

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

3 Lucy L. W. Owen¹, Thomas H. Chang^{1,2}, and Jeremy R. Manning^{1,†}

¹Department of Psychological and Brain Sciences,
Dartmouth College, Hanover, NH

³Amazon.com, Seattle, WA

[†]Address correspondence to jeremy.r.manning@dartmouth.edu

4 August 30, 2019

5 **Abstract**

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

20 **Introduction**

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

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

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

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

58 To gain insights into the above aspects of the neural code, we developed a computational framework
59 for estimating dynamic high-order correlations in timeseries data. This framework provides an important
60 advance, in that it enables us to examine patterns in higher-order correlations that are computationally
61 intractable to estimate via conventional methods. Given a multivariate timeseries, our framework provides
62 timepoint-by-timepoint estimates of the first-order correlations, second-order correlations, and so on (up to
63 tenth-order correlations in this manuscript). Our approach combines a kernel-based method for computing

64 dynamic correlations in timeseries data with a dimensionality reduction step that projects the resulting dy-
65 namic correlations into a low-dimensional space. We explored two dimensionality reduction approaches:
66 principle components analysis (PCA; Pearson, 1901), which preserves an approximately invertable transfor-
67 mation back to the original data; and a second non-invertible algorithm that explored patterns in eigenvector
68 centrality (Landau, 1895). This latter approach characterizes correlations between each feature dimension's
69 relative *position* in the network in favor of the specific activity histories of different features.

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

83 Methods

84 Our general approach to comprises four general steps (Fig. 1). First, we derive a kernel-based approach
85 to computing dynamic pairwise correlations in a T (timepoints) by K (features) multivariate timeseries,
86 \mathbf{X}_0 . This yields a T by $O(K^2)$ matrix of dynamic correlations, \mathbf{Y}_1 , where each row comprises the upper
87 triangle of the correlation matrix at a single timepoint, reshaped into a row vector (this reshaped vector is
88 $(\frac{K^2-K}{2})$ -dimensional). Second, we apply a dimensionality reduction step to project the matrix of dynamic
89 correlations back onto a K -dimensional space. This yields a T by K matrix, \mathbf{X}_1 , that reflects an approximation
90 of the dynamic correlations reflected in the original data. Third, we use repeated applications of the kernel-
91 based dynamic correlation step to \mathbf{X}_n and the dimensionality reduction step to the resulting \mathbf{Y}_{n+1} to estimate
92 high-order dynamnic correlations. Each application of these steps to a T by K time series \mathbf{X}_n yields a T by K
93 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
94 the *order* of the timeseries, where \mathbf{X}_0 (order 0) denotes the original data and \mathbf{X}_n denotes n^{th} -order dynamic

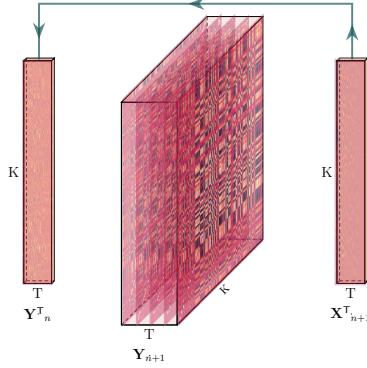


Figure 1: **Estimating dynamic high-order correlations.** Given a T by K matrix of multivariate timeseries data, \mathbf{Y}_n (where $n \in \mathbb{N}, n \geq 0$), we use Equation 5 to compute a timeseries of K by K correlation matrices, \mathbf{Y}_{n+1} . We then approximate \mathbf{Y}_{n+1} with the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

95 correlations between the columns of \mathbf{X}_0 . Finally, we use a cross-validation-based decoding approach to
 96 evaluate how well information contained in a given order (or weighted mixture of orders) may be used
 97 to decode relevant cognitive states. If including a given \mathbf{X}_n in the feature set yields higher classification
 98 accuracy on held-out data, we interpret this as evidence that the given cognitive states are reflected in
 99 patterns of n^{th} -order correlations. All of the code used to produce the figures and results in this manuscript,
 100 along with links to the corresponding datasets, may be found at github.com/ContextLab/timecorr-paper. In
 101 addition, we have released a Python toolbox for computing dynamic high-order correlations in timeseries
 102 data; our toolbox may be found at timecorr.readthedocs.io. **JRM NOTE: CHECK LINK**

103 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):

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

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

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

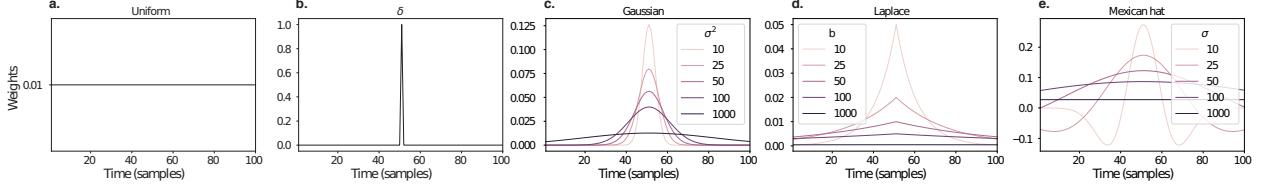


Figure 2: Examples of kernel functions. Each panel displays per-timepoint weights at $t = 50$, evaluated for 100 timepoints ($\tau \in [1, \dots, 100]$). **a. Uniform kernel.** The weights are timepoint-invariant; observations at all timepoints are weighted equally, and do not change as a function of 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 variance 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 scale 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.

104 We can generalize this formula to compute time-varying correlations by incorporating a *kernel function* that
 105 takes a time t as input, and returns how much the observed data at each timepoint $\tau \in [-\infty, \infty]$ contributes
 106 to the estimated instantaneous correlation at time t (Fig. 2).

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

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

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

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

107 Here $\text{timecorr}(\mathbf{X}(\cdot, i), \mathbf{X}(\cdot, j), \kappa_t)$ reflects the correlation at time t between columns i and j of \mathbf{X} , estimated using
 108 the kernel κ_t .

109 **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:

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

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

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

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

112 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-theoretic 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.

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

119 The modern library of dimensionality reduction algorithms include Principal Components Analysis (PCA;
120 Pearson, 1901), Probabilistic PCA (PPCA; Tipping & Bishop, 1999), Exploratory Factor Analysis (EFA;
121 Spearman, 1904), Independent Components Analysis (ICA; Comon et al., 1991; Jutten & Herault, 1991),
122 *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 &
124 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),
126 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
128 of the present manuscript, the general intuition driving these approaches is to compute the T by I matrix,
129 \mathbf{X} , that is closest to the original T by J matrix, \mathbf{Y} , where (typically) $I \ll J$. The different approaches place
130 different constraints on what properties \mathbf{X} must satisfy and which aspects of the data are compared (and
131 how) to characterize the match between \mathbf{X} and \mathbf{Y} .

132 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,
134 with each repeated dimensionality reduction, the resulting \mathbf{X}_n has lower and lower fidelity (with respect to
135 what the “true” \mathbf{Y}_n might have looked like without using dimensionality reduction to maintain scalability).
136 In other words, computing \mathbf{X}_n is a lossy operation. Second, whereas each columns of \mathbf{Y}_n may always be
137 mapped directly onto specific pairs of columns of \mathbf{Y}_{n-1} , the columns of \mathbf{X}_n reflect weighted combinations
138 and/or nonlinear transformations of the columns of \mathbf{Y}_n . Many dimensionality reduction algorithms are
139 invertible (or approximately invertible). However, attempting to map a given \mathbf{X}_n back onto the original
140 feature space of \mathbf{Y}_0 will usually require $O(TK^{2n})$ space and therefore quickly becomes intractable as n or K
141 grow large.

142 **Graph theoretic approaches to computing \mathbf{X}_n**

143 The above dimensionality reduction approaches to approximating a given \mathbf{Y}_n with a lower-dimensional
144 \mathbf{X}_n preserve a (potentially recombined and transformed) mapping back to the original data in \mathbf{Y}_0 . We
145 also explore graph theoretic approaches that forgo a preserved mapping back to the original data in favor
146 of preserving each feature’s relative *position* in the broader network of interactions and connections. To
147 illustrate the distinction between the two general approaches we explore, suppose a network comprises
148 nodes A , B , and C . If A and B exhibit uncorrelated activity patterns, the functional connection between

149 them will be (by definition) close to 0. However, if A and B each interact with C in similar ways, we might
150 attempt to capture those similarities using a measure that reflects the how A and B interact in the network.

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

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

172 Higher-order correlations

173 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
174 enables us to estimate high-order dynamic correlations in a scalable way. Given a T by K input matrix, the
175 output of Equation 5 requires $O(TK^2)$ space to store. Repeated applications of Equation 5 (i.e., computing
176 dynamic correlations between the columns of the outputted dynamic correlation matrix) each require
177 exponentially more space; in general the n^{th} -order dynamic correlations of a T by K timeseries occupies
178 $O(TK^{2n})$ space. However, when we approximate or summarize the output of Equation 5 with a T by K matrix
179 (as described above), it becomes feasible to compute even very high-order correlations in high-dimensional

180 data. Specifically, approximating the n^{th} -order dynamic correlations of a T by K timeseries requires only
 181 $O(TK^2)$ additional space— the same as would be required to compute first-order dynamic correlations. In
 182 other words, the space required to store $n + 1$ multivariate timeseries reflecting up to n^{th} order correlations
 183 in the original data scales linearly with n using our approach (Fig. 1).

184 Data

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

193 Synthetic data

194 We constructed a total of 40 multivariate timeseries, collectively reflecting a total of 4 different patterns of
 195 dynamic correlations (i.e., 10 datasets reflecting each type of dynamic pattern). Each timeseries comprised
 196 50 features (dimensions) that varied over 300 timepoints. The observations at each timepoint were drawn
 197 from a zero-mean multivariate Gaussian distribution with a covariance matrix defined for each timepoint
 198 as described below. We drew the observations at each timepoint independently from the draws at all other
 199 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 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, Σ_m :

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

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

200 $i, j \in [1, 2, \dots, 50]$. In other words, all of the observations (for each of the 300 timepoints) within each dataset
 201 were drawn from a multivariate Gaussian distribution with the same covariance matrix, and the 10 datasets

202 each used a different covariance matrix.

203 **Random.** We generated a second set of 10 synthetic datasets whose observations at each timepoint were
204 drawn from a Gaussian distribution with a new randomly constructed (using Eqn. 12) covariance matrix.
205 Because each timepoint’s covariance matrix was drawn independently of the covariance matrices for all
206 other timepoints, these datasets provided a test of reconstruction accuracy in the absence of any meaningful
207 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, Σ_{start} and Σ_{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

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

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

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

216 Functional neuroimaging data collected during story listening

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

225 **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
226 timeseries of voxel activations (44,415 voxels) using a much smaller number of radial basis function nodes (in
227 this case 700 nodes). This provides a convenient representation for examining full-brain network dynamics.
228 All of the analyses we carried out on the neuroimaging dataset were performed in this lower-dimensional
229 space. In other words, each participant’s data matrix, \mathbf{Y}_0 , was a number-of-timepoints by 700 matrix of
230 HTFA-derived factor weights (where the row and column labels were matched across participants). Code
231 for carrying out HTFA on fMRI data may be found as part of the BrainIAK toolbox (Capota et al., 2017),
232 which may be downloaded at brainiak.org.
233

234 **Temporal decoding**

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

247 **Forward inference and decoding accuracy**

248 We used an across-participants correlation-based classifier to decode which stimulus timepoint matched a
249 given neural pattern. We first divided the participants into two groups: a template group, $\mathcal{G}_{\text{template}}$, and a
250 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
251 $\bar{\mathbf{C}}_{\text{decode}}$, 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
252 number-of-timepoints decoding matrix, Λ . In this way, the rows of Λ reflected timepoints from the template
253 group, while the columns reflected timepoints from the to-be-decoded group. We assigned temporal labels
254 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

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

258 **Feature weighting and testing**

259 We sought to examine which types of neural features (i.e., activations, first-order dynamic correlations, and
260 higher-order dynamic correlations) were informative to the temporal decoders. Using the notation above,
261 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}).

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

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

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

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

286 neural features. In particular, we computed the decoding accuracy for each of 10 random group assignments
287 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.

288 **Identifying robust decoding results**

289 For synthetic data where the true underlying dynamic correlations are known, we are able to directly
290 evaluate how well the recovered correlations match the ground truth, for different choices of kernel shape
291 and width. For real-world data (where the ground truth is not known), we use decoding accuracy to estimate
292 whether a given set of neural features reliably reflects ongoing stimulus-driven cognitive processing. We
293 cannot know which choice of kernel most accurately reflects the “true” information content of the data. For
294 example, a given kernel might highlight particular temporal aspects of the data in a way that enhances or
295 diminishes decoding performance. **JRM STOPPED HERE**

296 Params: Gaussian, Laplace, mexican hat kernels. Widths of 5, 10, 20, and 50.

297 **Reverse inference**

298 top 10 strongest correlations for each level; identify endpoint nodes (RBFs); make map (threshold at θ)—JRM
299 **NOTE: what threshold did you use, Lucy?); decode using neurosynth (cite).**

300 **Results**

301 **Synthetic data**

302 To assess the performance of dynamic correlation recovery using timecorr, we varied width the kernel and
303 the specific structure of the data. We applied timecorr, using delta and gaussian kernels Fig. 2) to each of
304 the following synthetic datasets: constant, random, ramping, and block. We then correlated each recovered
305 correlation matrix with the ground truth.

306 For the constant synthetic dataset, a gaussian kernel (width=10) outperformed the delta kernel (Fig. 3,
307 a.). This is in contrast with the random synthetic dataset, for which the delta kernel best captures the rapidly
308 changing structure (Fig. 3, b.). For the ramping synthetic dataset, the slow changing strucutre within the
309 data is best captured by the gaussian kernel and the best recovery occurs in the middle (Ramping, Fig. 3,
310 c.). In addition to comparing the timecorr recovered correlation matrices to the ground truth, we further
311 compared the ramping recovered correlation matrices to only the first random covariance matrix K_1 (First,
312 Fig. 3, c.) and to only the last random covariance matrix K_2 (Last, Fig. 3, c.), both of which perform best at
313 the beginning and end respectively.

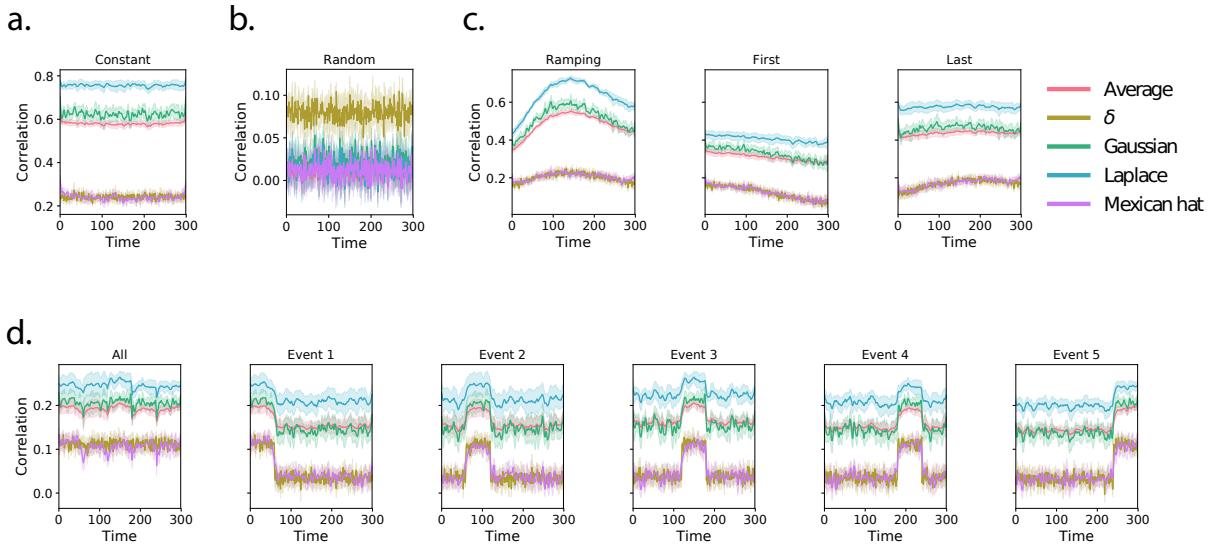


Figure 3: Dynamic correlation recovery with synthetic data. Using synthetic data containing different underlying correlational structure, we test how well we can recover dynamic correlation matrices using different kernels when compared to ground truth. We compare the results using a delta kernel with averaged results from several kernels (Gaussian, Laplace, and mexican hat) and several widths (5, 10, 20, and 50). We plot recovery using of datasets containing the following underlying structure: **a. Constant**. **b. Random**. **c. Ramping**. **d. Block**.

314 Similary for the block sythetic dataset, we compared the timecorr recovered correlation matrices to
 315 the ground truth as well as to each block-specific covariance matrix (Block 1-5, Fig. 3, d.). Although the
 316 structure is changing by block, the gaussian kernel once again outperforms the delta kernel. Performance
 317 does however drop near even boundaries for when using the gaussian kernel.

318 We next evaluated if our model of high-order correlations in brain activity can capture cognitively
 319 relevant brain patterns. We performed a decoding analysis, using cross validation to estimate (using other
 320 participants' data) which parts of the story each weighted-mixture of higher-order brain activity pattern
 321 corresponded to (see *Materials and methods*). We note that our primary goal was not to achieve perfect
 322 decoding accuracy, but rather to use decoding accuracy as a benchmark for assessing whether different
 323 neural features specifically capture cognitively relevant brain patterns.

324 Separately for each experimental condition, we divided participants into two groups. For the zeroth
 325 order, we computed the mean factor activity for each group. For all subsequent orders up to the tenth
 326 order, we computed the mean approximated dynamic ISFC of factor activity for each group (see *Materials*
 327 and *methods*), and combined in a weighted mixutre with all previous orders (i.e. cross-validation for the
 328 second order contained a weighted-mixture of zeroth, first, and second order (Fig. 4, c.&f.). For each
 329 order, we correlated the group 1 activity patterns with group 2 activity patterns. We then subdivided

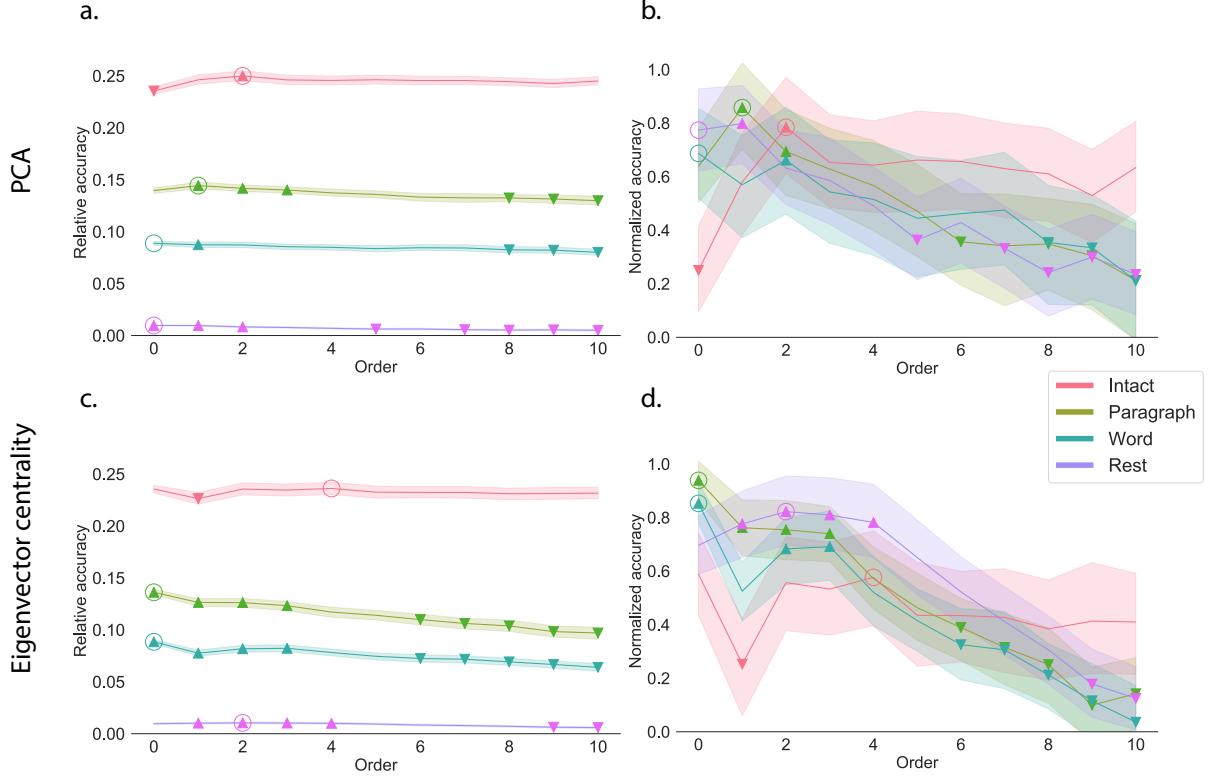


Figure 4: Decoding by order. a.&d. Relative decoding accuracy by order. Ribbons of each color display cross-validated decoding performance averaged over all parameters for each condition (intact, paragraph, word, and rest) using PCA (a.) or eigenvector centrality (d.) to approximate correlations. Decoders were trained using increasingly more higher-order information and this ribbons are displayed relative to chance at 0. The red dots indicates maximum decoding accuracy for each condition. **b.&e. Z-transformed decoding accuracy by order.** We Z-transformed the decoding accuracy by order to better visualize the order with the maximum decoding accuracy for each condition using PCA (b.) or eigenvector centrality (e.) to approximate correlations. **c.&f. Optimized weights.** Bar heights indicate the optimized mixing parameter ϕ of each contributing order up to and including the order with the maximum decoding accuracy for each contributing order using PCA (c.) or eigenvector centrality (f.) to approximate correlations. For the order with maximum decoding accuracy by condition, we show barplots of the optimized weights ϕ for each contributing order.

330 the group 1 to obtain an optimal weighting parameter for each order's correlation matrix using the same
331 cross validation method. We used the optimal weighting parameters to obtain a weighted-mixture (see
332 *Materials and methods*) of each order's correlation matrix. Using these correlations, we labeled the group 1
333 timepoints using the group 2 timepoints with which they were most highly correlated; we then computed
334 the proportion of correctly labeled group 1 timepoints. (We also performed the symmetric analysis whereby
335 we labeled the group 2 timepoints using the group 1 timepoints as a template.) We repeated this procedure
336 100 times (randomly re-assigning participants to the two groups each time) to obtain a distribution of
337 decoding accuracies for each experimental condition. There were 272 timepoints for paragraph condition,
338 300 timepoints for intact and word conditions, and 400 timepoints for rest condition, so chance performance
339 on this decoding test is was $\frac{1}{272}$, $\frac{1}{300}$, and $\frac{1}{400}$ respectively.

340 We repeated this process for each set of parameters, varying kernel type and width, and averaged over the
341 reduction technique used to approximate the higher-order correlations (PCA Fig. 4, a.-c. and eigenvector
342 centrality Fig. 4, d.-f.). Since there is no ground truth in these analyses, and we did not know which
343 parameters best capture the data, we instead report a robustness search by averaging over the parameters
344 and reporting which results consistently showed up across all parameters.

345 The two methods used to approximate the higher-order correlations (PCA Fig. 4, a.-c. and eigenvector
346 centrality Fig. 4, d.-f.) capture different facets of the activity patterns. Using PCA, the higher-order
347 correlations are all linked to the original activity patterns, whereas eigenvector centrality breaks the
348 immediate link with specific brain areas and instead characterizes the position of the nodes in the network
349 that are similar over time.

350 We found for both PCA and eigenvector centrality, during the intact condition in the experiment,
351 classifiers that incorporated higher-order correlations yielded consistently higher accuracy than classifiers
352 trained only on lower-order patterns (Fig. 4, a.&d.). We plot the average correlations for up to the fourth
353 order for the intact condition (Fig. 5) representing the degree of agreement by location pair over time. By
354 contrast, we found that incorporating higher-order (greater than first order) correlations did not further
355 improve decoding accuracy for the other listening conditions or rest condition. This suggests that the
356 cognitive processing that supported the most cognitively rich condition involved higher-order network
357 dynamics.

358 Discussion

- 359 • Methods advances: kernel-based dynamic correlations, extension to dynamic ISFC, efficient method
360 for estimating high-order dynamic correlations, identifying robust results by averaging

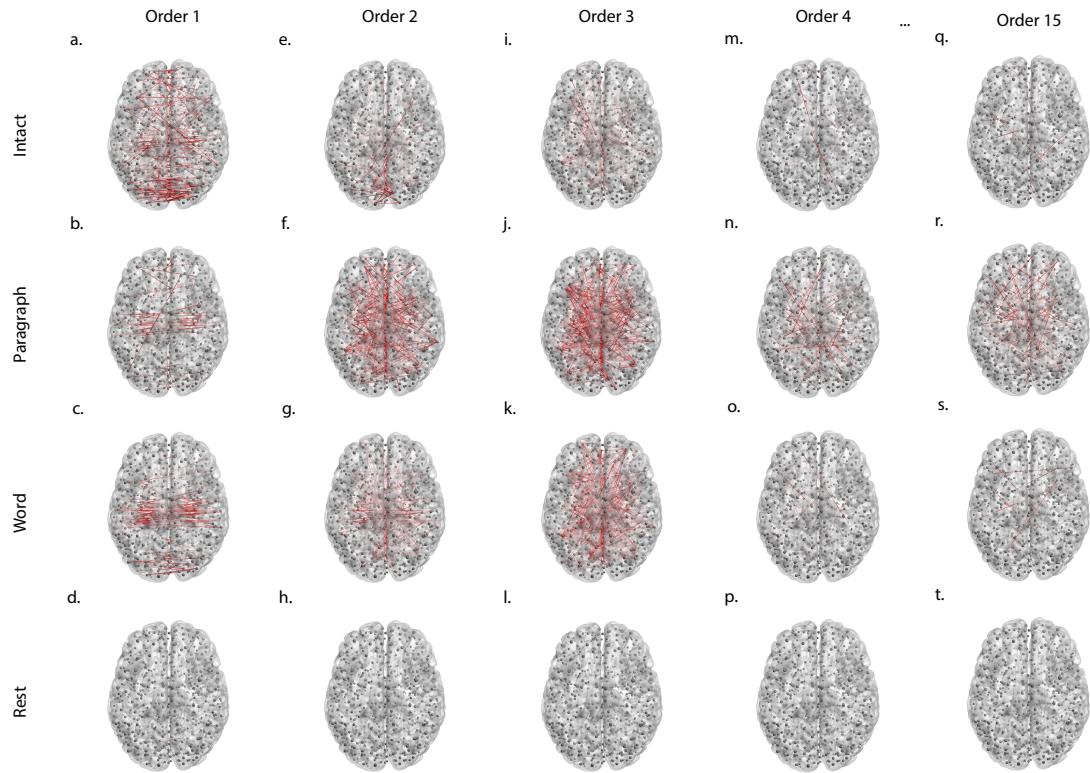


Figure 5: Average correlations by order for the intact listening condition. Using eigenvector centrality to approximate higher-order correlations for the intact, paragraph scrambled, word scrambled, and rest condition. We plot the strongest 50% absolute value mean correlation for **a.-d. first order, e.-h. second order, i.-l. third order, and m.-p. fourth order**, representing the degree of agreement by location pair over time. To demonstrate how this method is computationally scalable, we also approximated **a.-d. fifteenth order** dynamic correlation, which would be possible to compute using conventional methods since it would require more bits to represent the solution than there are molecules in the universe.

- 361 • Discoveries:
- 362 – Dimensionality reduction and graph theoretic approaches give different insights into the data
363 and identify different patterns as being relevant to cognition (different peak orders).
- 364 – An insight common to both approaches is that high-order (greater than first order) dynamic
365 correlations are informative about ongoing high-level cognitive processing. As the level of
366 cognitive processing decreases, cognition is reflected by lower-order correlations.
- 367 – Correlations at different orders are also associated with different networks of brain regions. How-
368 ever, which networks reflect which types of interactions depends on the current task. In general,
369 lower order correlations during auditory listening reflect processing of low-level (auditory) fea-
370 tures; mid-order correlations reflect speech and linguistic processing; higher-order correlations
371 reflect across-sensory integration (e.g. ties to visual areas) and cognitive control areas. This
372 hierarchy dissolves during lower-order cognitive processing.

373 Based on prior work (Demertzis et al., 2019) and following the direction of the field (Turk-Browne, 2013)
374 we think our thoughts might be encoded in dynamic network patterns, and possibly higher order network
375 patterns (Fig. 6). We sought to test this hypothesis by developing an approach to inferring high-order
376 network dynamics from timeseries data.

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

382 We first validated our model using synthetic data, and explored how recovery varied with different
383 underlying data structures and kernels. We then applied the approach to an fMRI dataset (Simony et al.,
384 2016) in which participants listened to an audio recording of a story, as well as scrambled versions of the
385 same story (where the scrambling was applied at different temporal scales). We trained classifiers to take
386 the output of the model and decode the timepoint in the story (or scrambled story) that the participants
387 were listening to. We found that, during the intact listening condition in the experiment, classifiers that
388 incorporated higher-order correlations yielded consistently higher accuracy than classifiers trained only
389 on lower-order patterns (Fig. 4, a.&d.). By contrast, these higher-order correlations were not necessary
390 to support decoding the other listening conditions and (minimally above chance) during a control rest
391 condition. This suggests that the cognitive processing that supported the most cognitively rich listening
392 conditions involved second-order (or higher) network dynamics.

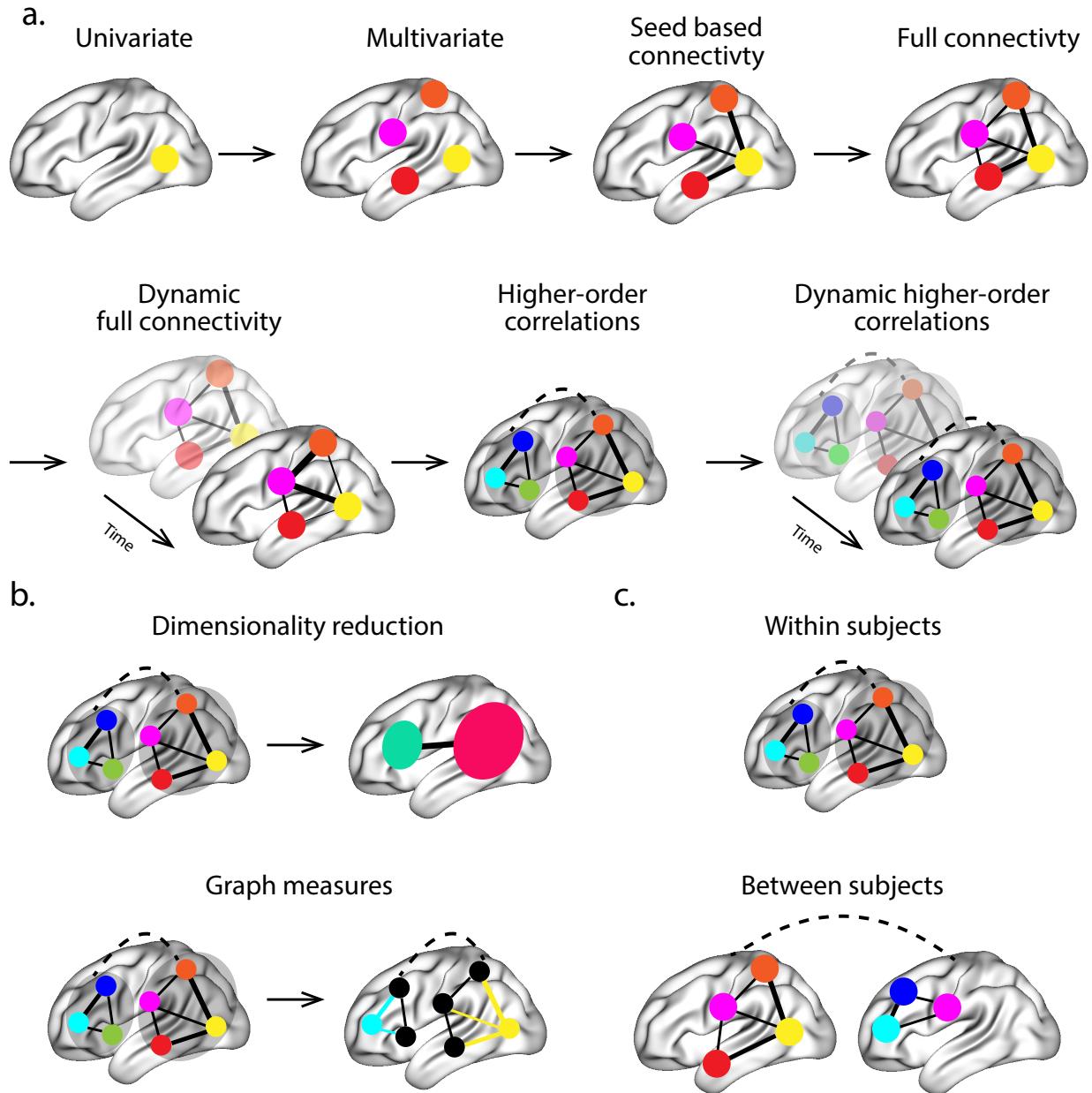


Figure 6: Direction of the field (adapted from (Turk-Browne, 2013)). The evolution of fMRI analyses started with univariate activation, which refers to the average amplitude of BOLD activity evoked by events of an experimental condition. Next, multivariate classifiers are trained on patterns of activation 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.

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

397 **Concluding remarks**

398 How can we better understand how brain patterns change over time? How can we quantify the potential
399 network dynamics that might be driving these changes? One way to judge the techniques of the future is
400 to look at the trajectory of the fMRI field so far has taken so far (Fig. 1). The field started with univariate
401 activation, measuring the average activity for each voxel. Analyses of multivariate activation followed,
402 looking at spatial patterns of activity over voxels. Next, correlations of activity were explored, first with
403 measures like resting connectivity that take temporal correlation between a seed voxel and all other voxels
404 then with full connectivity that measure all pairwise correlations. Additionally, this path of increasing
405 complexity also moved from static to dynamic measurements. One logical next step in this trajectory would
406 be dynamic higher-order correlations. We have created a method to support these calculations by scalably
407 approximating dynamic higher-order correlations.

408 **Acknowledgements**

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

415 **Author contributions**

416 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.

417 **References**

- 418 Alvarez-Hamelin, I., Dall'Asta, L., Barrat, A., & Vespignani, A. (2005). *k*-corr decomposition: a tool for the
419 visualization of large scale networks. *arXiv*, cs/0504107v2.
- 420 Barthélémy, M. (2004). Betweenness centrality in large complex networks. *European Physical Journal B*, 38,
421 163–168.
- 422 Bassett, D., Meyer-Lindenberg, A., Achard, S., Duke, T., & Bullmore, E. (2006, December). Adaptive
423 reconfiguration of fractal small-world human brain functional networks. *Proceedings of the National
424 Academy of Sciences, USA*, 103(51), 19518-23.
- 425 Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, 29(4), 555–564.
- 426 Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and
427 functional systems. *Nature Reviews Neuroscience*, 10(3), 186–198.
- 428 Capota, M., Turek, J., Chen, P.-H., Zhu, X., Manning, J. R., Sundaram, N., . . . Shin, Y. S. (2017). *Brain imaging
429 analysis kit*. Retrieved from <https://doi.org/10.5281/zenodo.59780>
- 430 Christakis, N. A., & Fowler, J. H. (2010). Social network sensors for early detection of contagious outbreaks.
431 *PLoS One*, 5(9), e12948.
- 432 Comon, P., Jutten, C., & Herault, J. (1991). Blind separation of sources, part II: Problems statement. *Signal
433 Processing*, 24(1), 11 - 20.
- 434 Demertzi, A., Tagliazucchi, E., Dehaene, S., Deco, G., Barttfeld, P., Raimondo, F., . . . Sitt, J. D. (2019). Human
435 consciousness is supported by dynamic complex patterns of brain signal coordination. *Science Advances*,
436 5(2), eaat7603.
- 437 Estrada, E., & Rodríguez-Velázquez, J. A. (2005). Subgraph centrality in complex networks. *Physical Review
438 E*, 71(5), 056103.
- 439 Etzel, J. A., Gazzola, V., & Keysers, C. (2009). An introduction to anatomical ROI-based fMRI classification.
440 *Brain Research*, 1282, 114–125.
- 441 Fong, A. H. C., Yoo, K., Rosenberg, M. D., Zhang, S., Li, C.-S. R., Scheinost, D., . . . Chun, M. M. (2019).
442 Dynamic functional connectivity during task performance and rest predicts individual differences in
443 attention across studies. *NeuroImage*, 188, 14–25.
- 444 Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 35–41.

- 445 Geisberger, R., Sanders, P., & Schultes, D. (2008). Better approximation of betweenness centrality. *Proceedings*
446 of the meeting on Algorithm Engineering and Experiments, 90–100.
- 447 Gershman, S., Blei, D., Pereira, F., & Norman, K. (2011). A topographic latent source model for fMRI data.
448 *NeuroImage*, 57, 89–100.
- 449 Halu, A., Mondragón, R. J., Panzarasa, P., & Bianconi, G. (2013). Multiplex PageRank. *PLoS One*, 8(10),
450 e78293.
- 451 Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and
452 overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430.
- 453 Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks.
454 *Science*, 313(5786), 504–507.
- 455 Honey, C. J., Kötter, R., Breakspear, M., & Sporns, O. (2007). Network structure of cerebral cortex shapes
456 functional connectivity on multiple time scales. *Proceedings of the National Academy of Science USA*, 104(24),
457 10240–10245.
- 458 Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals
459 the semantic maps that tile human cerebral cortex. *Nature*, 532, 453–458.
- 460 Huth, A. G., Nisimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes
461 the representation of thousands of object and action categories across the human brain. *Neuron*, 76(6),
462 1210–1224.
- 463 Jutten, C., & Herault, J. (1991). Blind separation of sources, part I: An adaptive algorithm based on
464 neuromimetic architecture. *Signal Processing*, 24(1), 1–10.
- 465 Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature*
466 *Neuroscience*, 8, 679–685.
- 467 Landau, E. (1895). Zur relativen Wertbemessung der Turnierresultate. *Deutsches Wochenschach*, 11, 366–369.
- 468 Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*,
469 401, 788–791.
- 470 Lin, J. (2009). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*,
471 37(1), 145–151.

- 472 Lohmann, G., Margulies, D. S., Horstmann, A., Pleger, B., Lepsius, J., Goldhahn, D., ... Turner, R. (2010).
473 Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain.
474 *PLoS One*, 5(4), e10232.
- 475 Mairal, J., Ponce, J., Sapiro, G., Zisserman, A., & Bach, F. R. (2009). Supervised dictionary learning. *Advances
476 in Neural Information Processing Systems*, 1033–1040.
- 477 Mairal, J. B., Bach, F., Ponce, J., & Sapiro, G. (2009). Online dictionary learning for sparse coding. *Proceedings
478 of the 26th annual international conference on machine learning*, 689–696.
- 479 Manning, J. R., Ranganath, R., Norman, K. A., & Blei, D. M. (2014). Topographic factor analysis: a Bayesian
480 model for inferring brain networks from neural data. *PLoS One*, 9(5), e94914.
- 481 Manning, J. R., Zhu, X., Willke, T. L., Ranganath, R., Stachenfeld, K., Hasson, U., ... Norman, K. A. (2018).
482 A probabilistic approach to discovering dynamic full-brain functional connectivity patterns. *NeuroImage*,
483 180, 243–252.
- 484 McInnes, L., & Healy, J. (2018). UMAP: Uniform manifold approximation and projection for dimension
485 reduction. *arXiv*, 1802(03426).
- 486 Mitchell, T., Shinkareva, S., Carlson, A., Chang, K., Malave, V., Mason, R., & Just, M. (2008). Predicting
487 human brain activity associated with the meanings of nouns. *Science*, 320(5880), 1191.
- 488 Newman, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, 27,
489 39–54.
- 490 Newman, M. E. J. (2008). The mathematics of networks. *The New Palgrave Encyclopedia of Economics*, 2, 1–12.
- 491 Nishimoto, S., Vu, A., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. (2011). Reconstructing visual
492 experience from brain activity evoked by natural movies. *Current Biology*, 21, 1 - 6.
- 493 Nocedal, J., & Wright, S. J. (2006). *Numerical optimization*. New York: Springer.
- 494 Norman, K. A., Newman, E., Detre, G., & Polyn, S. M. (2006). How inhibitory oscillations can train neural
495 networks and punish competitors. *Neural Computation*, 18, 1577–1610.
- 496 Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: generalizing degree
497 and shortest paths. *Social Networks*, 32, 245–251.
- 498 Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh,
499 and Dublin Philosophical Magazine and Journal of Science*, 2, 559-572.

- 500 Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., ... Fedorenko, E. (2018). Toward
501 a universal decoder of linguistic meaning from brain activation. *Nature Communications*, 9(963), 1–13.
- 502 Rao, C. R. (1982). Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology*,
503 21(1), 24–43.
- 504 Ricotta, C., & Szeidl, L. (2006). Towards a unifying approach to diversity measures: Bridging the gap
505 between the Shannon entropy and Rao's quadratic index. *Theoretical Population Biology*, 70(3), 237–243.
- 506 Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpreta-
507 tions. *NeuroImage*, 52, 1059–1069.
- 508 Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85(2), 461–464.
- 509 Simony, E., Honey, C. J., Chen, J., & Hasson, U. (2016). Uncovering stimulus-locked network dynamics
510 during narrative comprehension. *Nature Communications*, 7(12141), 1–13.
- 511 Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of
512 Psychology*, 15, 201–292.
- 513 Sporns, O., & Honey, C. J. (2006). Small worlds inside big brains. *Proceedings of the National Academy of
514 Science USA*, 103(51), 19219–19220.
- 515 Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of Royal Statistical
516 Society, Series B*, 61(3), 611–622.
- 517 Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology*, 63,
518 483–509.
- 519 Turk-Browne, N. B. (2013). Functional interactions as big data in the human brain. *Science*, 342, 580–584.
- 520 van der Maaten, L. J. P., & Hinton, G. E. (2008). Visualizing high-dimensional data using t-SNE. *Journal of
521 Machine Learning Research*, 9, 2579–2605.
- 522 Zar, J. H. (2010). *Biostatistical analysis*. Prentice-Hall/Pearson.