High-level cognition is supported by at least second order dynamic correlations in neural activity patterns

Lucy L. W. Owen[1], Thomas H. Chang[1,2], and Jeremy R. Manning[1,†]

[1]Department of Psychological and Brain Sciences,
Dartmouth College, Hanover, NH
[3]Amazon.com, Seattle, WA
[†]Address correspondence to jeremy.r.manning@dartmouth.edu

September 3, 2019

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 or 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 activity snapshot was collected. We found that classifiers trained to decode from high-order dynamic correlations yielded better 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 raw 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.

Introduction

A central goal in cognitive neuroscience is to elucidate the *neural code*: the mapping between (a) mental states or cognitive representations and (b) neural activity patterns. One means of testing models of the neural code is to ask how accurately that model is able to "translate" neural activity patterns into known (or hypothesized) mental states or cognitive representations (e.g., Haxby et al., 2001; Huth et al., 2016, 2012; Kamitani & Tong, 2005; Mitchell et al., 2008; Nishimoto et al., 2011; Norman et al., 2006; Pereira et al., 2018; Tong & Pratte, 2012). Training decoding models on different types of neural features can also help to elucidate which specific aspects of neural activity patterns are informative about cognition– and, by extension, which types of neural activity patterns might comprise the neural code. For example, prior work has used region of interest analyses to estimate the anatomical locations of specific neural representations (e.g., Etzel et al., 2009), or to compare the relative contributions to the neural code of multivariate activity patterns versus patterns of dynamic correlations between neural activity patterns (e.g., Fong et al., 2019; Manning et al., 2018). An emerging theme in this literature is that cognition is mediated by complex dynamic interactions between brain structures (Bassett et al., 2006; Demertzi et al., 2019; Sporns & Honey, 2006; Turk-Browne, 2013).

Studies of the neural code to date have primarily focused on univariate or multivariate neural patterns (for review see , NormEtal06) or (more recently) on patterns of dynamic first-order correlations (i.e., interactions between pairs of brain structures; Fong et al., 2019; Manning et al., 2018). We wondered what the future of this line of work might hold. For example, is the neural code mediated by higher-order interactions between brain structures? Second-order correlations reflect *homologous* patterns of correlation. In other words, if the changing patterns of correlations between two regions, A and B, are similar to those between two other regions, C and D, this would be reflected in the second-order correlations between (A–B) and (C–D). In this way, second-order correlations identify similarities and differences between subgraphs of the brain's connectome. Analogously, third-order correlations reflect homologies between second-order correlations– i.e., homologous patterns of homologous interactions between brain regions. More generally, higher-order correlations reflect homologies between patterns of lower-order correlations. We can then ask: which "orders" of interaction are most reflective of high-level cognitive processes?

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

To gain insights into the above aspects of the neural code, we developed a computational framework for estimating dynamic high-order correlations in timeseries data. This framework provides an important advance, in that it enables us to examine patterns in higher-order correlations that are computationally intractable to estimate via conventional methods. Given a multivariate timeseries, our framework provides timepoint-by-timepoint estimates of the first-order correlations, second-order correlations, and so on (up to tenth-order correlations in this manuscript). Our approach combines a kernel-based method for computing dynamic correlations in timeseries data with a dimensionality reduction step that projects the resulting dynamic correlations into a low-dimensional space. We explored two dimensionality reduction approaches: principle components analysis (PCA; Pearson, 1901), which preserves an approximately invertable transformation back to the original data; and a second non-invertible algorithm that explored patterns in eigenvector centrality (Landau, 1895). This latter approach characterizes correlations between each feature dimension's relative *position* in the network in favor of the specific activity histories of different features.

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 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. 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 scambled 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. 1). 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 dynamnic 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 n^{th}-order dynamic

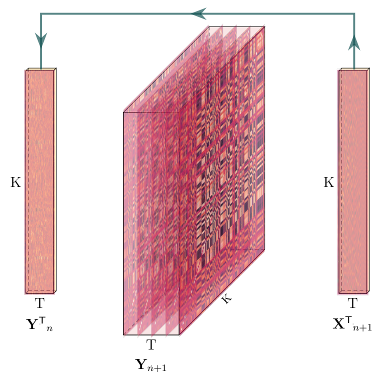

Figure 1: **Estimating dynamnic high-order correlations.** Given a T by K matrix of multivariate timeseries data, $\mathbf{Y}_n$ (where $n \in \mathbb{N}, n \geq 0$), we use Equation 5 to compute a timeseries of K by K correlation matrices, $\mathbf{Y}_{n+1}$. We then approximate $\mathbf{Y}_{n+1}$ with 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 $\mathbf{Y}_n$ and $\mathbf{X}_{n+1}$ are displayed in the figure for compactness.

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^{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 at 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. **JRM NOTE: CHECK LINK**

Kernel-based approach for computing dynamic correlations

Given a matrix of observations, 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):

$$\mathrm{corr}(\mathbf{X}(\cdot, i), \mathbf{X}(\cdot, j)) = \frac{\sum_{t=1}^{T} \left(\mathbf{X}(t, i) - \mathbf{X}(\bar{\cdot}, i)\right)\left(\mathbf{X}(t, j) - \mathbf{X}(\bar{\cdot}, j)\right)}{\sqrt{\sum_{t=1}^{T} \sigma^2_{\mathbf{X}(\cdot, i)} \sigma^2_{\mathbf{X}(\cdot, j)}}}, \text{ where} \tag{1}$$

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

$$\sigma^2_{\mathbf{X}(\cdot, k)} = \sum_{t=1}^{T} \left(\mathbf{X}(t, k) - \mathbf{X}(\bar{\cdot}, k)\right)^2 \tag{3}$$

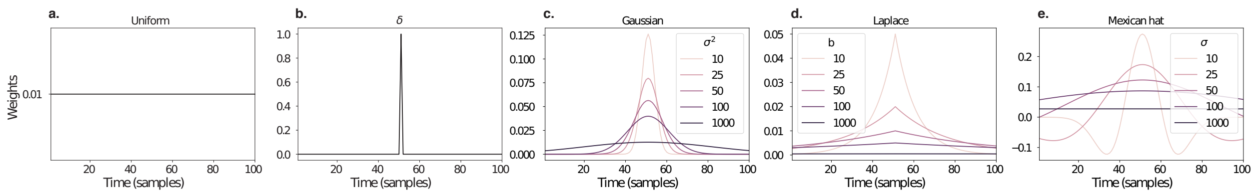

Figure 2: **Examples of kernel functions.** Each panel displays per-timepoint weights at $t = 50$, evaluated for 100 timepoints ($\tau \in [1, ..., 100]$). **a. Uniform kernel.** The weights are timepoint-invariant; observations at all timepoints are weighted equally, and do not change as a function of τ. This is a special case of weight function that reduces dynamic correlations to static correlations. **b. Dirac δ 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 (σ^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 (σ) are displayed.

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. 2).

Given a kernel function $\kappa_t(\cdot)$ for timepoint t, evaluated at timepoints $\tau \in [1, ..., T]$, we can update the static correlation formula in Equation 2 to estimate the *instantaneous correlation* at timepoint t:

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

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

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

Here $\mathrm{timecorr}(\mathbf{X}(\cdot, i), \mathbf{X}(\cdot, j), \kappa_t)$ reflects the correlation at time t between columns i and j of $\mathbf{X}$, estimated using the kernel κ_t. We evaluate Equation 5 in turn each pair of columns in $\mathbf{X}$ and for kernels centered on each timepoint in the timeseries, respectively, to obtain a T by K by K timeseries of dynamic correlations, $\mathbf{Y}$. For convenience, we then reshape the upper triangles of each timepoint's correlation matrix into a row vector to obtain an equivalent T by $\frac{K^2-K}{2}$ matrix.

Dynamic inter-subject functional connectivity (DISFC)

Equation 5 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 $\mathbf{X}_n^1$, $\mathbf{X}_n^2$, ..., $\mathbf{X}_n^P$ 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:

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

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

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

R is the inverse of Z:

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

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, $\bar{\mathbf{X}}_n^{\setminus p}$:

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

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

Low-dimensional representations of dynamic correlations

Given a T by $\frac{K^2-K}{2}$ 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 ($k \in [1, ..., K]$) in the network defined by the correlation matrix at each timepoint.

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

The modern library of dimensionality reduction algorithms include Principal Components Analysis (PCA; Pearson, 1901), Probabilistic PCA (PPCA; Tipping & Bishop, 1999), Exploratory Factor Analysis (EFA; Spearman, 1904), Independent Components Analysis (ICA; Comon et al., 1991; Jutten & Herault, 1991), *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 & Seung, 1999), Topographic Factor Analysis (TFA) Manning et al. (2014), Hierarchical Topographic Factor analysis (HTFA) Manning et al. (2018), Topographic Latent Source Analysis (TLSA) Gershman et al. (2011), Dictionary learning (J. Mairal et al., 2009; J. B. Mairal et al., 2009), deep autoencoders (Hinton & Salakhutdinov, 2006), among others. While complete characterizations of each of these algorithms is beyond the scope of the present manuscript, the general intuition driving these approaches is to compute the T by I matrix, $\mathbf{X}$, that is closest to the original T by J matrix, $\mathbf{Y}$, where (typically) $I \ll J$. The different approaches place different constraints on what properties $\mathbf{X}$ must satisfy and which aspects of the data are compared (and how) to characterize the match between $\mathbf{X}$ and $\mathbf{Y}$.

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

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

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

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

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

Dynamic higher-order correlations

Because $\mathbf{X}_n$ has the same shape as the original data $\mathbf{X}_0$, approximating $\mathbf{Y}_n$ with a lower-dimensional $\mathbf{X}_n$ enables us to estimate high-order dynamic correlations in a scalable way. Given a T by K input matrix, the output of Equation 5 requires $\mathcal{O}(TK^2)$ space to store. Repeated applications of Equation 5 (i.e., computing dynamic correlations between the columns of the outputted dynamic correlation matrix) each require exponentially more space; in general the n^{th}-order dynamic correlations of a T by K timeseries occupies $O(TK^{2n})$ space. However, when we approximate or summarize the output of Equation 5 with a T by K matrix (as described above), it becomes feasible to compute even very high-order correlations in high-dimensional data. Specifically, approximating the n^{th}-order dynamic correlations of a T by K timeseries requires only $O(TK^2)$ additional space– the same as would be required to compute first-order dynamic correlations. In other words, the space required to store $n + 1$ multivariate timeseries reflecting up to n^{th} order correlations in the original data scales linearly with n using our approach (Fig. 1).

Data

We examined two types of data: synthetic data and human functional neuroimaging data. We constructed and leveraged the synthetic data to evaluate our general approach. Specifically, we tested how well Equation 5 could be used to recover known dynamic correlations using different choices of kernel (κ; Fig. 2), for each of several synthetic datasets that exhibited different temporal properties. We applied our approach to a functional neuroimaging dataset to test the hypothesis that ongoing cognitive processing is reflected in high-order dynamic correlations. We used an across-participant classification test to estimate whether dynamic correlations of different orders contain information about which timepoint in a story participants were listening to.

Synthetic data

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

Constant. We generated data with stable underlying correlations to evaluate how Equation 5 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, $\mathbf{\Sigma}_m$:

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

$$\mathbf{\Sigma}_m = \mathbf{C}\mathbf{C}^\top \text{ , where} \tag{12}$$

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

Random. We generated a second set of 10 synthetic datasets whose observations at each timepoint were drawn from a Gaussian distribution with a new randomly constructed (using Eqn. 12) covariance matrix. Because each timepoint's covariance matrix was drawn independently of the covariance matrices for all other timepoints, these datasets provided a test of reconstruction accuracy in the absence of any meaningful 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, $\mathbf{\Sigma}_{\text{start}}$ and $\mathbf{\Sigma}_{\text{end}}$. For each of the 300 timepoints in each dataset, we drew the observations from a multivariate Gaussian distribution whose covariance matrix at each timepoint $t \in [0, ..., 299]$ was given by

$$\mathbf{\Sigma}_t = \left(1 - \frac{1-t}{299}\right)\mathbf{\Sigma}_{\text{start}} + \frac{t}{299}\mathbf{\Sigma}_{\text{end}}. \tag{13}$$

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

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

Functional neuroimaging data collected during story listening

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

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 timeseries of voxel activations (44,415 voxels) using a much smaller number of radial basis function (RBF) nodes (in this case 700 nodes). This provides a convenient representation for examining full-brain network dynamics. All of the analyses we carried out on the neuroimaging dataset were performed in this lower-dimensional space. In other words, each participant's data matrix, $\mathbf{Y}_0$, was a number-of-timepoints by 700 matrix of HTFA-derived factor weights (where the row and column labels were matched across participants). Code for carrying out HTFA on fMRI data may be found as part of the BrainIAK toolbox (Capota et al., 2017), which may be downloaded at brainiak.org.

Temporal decoding

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

Forward inference and decoding accuracy

We used an across-participants correlation-based classifier to decode which stimulus timepoint matched a given neural pattern. We first divided the participants into two groups: a template group, $\mathcal{G}_{\text{template}}$, and a 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 $\bar{\mathbf{C}}_{\text{decode}}$, 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 number-of-timepoints decoding matrix, $\mathbf{\Lambda}$. In this way, the rows of $\mathbf{\Lambda}$ reflected timepoints from the template group, while the columns reflected timepoints from the to-be-decoded group. We assigned temporal labels 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 this decoding procedure, but using $\mathcal{G}_{\text{decode}}$ as the template group and $\mathcal{G}_{\text{template}}$ as the to-be-decoded group. Given the true timepoint labels (for each group), we defined the *decoding accuracy* as the proportion of correctly decoded timepoints, across both groups.

Feature weighting and testing

We sought to examine which types of neural features (i.e., activations, first-order dynamic correlations, and higher-order dynamic correlations) were informative to the temporal decoders. Using the notation above, 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 $\mathbf{X}_{10}$).

One challenge to fairly evaluating high-order correlations is that if the kernel used in Equation 5 is wider than a single timepoint, each repeated application of the equation will result in further temporal blur. Because our primary assessment metric is temporal decoding accuracy, this unfairly biases against detecting meaningful signal in higher-order correlations (relative to lower-order correlations). We attempted to mitigate temporal blur in estimating each $\mathbf{X}_n$ by using a Dirac δ function kernel (which places all of its mass over a single timepoint; Fig. 2b) to compute each lower-order correlation ($\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_{n-1}$). We then used a (potentially wider, as described below) kernel to compute $\mathbf{X}_n$ from $\mathbf{X}_{n-1}$. In this way, temporal blurring was applied only in the last step of computing $\mathbf{X}_n$. We note that, because each $\mathbf{X}_n$ is a low-dimensional representation of the corresponding $\mathbf{Y}_n$, the higher-order correlations we estimated reflect true correlations in the data with lower-fidelity than estimates of lower-order correlations.

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

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

We used a permutation-based procedure to form a stable estimate of decoding accuracy for each set of neural features. In particular, we computed the decoding accuracy for each of 10 random group assignments 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.

Identifying robust decoding results

The temporal decoding procedure we use to estimate which neural features support ongoing cognitive processing is goverened by many parameters. For example, Equation 5 requires defining a kernel function, which can take on different shapes and widths. For a fixed set of neural features, each of these parameters can yield different decoding accuracies. Further, the best decoding accuracy for a given timepoint may be reliably achieved by one set of parameters, whereas the best decoding accuracy for another timepoint might be reliably achieved by a different set of parameters, and the best decoding accuracy across *all* timepoints might be reliably achieved by still another different set of parameters. Rather than attempting to maximize decoding accuracy, we sought to discover the trends in the data that were robust to specific classifier parameters choices. Specifically, we sought to characterize how decoding accuracy varied (under different experimental conditions) as a function of which neural features were considered.

To identify decoding results that were robust to specific classifier parameter choices, we repeated our decoding analyses that substituted in a variety of kernel shapes and widths for Equation 5. We examined Gaussian (Fig. 2c), Laplace (Fig. 2d), and Mexican Hat (Fig. 2e) kernels, each with widths of 5, 10, 20, and 50 samples. We then report the average decoding accuracies across all of these parameter choices. This enabled us to (roughly) factor out performance characteristics that were parameter dependent (within the space of parameters we examined).

Reverse inference

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

Results

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

Recovering known dynamic correlations from synthetic data

We generated synthetic datasets that differed in how the underlying correlations changed over time. For each dataset, we applied Equation 5 with a variety of kernel shapes and widths. We assessed how well the true underlying correlations at each timepoint matched the recovered correlations (Fig. 3). For every kernel and dataset we tested, our approach recovered the correlation dynamics we embedded into the data. However, the quality of these recoveries varied across different synthetic datasets in a kernel-dependent way.

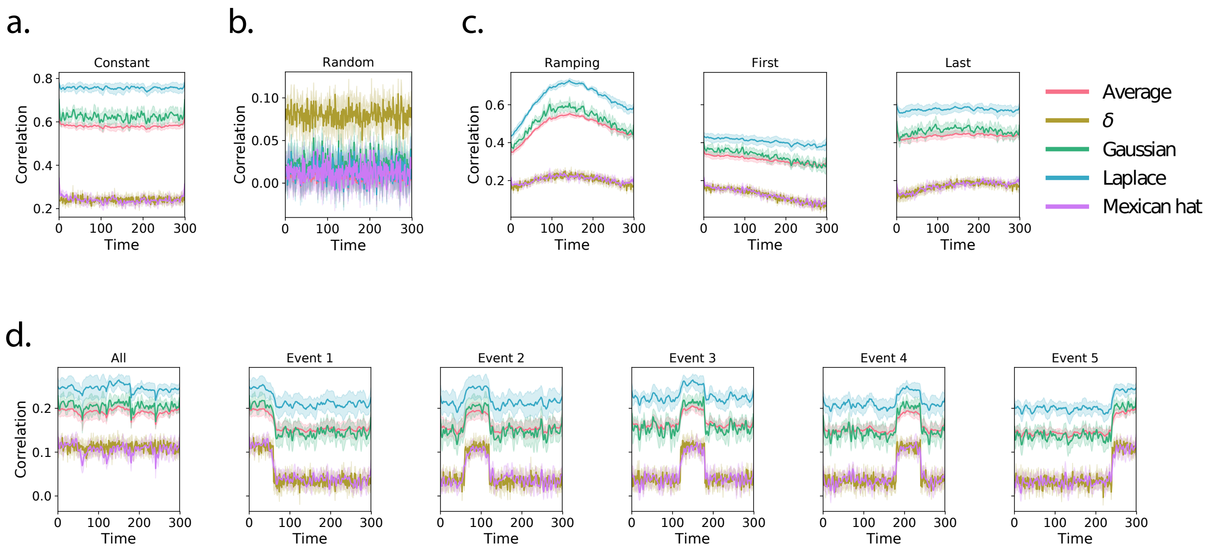

Figure 3: **Recovering known dynamic correlations from synthetic data.** Each panel displays the average correlations between the vectorized upper triangles of the recovered correlation matrix at each timepoint and either the true underlying correlation at each timepoint or a reference correlation matrix. (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.

In general, wide monotonic kernels (Laplace, Gaussian) performed best when the correlations varied gradually from moment-to-moment (Figs. 3a, c, and d). **TODO: Say something about kernel widths within a shape.** In the extreme, as the rate of change in correlations approaches 0 (Fig. 3a), an infinitely wide kernel would exactly recover the Pearson's correlation (e.g., compare Eqns. 2 and 5).

When the correlation dynamics were unstructured in time (Fig. 3b), a Dirac δ kernel (infinitely narrow) performed best. This is because, when every timepoint's correlations are independent of the correlations in every other timepoint, averaging data over time dilutes the available signal. **TODO: Say something about kernel widths within a shape.**

Cognitively relevant dynamic high-order correlations in fMRI data

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

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

Our approach to estimating dynamic high-order correlations requires mapping the high-dimensional feature space of correlations (a T by $O(K^2)$ matrix) onto a lower-dimensional T by K matrix. We carried out two sets of analyses that differed in how this mapping was computed. The first set of analyses used PCA to find a low-dimensional embedding of the original dynamic correlation matrices (Fig. 4a,b). The second set of analyses characterized correlations in dynamics of each feature's eigenvector centrality, but did not preserve the underlying activity dynamics (Fig. 4c,d).

Both sets of temporal decoding analyses yielded qualitatively similar results for the auditory (non-rest) conditions of the experiment. The highest decoding accuracy for participants who listened to the intact (unscrambled) story was achieved using high-order dynamic correlations (PCA: second-order; eigenvector-centrality: fourth-order). Scambled 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 results on resting state data (which we note could be decoded only very slightly above chance). 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.

JRM STOPPED HERE

Discussion

- Methods advances: kernel-based dynamic correlations, extension to dynamic ISFC, efficient method for estimating high-order dynamic correlations, identifying robust results by averaging
- Discoveries:
  - Dimensionality reduction and graph theoretic approaches give different insights into the data and identify different patterns as being relevant to cognition (different peak orders).
  - An insight common to both approaches is that high-order (greater than first order) dynamic correlations are informative about ongoing high-level cognitive processing. As the level of cognitive processing decreases, cognition is reflected by lower-order correlations.
  - Correlations at different orders are also associated with different networks of brain regions. However, which networks reflect which types of interactions depends on the current task. In general,

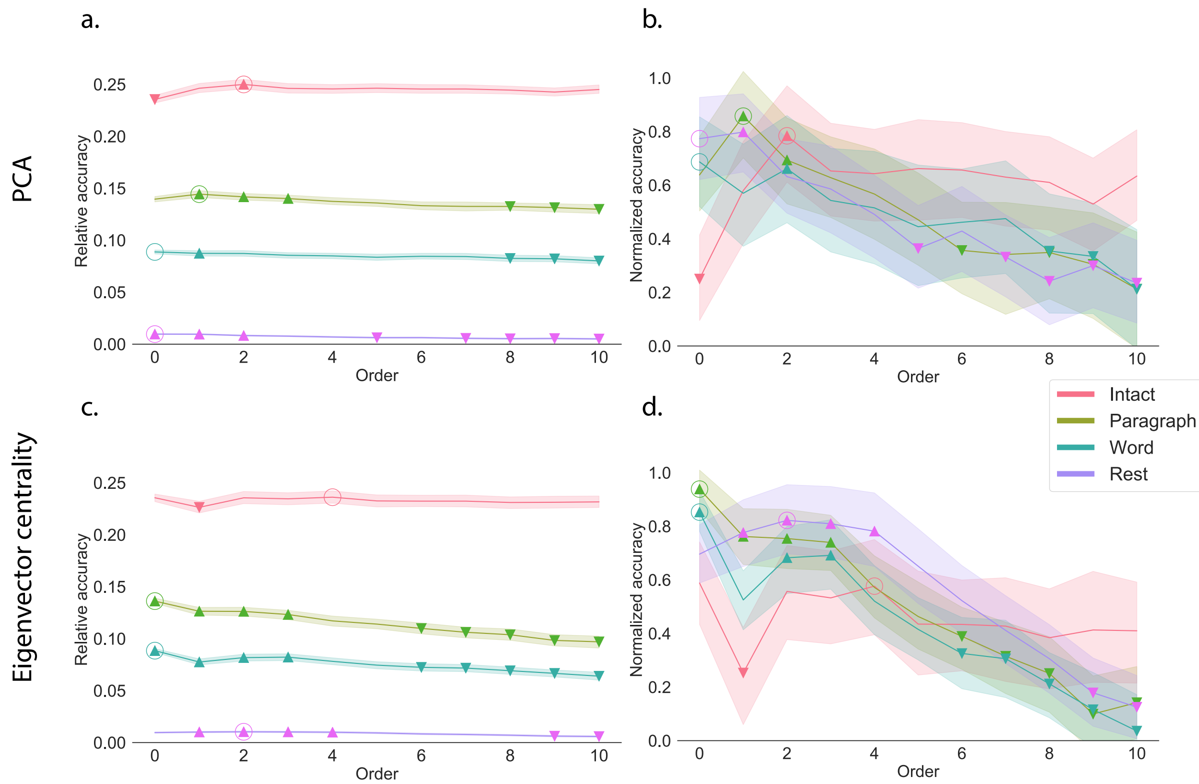

Figure 4: **Across-participant decoding accuracy varies with correlation order and cognitive engagement. a. Decoding accuracy as a function of order: PCA.** *Order* (x-axis) refers to the maximum order of dynamic correlations that were available to the classifiers (see *Feature weighting and testing*). 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. 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 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. Noramlized 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.

lower order correlations during auditory listening reflect processing of low-level (auditory) features; mid-order correlations reflect speech and linguistic processing; higher-order correlations reflect across-sensory integration (e.g. ties to visual areas) and cognitive control areas. This hierarchy dissolves during lower-order cognitive processing.

Based on prior work (Demertzi et al., 2019) and following the direction of the field (Turk-Browne, 2013) we think our thoughts might be encoded in dynamic network patterns, and possibly higher order network patterns (Fig. 5). We sought to test this hypothesis by developing an approach to inferring high-order network dynamics from timeseries data.

One challenge in studying dynamic interactions is the computational resources required to calculate higher-order correlations. We developed a computationally tractable model of network dynamics (Fig. 1) that takes in a feature timeseries and outputs approximated first-order dynamics (i.e., dynamic functional correlations), second-order dynamics (reflecting homologous networks that dynamically form and disperse), and higher-order network dynamics (up to tenth-order dynamic correlations).

We first validated our model using synthetic data, and explored how recovery varied with different underlying data structures and kernels. We then applied the approach to an fMRI dataset (Simony et al., 2016) in which participants listened to an audio recording of a story, as well as scrambled versions of the same story (where the scrambling was applied at different temporal scales). We trained classifiers to take the output of the model and decode the timepoint in the story (or scrambled story) that the participants were listening to. We found that, during the intact listening condition in the experiment, classifiers that incorporated higher-order correlations yielded consistently higher accuracy than classifiers trained only on lower-order patterns (Fig. **??**, a.&d.). By contrast, these higher-order correlations were not necessary to support decoding the other listening conditions and (minimally above chance) during a control rest condition. This suggests that the cognitive processing that supported the most cogntively rich listening conditions involved second-order (or higher) network dynamics.

Although we found decoding accuracy was best when incorporating higher-order network dynamics for all but rest condition, it is unclear if this is a product of the brain or the data collection technique. It could be that the brain is second-order or it could be that fMRI can only reliably give second-order interactions. Exploring this method with other data collection technique will be important to disentangle this question.

Concluding remarks

How can we better understand how brain patterns change over time? How can we quantify the potential network dynamics that might be driving these changes? One way to judge the techniques of the future is

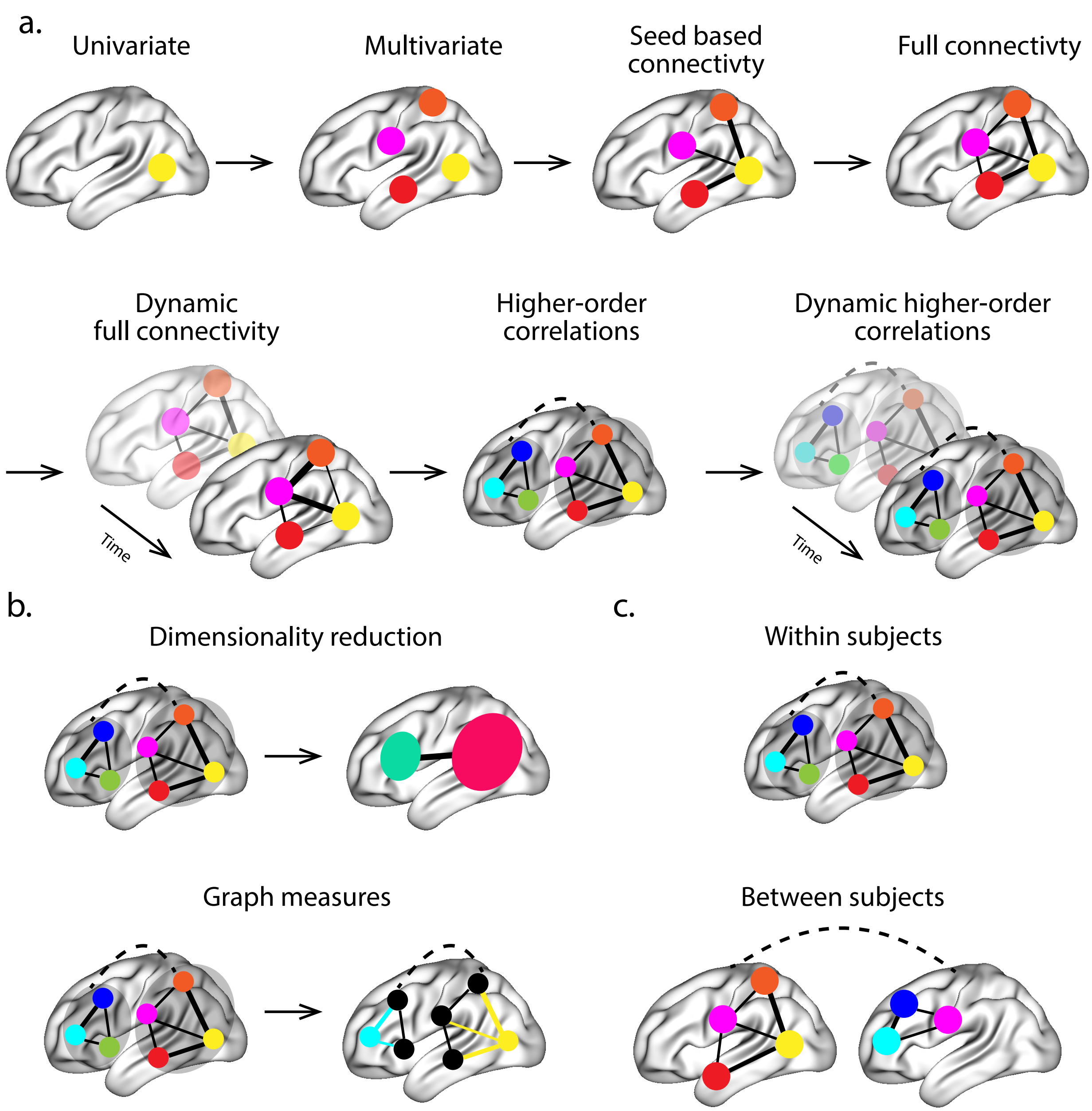

Figure 5: **Direction of the field (adapted from (Turk-Browne, 2013)).** The evolution of fMRI analyses started with univariate activation, which refers to the average amplitude of BOLD activity evoked by events of an experimental condition. Next, multivariate classifiers are trained on patterns of activation across voxels to decode distributed representations for specific events. The next, resting connectivity, is the temporal correlation of one or more seed regions with the remainder of the brain during rest. Additionally, task-based connectivity examines how these correlations differ by cognitive state. Following this increasing trajectory of increasing complexity, full connectivity considers all pairwise correlations in the brain, most commonly at rest. Next, dynamic full connectivity considers how full connectivity changes over time. Continuing this line of reasoning, we expect higher-order network dynamics might provide even richer insights into the neural basis of cognition.

to look at the trajectory of the fMRI field so far has taken so far (Fig. 1). The field started with univariate activation, measuring the average activity for each voxel. Analyses of multivariate activation followed, looking at spatial patterns of activity over voxels. Next, correlations of activity were explored, first with measures like resting connectivity that take temporal correlation between a seed voxel and all other voxels then with full connectivty that measure all pairwise correlations. Additionally, this path of increasing complexity also moved from static to dynamic measurements. One logical next step in this trajectory would be dynamic higher-order correlations. We have created a method to support these calculations by scalably approximating dynamic higher-order correlations.

Acknowledgements

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

Author contributions

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.

References

Alvarez-Hamelin, I., Dall'Asta, L., Barrat, A., & Vespignani, A. (2005). *k*-corr decomposition: a tool for the visualiztion of large scale networks. *arXiv*, cs/0504107v2.

Barthélemy, M. (2004). Betweenness centrality in large complex networks. *European Physical Journal B*, *38*, 163–168.

Bassett, D., Meyer-Lindenberg, A., Achard, S., Duke, T., & Bullmore, E. (2006, December). Adaptive reconfiguration of fractal small-world human brain functional networks. *Proceedings of the National Academy of Sciences, USA*, *103*(51), 19518-23.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, *29*(4), 555–564.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198.

Capota, M., Turek, J., Chen, P.-H., Zhu, X., Manning, J. R., Sundaram, N., . . . Shin, Y. S. (2017). *Brain imaging analysis kit.* Retrieved from `https://doi.org/10.5281/zenodo.59780`

Christakis, N. A., & Fowler, J. H. (2010). Social network sensors for early detection of contagious outbreaks. *PLoS One*, *5*(9), e12948.

Comon, P., Jutten, C., & Herault, J. (1991). Blind separation of sources, part II: Problems statement. *Signal Processing*, *24*(1), 11 - 20.

Demertzi, A., Tagliazucchi, E., Dehaene, S., Deco, G., Barttfeld, P., Raimondo, F., . . . Sitt, J. D. (2019). Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Science Advances*, *5*(2), eaat7603.

Estrada, E., & Rodríguez-Velázquez, J. A. (2005). Subraph centrality in complex networks. *Physical Review E*, *71*(5), 056103.

Etzel, J. A., Gazzola, V., & Keysers, C. (2009). An introduction to anatomical ROI-based fMRI classification. *Brain Research*, *1282*, 114–125.

Fong, A. H. C., Yoo, K., Rosenberg, M. D., Zhang, S., Li, C.-S. R., Scheinost, D., . . . Chun, M. M. (2019). Dynamic functional connectivity during task performance and rest predictsindividual differences in attention across studies. *NeuroImage*, *188*, 14–25.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*(1), 35–41.

Geisberger, R., Sanders, P., & Schultes, D. (2008). Better approximation of betweenness centrality. *Proceedings of the meeting on Algorithm Engineering and Experiments*, 90–100.

Gershman, S., Blei, D., Pereira, F., & Norman, K. (2011). A topographic latent source model for fMRI data. *NeuroImage*, *57*, 89–100.

Halu, A., Mondragón, R. J., Panzarasa, P., & Bianconi, G. (2013). Multiplex PageRank. *PLoS One*, *8*(10), e78293.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507.

Honey, C. J., Kötter, R., Breakspear, M., & Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Science USA*, *104*(24), 10240–10245.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*, 453–458.

Huth, A. G., Nisimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210-1224.

Jutten, C., & Herault, J. (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, *24*(1), 1–10.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*, 679–685.

Landau, E. (1895). Zur relativen Wertbemessung der Turnierresultate. *Deutsches Wochenschach*, *11*, 366–369.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*, 788–791.

Lin, J. (2009). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, *37*(1), 145–151.

Lohmann, G., Margulies, D. S., Horstmann, A., Pleger, B., Lepsien, J., Goldhahn, D., . . . Turner, R. (2010). Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain. *PLoS One*, *5*(4), e10232.

Mairal, J., Ponce, J., Sapiro, G., Zisserman, A., & Bach, F. R. (2009). Supervised dictionary learning. *Advances in Neural Information Processing Systems*, 1033–1040.

Mairal, J. B., Bach, F., Ponce, J., & Sapiro, G. (2009). Online dictionary learning for sparse coding. *Proceedings of the 26th annual international conference on machine learning*, 689–696.

Manning, J. R., Ranganath, R., Norman, K. A., & Blei, D. M. (2014). Topographic factor analysis: a Bayesian model for inferring brain networks from neural data. *PLoS One*, *9*(5), e94914.

Manning, J. R., Zhu, X., Willke, T. L., Ranganath, R., Stachenfeld, K., Hasson, U., . . . Norman, K. A. (2018). A probabilistic approach to discovering dynamic full-brain functional connectivity patterns. *NeuroImage*, *180*, 243–252.

McInnes, L., & Healy, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv*, *1802*(03426).

Mitchell, T., Shinkareva, S., Carlson, A., Chang, K., Malave, V., Mason, R., & Just, M. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191.

Newman, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, *27*, 39–54.

Newman, M. E. J. (2008). The mathematics of networks. *The New Palgrave Encyclopedia of Economics*, *2*, 1–12.

Nishimoto, S., Vu, A., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. (2011). Reconstructing visual experience from brain activity evoked by natural movies. *Current Biology*, *21*, 1 - 6.

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization*. New York: Springer.

Norman, K. A., Newman, E., Detre, G., & Polyn, S. M. (2006). How inhibitory oscillations can train neural networks and punish competitors. *Neural Computation*, *18*, 1577–1610.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: generalizing degree and shortest paths. *Social Networks*, *32*, 245–251.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*, 559-572.

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., . . . Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nature Communications*, *9*(963), 1–13.

Rao, C. R. (1982). Diversity and dissimilarity coeficients: a unified approach. *Theoretical Population Biology*, *21*(1), 24–43.

Ricotta, C., & Szeidl, L. (2006). Towards a unifying approach to diversity measures: Bridging the gap between the Shannon entropy and Rao's quadratic index. *Theoretical Population Biology*, *70*(3), 237–243.

Rubin, T. N., Kyoejo, O., Gorgolewski, K. J., Jones, M. N., Poldrack, R. A., & Yarkoni, T. (2017). Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. *PLoS Computational Biology*, *13*(10), e1005649.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, *52*, 1059–1069.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, *85*(2), 461–464.

Simony, E., Honey, C. J., Chen, J., & Hasson, U. (2016). Uncovering stimulus-locked network dynamics during narrative comprehension. *Nature Communications*, *7*(12141), 1–13.

Spearman, C. (1904). General intelligence, objectively determined and measured. *Americal Journal of Psychology*, *15*, 201–292.

Sporns, O., & Honey, C. J. (2006). Small worlds inside big brains. *Proceedings of the National Academy of Science USA*, *103*(51), 19219–19220.

Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of Royal Statistical Society, Series B*, *61*(3), 611–622.

Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology*, *63*, 483–509.

Turk-Browne, N. B. (2013). Functional interactions as big data in the human brain. *Science*, *342*, 580–584.

van der Maaten, L. J. P., & Hinton, G. E. (2008). Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579-2605.

Zar, J. H. (2010). *Biostatistical analysis*. Prentice-Hall/Pearson.