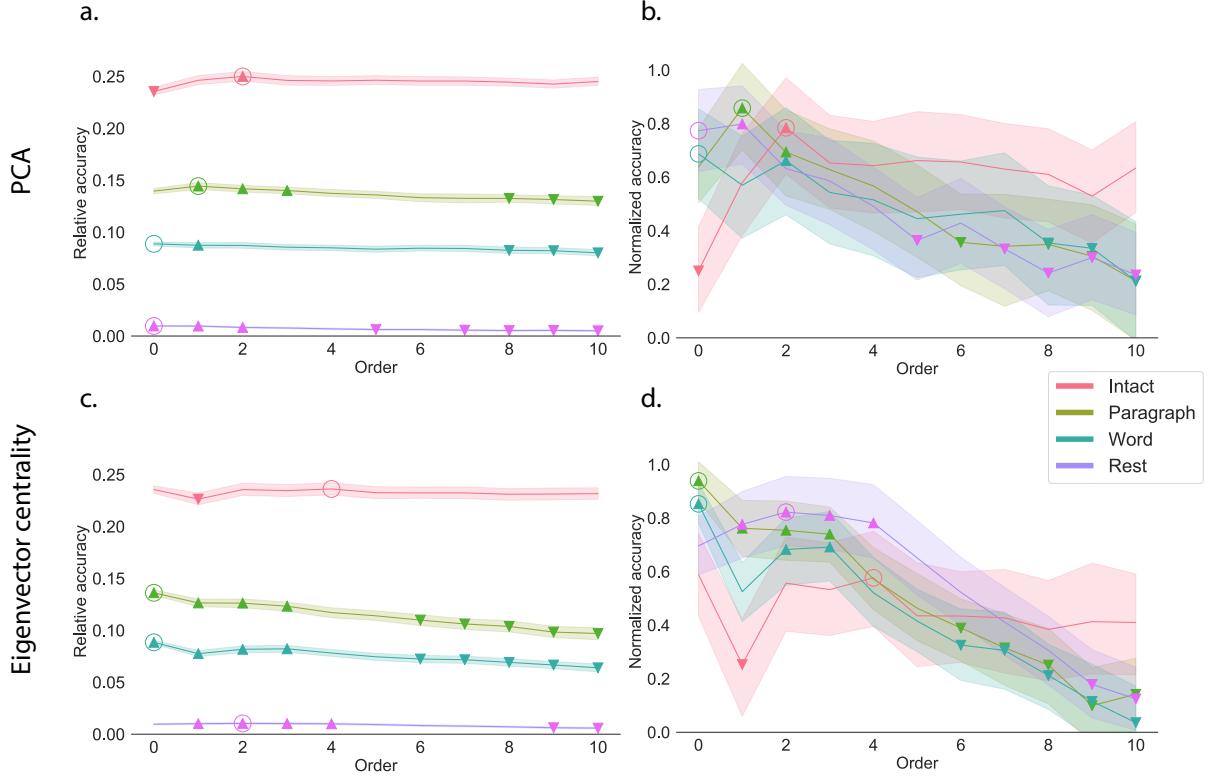
351 **Cognitively relevant dynamic high-order correlations in fMRI data**

352 We used across-participant temporal decoders to identify cognitively relevant neural patterns in fMRI data  
353 (see *Forward inference and decoding accuracy*). The dataset we examined (collected by Simony et al., 2016)  
354 comprised four experimental conditions that exposed participants to stimuli that varied systematically in  
355 how cognitively engaging they were. The *intact* experimental condition had participants listen to an audio  
356 recording of a 10-minute story. The *paragraph*-scrambled experimental condition had participants listen to a  
357 temporally scrambled version of the story, where the paragraphs occurred out of order (but where the same  
358 total set of paragraphs were presented over the full listening interval). All participants in this condition  
359 experienced the scrambled paragraphs in the same order. The *word*-scrambled experimental condition had  
360 participants listen to a temporally scrambled version of the story where the words in the story occurred in a  
361 random order. All participants in the word conditions experienced the scrambled words in the same order.  
362 Finally, in a *rest* experimental condition participants lay in the scanner with no overt stimulus, with their  
363 eyes open (blinking as needed). This dataset provided a convenient means of testing our hypothesis that  
364 different levels of cognitive engagement might be supported by different orders of complex brain activity  
365 dynamics.

366 In brief, we computed timeseries of dynamic high-order correlations that were similar across participants  
367 in each of two randomly assigned groups: a training group and a test group. We then trained classifiers  
368 on the training group's data to match each sample from the test group with a stimulus timepoint. Each  
369 classifier comprised a weighted blend of neural patterns that reflected up to  $n^{\text{th}}$ -order dynamic correlations  
370 (see *Feature weighting and testing*). We repeated this process for  $n \in \{0, 1, 2, \dots, 10\}$ . Our examinations of  
371 synthetic data suggested that none of the kernels we examined were "universal" in the sense of optimally  
372 recovering underlying correlations regardless of the structure of those correlations. In our analyses of neural  
373 data, we therefore averaged our decoding results over a variety of kernel shapes and widths in order to  
374 identify results that were robust to specific kernel parameters (also see *Identifying robust decoding results*).

375 Our approach to estimating dynamic high-order correlations requires mapping the high-dimensional  
376 feature space of correlations (a  $T$  by  $O(K^2)$  matrix) onto a lower-dimensional  $T$  by  $K$  matrix. We carried out  
377 two sets of analyses that differed in how this mapping was computed. The first set of analyses used PCA  
378 to find a low-dimensional embedding of the original dynamic correlation matrices (Fig. 5a,b). The second  
379 set of analyses characterized correlations in dynamics of each feature's eigenvector centrality, but did not  
380 preserve the underlying activity dynamics (Fig. 5c,d).

381 Both sets of temporal decoding analyses yielded qualitatively similar results for the auditory (non-  
382 rest) conditions of the experiment (Fig. 5: pink, yellow, and teal lines). The highest decoding accuracy



**Figure 5: Across-participant decoding accuracy varies with correlation order and cognitive engagement.**

**a. Decoding accuracy as a function of order: PCA.** Order (x-axis) refers to the maximum order of dynamic correlations that were available to the classifiers (see *Feature weighting and testing*). The reported across-participant decoding accuracies are averaged over all kernel shapes and widths (see *Identifying robust decoding results*). The y-values are displayed relative to chance accuracy (intact:  $\frac{1}{300}$ ; paragraph:  $\frac{1}{272}$ ; word:  $\frac{1}{300}$ ; rest:  $\frac{1}{400}$ ). The error bars denote 95% confidence intervals across cross-validation folds (i.e., random assignments of participants to the training and test sets). The colors denote the experimental condition. Arrows denote sets of features that yielded reliably higher (upwards facing) or lower (downward facing) decoding accuracy than the mean of all other features (via a two-tailed test, thresholded at  $p < 0.05$ ). The circled values represent the maximum decoding accuracy within each experimental condition.

**b. Normalized decoding accuracy as a function of order: PCA.** This panel displays the same results as Panel a, but here each curve has been normalized to have a maximum value of 1 and a minimum value of 0 (including the upper and lower bounds of the respective 95% confidence intervals). Panels a and b used PCA to project each high-dimensional pattern of dynamic correlations onto a lower-dimensional space.

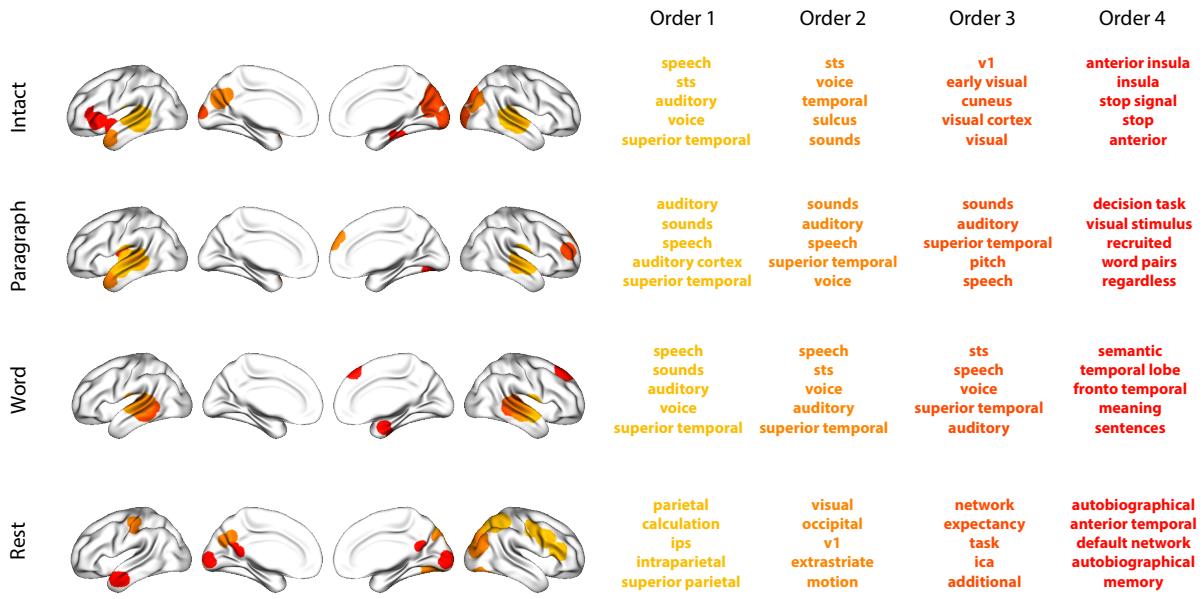
**c. Decoding accuracy as a function of order: eigenvector centrality.** This panel is in the same format as Panel a, but here eigenvector centrality has been used to project the high-dimensional patterns of dynamic correlations onto a lower-dimensional space.

**d. Normalized decoding accuracy as a function of order: eigenvector centrality.** This panel is in the same format as Panel b, but here eigenvector centrality has been used to project the high-dimensional patterns of dynamic correlations onto a lower-dimensional space.

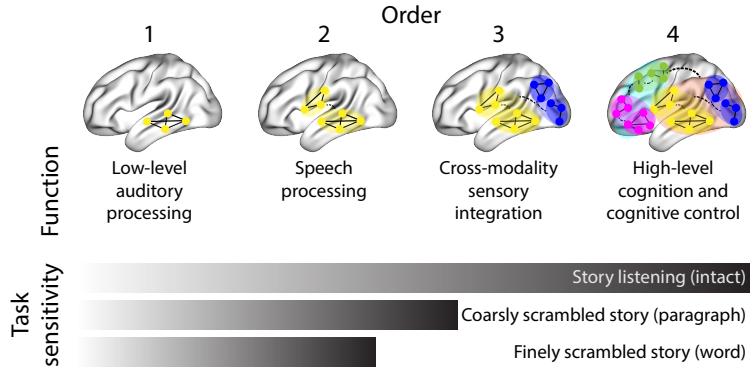
for participants who listened to the intact (unscrambled) story was achieved using high-order dynamic correlations (PCA: second-order; eigenvector-centrality: fourth-order). Scrambled versions of the story were best decoded by lower-order correlations (PCA/paragraph: first-order; PCA/word: order zero; eigenvector centrality/paragraph: order zero; eigenvector centrality/word: order zero). The two sets of analyses yielded different decoding results on resting state data (Fig. 5: purple lines). We note could be decoded only very slightly above chance; we speculate that the decoders might have picked up on attentional drift, boredom, or tiredness; we hypothesize these all increase throughout the resting state scan. Our decoders might be picking up on aspects of these loosely defined cognitive states that are common across individuals. The PCA-based approach achieved the highest resting state decoding accuracy using order zero features (non-correlational activation-based), whereas the eigenvector centrality-based approach achieved the highest resting state decoding accuracy using second-order correlations. Taken together, these analyses indicate that high-level cognitive processing (while listening to the intact story) is reflected in the dynamics of high-order correlations in brain activity, whereas lower-level cognitive processing (while listening to scrambled versions of the story that lack rich meaning) is reflected in the dynamics of lower-order correlations and non-correlational activity dynamics. Further, these patterns are associated both with the underlying activity patterns (characterized using PCA) and also with the changing relative positions that different brain areas occupy in their associated networks (characterized using eigenvector centrality).

Having established that patterns of high-order correlations are informative to decoders, we next wondered which specific networks of brain regions contributed most to these patterns. As a representative example, we selected the kernel parameters that yielded decoding accuracies that best matched the average accuracies across all of the kernel parameters we examined. Using Figure 5c as a template, the best-matching kernel was a Laplace kernel with a width of 50 (Fig. 3d). We used this kernel to compute a single  $K$  by  $K$   $n^{\text{th}}$ -order correlation matrix for each experimental condition, along with the Neurosynth (Rubin et al., 2017) terms most highly associated with each of these matrices (Fig. 6; see *Reverse inference*).

For all of the story listening conditions (intact, paragraph, and word), we found that first- and second-order correlations were most strongly associated with auditory and speech processing. During intact story listening, third-order correlations reflected integration with visual areas, and fourth-order correlations reflected integration with areas associated with high-level cognition and cognitive control, such as the ventrolateral prefrontal cortex. However, during listening to temporally scrambled stories, these higher-order correlations instead involved interactions with additional regions associated with speech and semantic processing. By contrast, we found a much different set of patterns in the resting state data. First-order resting state correlations were most strongly associated with regions involved in counting and numerical understanding. Second-order resting state correlations were strongest in visual areas; third-order correlations



**Figure 6: Top terms associated with the endpoints of the strongest correlations.** Each color corresponds to one order of correlations, averaged across participants and time. The inflated brain plots display the locations of the endpoints of the 10 strongest (absolute value) correlations at each order, projected onto the cortical surface (Combrisson et al., 2019). The lists of terms on the right display the top five Neurosynth terms (Rubin et al., 2017) decoded from the corresponding brain maps for each order. Each row displays data from a different experimental condition. Additional maps and their corresponding Neurosynth terms may be found in the *Supplementary materials* (intact: Figure S1; paragraph: Figure S2; word: Figure S3; rest: Figure S4).



**Figure 7: Proposed high-order network dynamics underlying high-level cognition.** Higher orders of network interactions support higher-level aspects of cognitive processing. When tasks evoke richer, deeper, and/or higher-level processing, this is reflected in higher-order network interactions.

416 were strongest in task-positive areas; and fourth-order correlations were strongest in regions associated  
 417 with autobiographical and episodic memory. We carried out analogous analyses to create maps (and de-  
 418 code the top associated Neurosynth terms) for up to fifteenth-order correlations (Figs. S1, S2, S3, and S4).  
 419 Of note, examining fifteenth-order correlations between 700 nodes using conventional methods would have  
 420 required storing roughly  $\frac{700^{2 \times 15}}{2} \approx 1.13 \times 10^{85}$  floating point numbers—assuming single-precision (32 bits  
 421 each), this would require roughly 32 times as many bits as there are molecules in the known universe!  
 422 Although these fifteenth-order correlations do appear (visually) to have some well-formed structure, we  
 423 provide this latter example primarily as a demonstration of the efficiency and scalability of our approach.

## 424 Discussion

425 We tested the hypothesis that high-level cognition is supported by high-order brain network dynamics (e.g.,  
 426 see Reimann et al., 2017; Solomon et al., 2019). We examined high-order network dynamics in functional  
 427 neuroimaging data collected during a story listening experiment. When participants listened to an auditory  
 428 recording of the story, participants exhibited similar high-order brain network dynamics. By contrast,  
 429 when participants instead listened to temporally scrambled recordings of the story, only lower-order brain  
 430 network dynamics were similar across participants. Our results indicate that higher orders of network  
 431 interactions support higher-level aspects of cognitive processing (Fig. 7).

432 The notion that cognition is reflected in (and possibly mediated by) patterns of network dynamics  
 433 has been suggested by or proposed in myriad empirical studies and reviews (e.g., Chang & Glover, 2010;  
 434 Demertzi et al., 2019; Fong et al., 2019; Gonzalez-Castillo et al., 2019; Liégeois et al., 2019; Lurie et al.,  
 435 2018; Park et al., 2018; Preti et al., 2017; Roy et al., 2019; Turk-Browne, 2013; Zou et al., 2019). Our study

436 extends this line of work by finding cognitively relevant higher-order network dynamics that reflect ongoing  
437 cognition. Our findings complement other work that uses graph theory and topology to characterize how  
438 brain networks reconfigure during cognition (e.g., Bassett et al., 2006; Betzel et al., 2019; McIntosh & Jirsa,  
439 2019; Reimann et al., 2017; Sizemore et al., 2018; Toker & Sommer, 2019; Zheng et al., 2019).

440 An open question not addressed by our study pertains to how different structures integrate incoming  
441 information with different time constants. For example, one line of work suggests that the cortical surface  
442 comprises a structured map such that nearby brain structures process incoming information at similar  
443 timescales. Low-level sensory areas integrate information relatively quickly, whereas higher-level regions  
444 integrate information relatively slowly (Baldassano et al., 2017; Hasson et al., 2015, 2008; Honey et al., 2012;  
445 Lerner et al., 2014, 2011). Other related work in human and mouse brains indicates that the temporal  
446 response profile of a given brain structure may relate to how strongly connected that structure is with other  
447 brain areas (Fallon et al., 2019). Further study is needed to understand the role of temporal integration at  
448 different scales of network interaction, and across different anatomical structures.

449 Another potential limitation of our approach relates to recent work suggesting that the brain undergoes  
450 rapid state changes, for example across event boundaries (e.g., Baldassano et al., 2017). Shappell et al. (2019)  
451 used hidden semi-Markov models to estimate state-specific network dynamics. Our general approach might  
452 be extended by considering putative state transitions. For example, rather than weighting all timepoints  
453 using a similar kernel (Eqn. 4), the kernel function could adapt on a timepoint-by-timepoint basis such that  
454 only timepoints determined to be in the same “state” were given non-zero weight.

455 Identifying higher-order network dynamics associated with high-level cognition required several im-  
456 portant methods advances. First, we used kernel-based dynamic correlations to extended the notion of  
457 (static) inter-subject functional connectivity (Simony et al., 2016) to a dynamic measure of inter-subject func-  
458 tional connectivity (DISFC) that does not rely on sliding windows, and that may be computed at individual  
459 timepoints. This allowed us to precisely characterize stimulus-evoked network dynamics that were similar  
460 across individuals. Second, we developed a computational framework for efficiently and scalably estimat-  
461 ing high-order dynamic correlations. Our approach uses dimensionality reduction algorithms and graph  
462 measures to obtain low-dimensional embeddings of patterns of network dynamics. Third, we developed  
463 an analysis framework for identifying robust decoding results by carrying out our analyses using a range  
464 of parameter values and then identifying which results were robust to specific parameter choices.

465 **Concluding remarks**

466 The complex hierarchy of dynamic interactions that underlie our thoughts is perhaps the greatest mystery in  
467 modern science. Methods for characterizing the dynamics of high-order correlations in neural data provides  
468 a window into the neural basis of cognition. **Need a good closing sentence.**

469 **Acknowledgements**

470 We acknowledge discussions with Luke Chang, Hany Farid, Paxton Fitzpatrick, Andrew Heusser, Eshin  
471 Jolly, Qiang Liu, Matthijs van der Meer, Judith Mildner, Gina Notaro, Stephen Satterthwaite, Emily Whitaker,  
472 Weizhen Xie, and Kirsten Ziman. Our work was supported in part by NSF EPSCoR Award Number 1632738  
473 to J.R.M. and by a sub-award of DARPA RAM Cooperative Agreement N66001-14-2-4-032 to J.R.M. The  
474 content is solely the responsibility of the authors and does not necessarily represent the official views of our  
475 supporting organizations.

476 **Author contributions**

477 Concept: J.R.M. Implementation: T.H.C., L.L.W.O., and J.R.M. Analyses: L.L.W.O and J.R.M.

478 **References**

- 479 Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. (2012). Tracking  
480 whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*, 24(3), 663–676.
- 481 Alvarez-Hamelin, I., Dall'Asta, L., Barrat, A., & Vespignani, A. (2005). *k*-corr decomposition: a tool for the  
482 visualiztion of large scale networks. *arXiv*, cs/0504107v2.
- 483 Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event  
484 structure in continuous narrative perception and memory. *Neuron*, 95(3), 709–721.
- 485 Barthélémy, M. (2004). Betweenness centrality in large complex networks. *European Physical Journal B*, 38,  
486 163–168.
- 487 Bassett, D., Meyer-Lindenberg, A., Achard, S., Duke, T., & Bullmore, E. (2006, December). Adaptive  
488 reconfiguration of fractal small-world human brain functional networks. *Proceedings of the National  
489 Academy of Sciences, USA*, 103(51), 19518-23.

- 490 Betzel, R. F., Byrge, L., Esfahlani, F. Z., & Kennedy, D. P. (2019). Temporal fluctuations in the brain's modular  
491 architecture during movie-watching. *bioRxiv*, doi.org/10.1101/750919.
- 492 Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, 29(4), 555–564.
- 493 Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and  
494 functional systems. *Nature Reviews Neuroscience*, 10(3), 186–198.
- 495 Capota, M., Turek, J., Chen, P.-H., Zhu, X., Manning, J. R., Sundaram, N., ... Shin, Y. S. (2017). *Brain imaging  
496 analysis kit*. Retrieved from <https://doi.org/10.5281/zenodo.59780>
- 497 Chang, C., & Glover, G. H. (2010). Time-frequency dynamics of resting-state brain connectivity measured  
498 with fMRI. *NeuroImage*, 50, 81–98.
- 499 Christakis, N. A., & Fowler, J. H. (2010). Social network sensors for early detection of contagious outbreaks.  
500 *PLoS One*, 5(9), e12948.
- 501 Combrisson, E., Vallat, R., O'Reilly, C., Jas, M., Pascarella, A., l Saive, A., ... Jerbi, K. (2019). Visbrain: a  
502 multi-purpose GPU-accelerated open-source suite for multimodal brain data visualization. *Frontiers in  
503 Neuroinformatics*, 13(14), 1–14.
- 504 Comon, P., Jutten, C., & Herault, J. (1991). Blind separation of sources, part II: Problems statement. *Signal  
505 Processing*, 24(1), 11 - 20.
- 506 Demertzi, A., Tagliazucchi, E., Dehaene, S., Deco, G., Barttfeld, P., Raimondo, F., ... Sitt, J. D. (2019). Human  
507 consciousness is supported by dynamic complex patterns of brain signal coordination. *Science Advances*,  
508 5(2), eaat7603.
- 509 Estrada, E., & Rodríguez-Velázquez, J. A. (2005). Subgraph centrality in complex networks. *Physical Review  
510 E*, 71(5), 056103.
- 511 Etzel, J. A., Gazzola, V., & Keysers, C. (2009). An introduction to anatomical ROI-based fMRI classification.  
512 *Brain Research*, 1282, 114–125.
- 513 Fallon, J., Ward, P., Parkes, L., Oldham, S., Arnatkevičiūtė, A., Fornito, A., & Fulcher, B. D. (2019).  
514 Timescales of spontaneous activity fluctuations relate to structural connectivity in the brain. *bioRxiv*,  
515 doi.org/10.1101/655050.
- 516 Fong, A. H. C., Yoo, K., Rosenberg, M. D., Zhang, S., Li, C.-S. R., Scheinost, D., ... Chun, M. M. (2019).  
517 Dynamic functional connectivity during task performance and rest predicts individual differences in  
518 attention across studies. *NeuroImage*, 188, 14–25.

- 519 Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 35–41.
- 520 Geisberger, R., Sanders, P., & Schultes, D. (2008). Better approximation of betweenness centrality. *Proceedings*  
521 *of the meeting on Algorithm Engineering and Experiments*, 90–100.
- 522 Gershman, S., Blei, D., Pereira, F., & Norman, K. (2011). A topographic latent source model for fMRI data.  
523 *NeuroImage*, 57, 89–100.
- 524 Gonzalez-Castillo, J., Caballero-Gaudes, C., Topolski, N., Handwerker, D. A., Pereira, F., & Bandettini, P. A.  
525 (2019). Imaging the spontaneous flow of thought: distinct periods of cognition contribute to dynamic  
526 functional connectivity during rest. *NeuroImage*, 202(116129).
- 527 Halu, A., Mondragón, R. J., Panzarasa, P., & Bianconi, G. (2013). Multiplex PageRank. *PLoS One*, 8(10),  
528 e78293.
- 529 Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: memory as an integral component  
530 of information processing. *Trends in Cognitive Science*, 19(6), 304-315.
- 531 Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows  
532 in human cortex. *Journal of Neuroscience*, 28(10), 2539-2550. doi: 10.1523/JNEUROSCI.5487-07.2008
- 533 Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and  
534 overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430.
- 535 Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks.  
536 *Science*, 313(5786), 504–507.
- 537 Honey, C. J., Kotter, R., Breakspear, M., & Sporns, O. (2007). Network structure of cerebral cortex shapes  
538 functional connectivity on multiple time scales. *Proceedings of the National Academy of Science USA*, 104(24),  
539 10240–10245.
- 540 Honey, C. J., Thesen, T., Donner, T. H., Silbert, L. J., Carlson, C. E., Devinsky, O., ... Hasson, U. (2012). Slow  
541 cortical dynamics and the accumulation of information over long timescales. *Neuron*, 76, 423-434. doi:  
542 10.1016/j.neuron.2012.08.011
- 543 Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals  
544 the semantic maps that tile human cerebral cortex. *Nature*, 532, 453–458.
- 545 Huth, A. G., Nisimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes  
546 the representation of thousands of object and action categories across the human brain. *Neuron*, 76(6),  
547 1210-1224.

- 548 Jutten, C., & Herault, J. (1991). Blind separation of sources, part I: An adaptive algorithm based on  
549 neuromimetic architecture. *Signal Processing*, 24(1), 1–10.
- 550 Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature*  
551 *Neuroscience*, 8, 679–685.
- 552 Landau, E. (1895). Zur relativen Wertbemessung der Turnierresultate. *Deutsches Wochenschach*, 11, 366–369.
- 553 Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*,  
554 401, 788–791.
- 555 Lerner, Y., Honey, C. J., Katkov, M., & Hasson, U. (2014). Temporal scaling of neural responses to compressed  
556 and dilated natural speech. *Journal of Neurophysiology*, 111, 2433–2444.
- 557 Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of  
558 temporal receptive windows using a narrated story. *Journal of Neuroscience*, 31(8), 2906–2915. doi: 10.1523/  
559 JNEUROSCI.3684-10.2011
- 560 Liégeois, R., Li, J., Kong, R., Orban, C., De Ville, D. V., Ge, T., … Yeo, B. T. T. (2019). Resting brain  
561 dynamics at different timescales capture distinct aspects of human behavior. *Nature Communications*,  
562 10(2317), doi.org/10.1038/s41467-019-10317-7.
- 563 Lin, J. (2009). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*,  
564 37(1), 145–151.
- 565 Lohmann, G., Margulies, D. S., Horstmann, A., Pleger, B., Lepsien, J., Goldhahn, D., … Turner, R. (2010).  
566 Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain.  
567 *PLoS One*, 5(4), e10232.
- 568 Lurie, D., Kessler, D., Bassett, D., Betzel, R., Breakspear, M., Keilholz, S., … Calhoun, V. (2018). On the  
569 nature of time-varying functional connectivity in resting fMRI. *PsyArXiv*, doi.org/10.31234/osf.io/xtzre.
- 570 Mack, M. L., Preston, A. R., & Love, B. C. (2017). Medial prefrontal cortex compresses concept representations  
571 through learning. *bioRxiv*, doi.org/10.1101/178145.
- 572 Mairal, J., Ponce, J., Sapiro, G., Zisserman, A., & Bach, F. R. (2009). Supervised dictionary learning. *Advances*  
573 *in Neural Information Processing Systems*, 1033–1040.
- 574 Mairal, J. B., Bach, F., Ponce, J., & Sapiro, G. (2009). Online dictionary learning for sparse coding. *Proceedings*  
575 *of the 26th annual international conference on machine learning*, 689–696.

- 576 Manning, J. R., Ranganath, R., Norman, K. A., & Blei, D. M. (2014). Topographic factor analysis: a Bayesian  
577 model for inferring brain networks from neural data. *PLoS One*, 9(5), e94914.
- 578 Manning, J. R., Zhu, X., Willke, T. L., Ranganath, R., Stachenfeld, K., Hasson, U., ... Norman, K. A. (2018).  
579 A probabilistic approach to discovering dynamic full-brain functional connectivity patterns. *NeuroImage*,  
580 180, 243–252.
- 581 McInnes, L., & Healy, J. (2018). UMAP: Uniform manifold approximation and projection for dimension  
582 reduction. *arXiv*, 1802(03426).
- 583 McIntosh, A. R., & Jirsa, V. K. (2019). The hidden repertoire of brain dynamics and dysfunction. *Network  
584 Neuroscience*, doi.org/10.1162/netn\_a\_00107.
- 585 Mitchell, T., Shinkareva, S., Carlson, A., Chang, K., Malave, V., Mason, R., & Just, M. (2008). Predicting  
586 human brain activity associated with the meanings of nouns. *Science*, 320(5880), 1191.
- 587 Newman, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, 27,  
588 39–54.
- 589 Newman, M. E. J. (2008). The mathematics of networks. *The New Palgrave Encyclopedia of Economics*, 2, 1–12.
- 590 Nishimoto, S., Vu, A., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. (2011). Reconstructing visual  
591 experience from brain activity evoked by natural movies. *Current Biology*, 21, 1 - 6.
- 592 Nocedal, J., & Wright, S. J. (2006). *Numerical optimization*. New York: Springer.
- 593 Norman, K. A., Newman, E., Detre, G., & Polyn, S. M. (2006). How inhibitory oscillations can train neural  
594 networks and punish competitors. *Neural Computation*, 18, 1577–1610.
- 595 Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: generalizing degree  
596 and shortest paths. *Social Networks*, 32, 245–251.
- 597 Park, H.-J., Friston, K. J., Pae, C., Park, B., & Razi, A. (2018). Dynamic effective connectivity in resting state  
598 fMRI. *NeuroImage*, 180, 594–608.
- 599 Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh,  
600 and Dublin Philosophical Magazine and Journal of Science*, 2, 559-572.
- 601 Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., ... Fedorenko, E. (2018). Toward  
602 a universal decoder of linguistic meaning from brain activation. *Nature Communications*, 9(963), 1–13.

- 603 Preti, M. G., Bolton, T. A. W., & Van De Ville, D. (2017). The dynamic functional connectome: state-of-the-art  
604 and perspectives. *NeuroImage*, 160, 41–54.
- 605 Rao, C. R. (1982). Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology*,  
606 21(1), 24–43.
- 607 Reimann, M. W., Nolte, M., Scolamiero, M., Turner, K., Perin, R., Chindemi, G., ... Markram, H. (2017).  
608 Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers*  
609 in *Computational Neuroscience*, 11(48), 1–16.
- 610 Ricotta, C., & Szeidl, L. (2006). Towards a unifying approach to diversity measures: Bridging the gap  
611 between the Shannon entropy and Rao's quadratic index. *Theoretical Population Biology*, 70(3), 237–243.
- 612 Roy, D. S., Park, Y.-G., Ogawa, S. K., Cho, J. H., Choi, H., Kamensky, L., ... Tonegawa, S. (2019). Brain-  
613 wide mapping of contextual fear memory engram ensembles supports the dispersed engram complex  
614 hypothesis. *bioRxiv*, doi.org/10.1101/668483.
- 615 Rubin, T. N., Kyoejo, O., Gorgolewski, K. J., Jones, M. N., Poldrack, R. A., & Yarkoni, T. (2017). Decoding  
616 brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. *PLoS*  
617 *Computational Biology*, 13(10), e1005649.
- 618 Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpreta-  
619 tions. *NeuroImage*, 52, 1059–1069.
- 620 Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85(2), 461–464.
- 621 Shappell, H., Caffo, B. S., Pekar, J. J., & Lindquist, M. A. (2019). Improved state change estimation in  
622 dynamic functional connectivity using hidden semi-Markov models. *NeuroImage*, 191, 243–257.
- 623 Simony, E., Honey, C. J., Chen, J., & Hasson, U. (2016). Uncovering stimulus-locked network dynamics  
624 during narrative comprehension. *Nature Communications*, 7(12141), 1–13.
- 625 Sizemore, A. E., Giusti, C., Kahn, A., Vettel, J. M., Betzel, R. F., & Bassett, D. S. (2018). Cliques and cavities  
626 in the human connectome. *Journal of Computational Neuroscience*, 44(1), 115–145.
- 627 Solomon, S. H., Medaglia, J. D., & Thompson-Schill, S. L. (2019). Implementing a concept network model.  
628 *Behavior Research Methods*, 51, 1717–1736.
- 629 Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of*  
630 *Psychology*, 15, 201–292.

- 631 Sporns, O., & Honey, C. J. (2006). Small worlds inside big brains. *Proceedings of the National Academy of*  
632 *Science USA*, 103(51), 19219–19220.
- 633 Thompson, W. H., Richter, C. G., Plavén-Sigray, P., & Fransson, P. (2018). Simulations to benchmark  
634 time-varying connectivity methods for fMRI. *PLoS Computational Biology*, 14(5), e1006196.
- 635 Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of Royal Statistical*  
636 *Society, Series B*, 61(3), 611–622.
- 637 Toker, D., & Sommer, F. T. (2019). Information integration in large brain networks. *PLoS Computational*  
638 *Biology*, 15(2), e1006807.
- 639 Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology*, 63,  
640 483–509.
- 641 Turk-Browne, N. B. (2013). Functional interactions as big data in the human brain. *Science*, 342, 580–584.
- 642 van der Maaten, L. J. P., & Hinton, G. E. (2008). Visualizing high-dimensional data using t-SNE. *Journal of*  
643 *Machine Learning Research*, 9, 2579-2605.
- 644 Zar, J. H. (2010). *Biostatistical analysis*. Prentice-Hall/Pearson.
- 645 Zheng, M., Allard, A., Hagmann, P., & Serrano, M. A. (2019). Geometric renormalization unravels self-  
646 similarity of the multiscale human connectome. *arXiv*, 1904.11793.
- 647 Zou, Y., Donner, R. V., Marwan, N., Donges, J. F., & Kurths, J. (2019). Complex network approaches to  
648 nonlinear time series analysis. *Physics Reports*, 787, 1–97.