

# An examination of the high-order dynamic interactions underlying multi-dimensional timeseries data

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

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

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

<sup>†</sup>Address correspondence to [jeremy.r.manning@dartmouth.edu](mailto:jeremy.r.manning@dartmouth.edu)

March 20, 2019

## Abstract

Most complex systems reflect dynamic interactions between myriad evolving components (e.g., interacting molecules, interacting brain systems, interacting individuals within a social network or ecological system, coordinated components within a mechanical or digital device, etc.). Despite that these interactions are central to the full system’s behavior (e.g., removing a component from the full system can change the entire system’s behavior), dynamic interactions cannot typically be directly measured. Rather, the interactions must be inferred through their hypothesized role in guiding the dynamics of system components. Here we use a model-based approach to inferring dynamic interactions from timeseries data. In addition to examining first-order interactions (e.g., between pairs of components) we also examine higher-order interactions (e.g., that characterize mirrored structure in the patterns of interaction dynamics displayed by different subsets of components). We apply our approach to two datasets. First, we use a synthetic dataset, for which the underlying dynamic interactions are known, to show that our model recovers those ground-truth dynamic interactions. We also apply our model to a neuroimaging dataset and show that the high-order dynamic interactions exhibited by brain data vary meaningfully as a function of the cognitive “richness” of the stimulus people are experiencing.

## Introduction

The dynamics of the observable universe are meaningful in three respects. First, the behaviors of the *atomic units* that exhibit those dynamics are highly interrelated. The actions of one unit typically have implications for one or more other units. In other words, there is non-trivial *correlational structure* defining how different units interact with and relate to each other. Second, that correlational structure is *hierarchical* in the sense that it exists on many spatiotemporal scales. The way one group of units interacts may relate to how another group of units interact, and the interactions between those groups may exhibit some rich structure. Third, the structure at each level of this correlational hierarchy changes from moment to moment, reflecting the “behavior” of the full system.

These three properties (rich correlations, hierarchical organization, and dynamics) are major hallmarks of many complex systems. For example, within a single cell, the cellular components interact at many

spatiotemporal scales, and those interactions change according to what that single cell is doing. Within a single human brain, the individual neurons interact within each brain structure, and the structures interact to form complex networks. The interactions at each scale vary according to the functions our brains are carrying out. And within social groups, interactions at different scales (e.g., between individuals, family units, communities, etc.) vary over time according to changing goals and external constraints.

Although many systems exhibit rich dynamic correlations at many scales, a major challenge to studying such patterns is that typically neither the correlations nor the hierarchical organizations of those correlations may be directly observed. Rather, these fundamental properties must be inferred indirectly by examining the observable parts of the system— e.g., the behaviors of the individual atomic units of that system. In the *Methods* section, we propose a series of mathematical operations that may be used to recover dynamic correlations at a range of scales (i.e., orders of interaction). In the *Results* section, we demonstrate how our approach may be applied to multi-dimensional timeseries data: a synthetic dataset where the underlying dynamic correlations are known (we use this dataset to validate our approach), and a neuroimaging dataset comprising data collected as participants listened to a story (Simony et al., 2016). In different experimental conditions in the neuroimaging study, participants listened to altered versions of the story that varied in cognitive richness: the intact story (fully engaging), a scrambled version of the story where the paragraphs were presented in a randomized order (moderately engaging), a second scrambled condition where the words were presented in a random order (minimally engaging), and a “rest” condition where the participants did not listen to any version of the story (control condition). We use the neuroimaging dataset to examine how higher-order structure in brain data varies as a function of the cognitive richness of the stimulus.

## Methods

There are two basic steps to our approach. In the first step, we take a number-of-timepoints ( $T$ ) by number-of-features ( $F$ ) *matrix of observations* ( $\mathbf{X}$ ) and we return a  $T$  by  $\frac{F^2-F}{2}$  *matrix of dynamic correlations* ( $\mathbf{Y}$ ). Here  $\mathbf{Y}_0$  describes, at each moment, how all of the features (columns of  $\mathbf{X}$ ) are inferred to be interacting. (Since the interactions are assumed to be non-recurrent and symmetric, only the upper triangle of the full correlation matrix is computed.) In the second step, we project  $\mathbf{Y}_0$  onto an  $F$ -dimensional space, resulting in a new  $T$  by  $F$  matrix  $\mathbf{Y}_1$ . Note that  $\mathbf{Y}_1$  contains information about the correlation dynamics present in  $\mathbf{X}$ , but represented in a compressed number of dimensions. By repeatedly applying these two steps in sequence, we can examine and explore higher order dynamic correlations in  $\mathbf{X}$ .

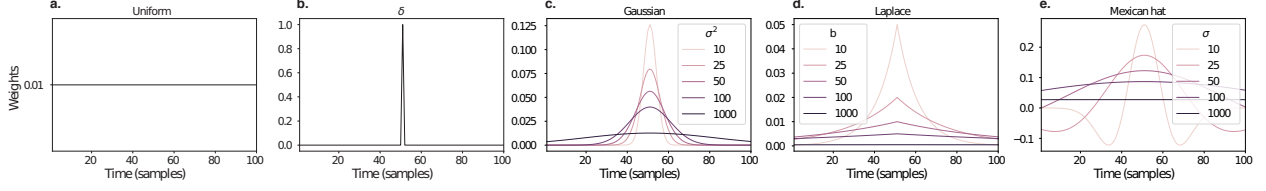


Figure 1: **Examples of time-varying weights.** Each panel displays per-timepoint weights at  $t = 50$ , evaluated for 100 timepoints (1, ..., 100). **a. Uniform weights.** The weights are timepoint-invariant; observations at all timepoints are weighted equally, and do not change as a function of  $t$ . This is a special case of weight function that reduces dynamic correlations to static correlations. **b. Dirac delta function.** Only the observation at timepoint  $t$  is given weight (of 1), and weights for observations at all other timepoints are set to 0. **c. Gaussian weights.** Each observation's weights fall off in time according to a Gaussian probability density function centered on  $\mu = t$ . Weights derived using several different example variance parameters ( $\sigma^2$ ) are displayed. **d. Laplace weights.** Each observation's weights fall off in time according to a Laplace probability density function centered on  $\mu = t$ . Weights derived using several different example scale parameters ( $b$ ) are displayed. **e. Mexican hat (Ricker wavelet) weights.** Each observation's weights fall off in time according to a Ricker wavelet centered on  $t$ . This function highlights the *contrasts* between local versus surrounding activity patterns in estimating dynamic correlations. Weights derived using several different example width parameters ( $\sigma$ ) are displayed.

## 60 Dynamic correlations

Given a matrix of observations, we can compute the (static) correlations between any pair of observations,  $\mathbf{X}_i$  and  $\mathbf{X}_j$  using:

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

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

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

61 We can generalize this formula to compute time-varying correlations by incorporating a *weight function*  
 62 that takes a time  $t$  as input, and returns how much the observed data every timepoint (including  $t$ ) contribute  
 63 to the correlations at time  $t$  (Fig. 1).

Given a weight function  $w(t)$  for timepoint  $t$ , evaluated at timepoints in the interval  $[1, \dots, T]$ , we can

extend the static correlation formula in Equation 2 to reflect an *instantaneous correlation* at timepoint  $t$ :

$$\text{timecorr}(\mathbf{X}_i, \mathbf{X}_j, t) = \frac{\sum_{t=1}^T (\mathbf{X}_i(t) - \tilde{\mathbf{X}}_i(t)) (\mathbf{X}_j(t) - \tilde{\mathbf{X}}_j(t))}{\sqrt{\sum_{t=1}^T \tilde{\sigma}_{\mathbf{X}_i}^2(t) \tilde{\sigma}_{\mathbf{X}_j}^2(t)}}, \text{ where} \quad (4)$$

$$\tilde{\mathbf{X}}_k(t) = \sum_{i=1}^T w(t, i) \mathbf{X}_k(i), \quad (5)$$

$$\tilde{\sigma}_{\mathbf{X}_k}^2(t) = \sum_{i=1}^T (\mathbf{X}_k(i) - \tilde{\mathbf{X}}_k(t))^2, \quad (6)$$

64 and  $w(t, i)$  is shorthand for  $w(t)$  evaluated at timepoint  $i$ . Equation 5 may be used to estimate the instanta-  
 65 neous correlations between every pair of observations, at each timepoint (i.e.,  $\mathbf{Y}$ ).

## 66 Inter-subject dynamic correlations

Equation 5 provides a means of taking a single observation matrix,  $\mathbf{X}$  and estimating the dynamic correlations from moment to moment,  $\mathbf{Y}$ . 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  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_P\}$  reflect the  $T$  by  $F$  observation matrices 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:

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

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  $\mathbf{Y}_i(t)$  denotes the correlation matrix (Eqn. 2) between each column of  $\mathbf{X}_i$  and each column of the average observations from all *other* participants,  $\bar{\mathbf{X}}_{\setminus i}$ :

$$\bar{\mathbf{X}}_{\setminus i} = R \left( \frac{1}{P-1} \sum_{i \in \setminus i} Z(\mathbf{X}_i) \right), \quad (10)$$

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

## Higher-order correlations

Given a timeseries of dynamic correlations (e.g., obtained using Eqn. 5), higher-order correlations reflect the dynamic correlations between columns of  $\mathbf{Y}$ . Given unlimited computing resources, one could use repeated applications of Equation 5 to estimate these higher-order correlations (i.e., substituting in the previous output,  $\mathbf{Y}$ , for the input,  $\mathbf{X}$  in the equation). However, because each output  $\mathbf{Y}$  has  $O(F^2)$  columns relative to  $F$  columns in the input  $\mathbf{X}$ , the output of Equation 5 grows with the square of the number of repeated applications (total cost of computing  $n^{\text{th}}$  order correlations is  $O(F^{2n})$  for  $n \in \mathcal{J}, n > 0$ ). When  $F$  or  $n$  is large, this approach quickly becomes intractable.

To make progress in computing  $\mathbf{Y}_{n+1}$ , we can approximate  $\mathbf{Y}_n$  by computing an  $O(F)$ -dimensional embedding of  $\mathbf{Y}_n$ , termed  $\hat{\mathbf{Y}}_n$ , and then we can apply Equation 5 to  $\hat{\mathbf{Y}}_n$  rather than directly to  $\mathbf{Y}_n$ . This enables us to maintain  $O(n)$  scaling with respect to  $n$ , rather than exponential scaling via the direct approach.

There are many possible methods for computing  $\hat{\mathbf{Y}}_n$  from  $\mathbf{Y}_n$ , including traditional dimensionality reduction approaches and graph theory based approaches as described next. In the *Discussion* section we elaborate on other potential approaches.

## Dimensionality reduction-based approaches to computing $\hat{\mathbf{Y}}_n$

Commonly used 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; Jutten & Herault, 1991; Comon et al., 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. B. Mairal et al., 2009; J. Mairal et al., 2009), deep autoencoders (Hinton & Salakhutdinov,

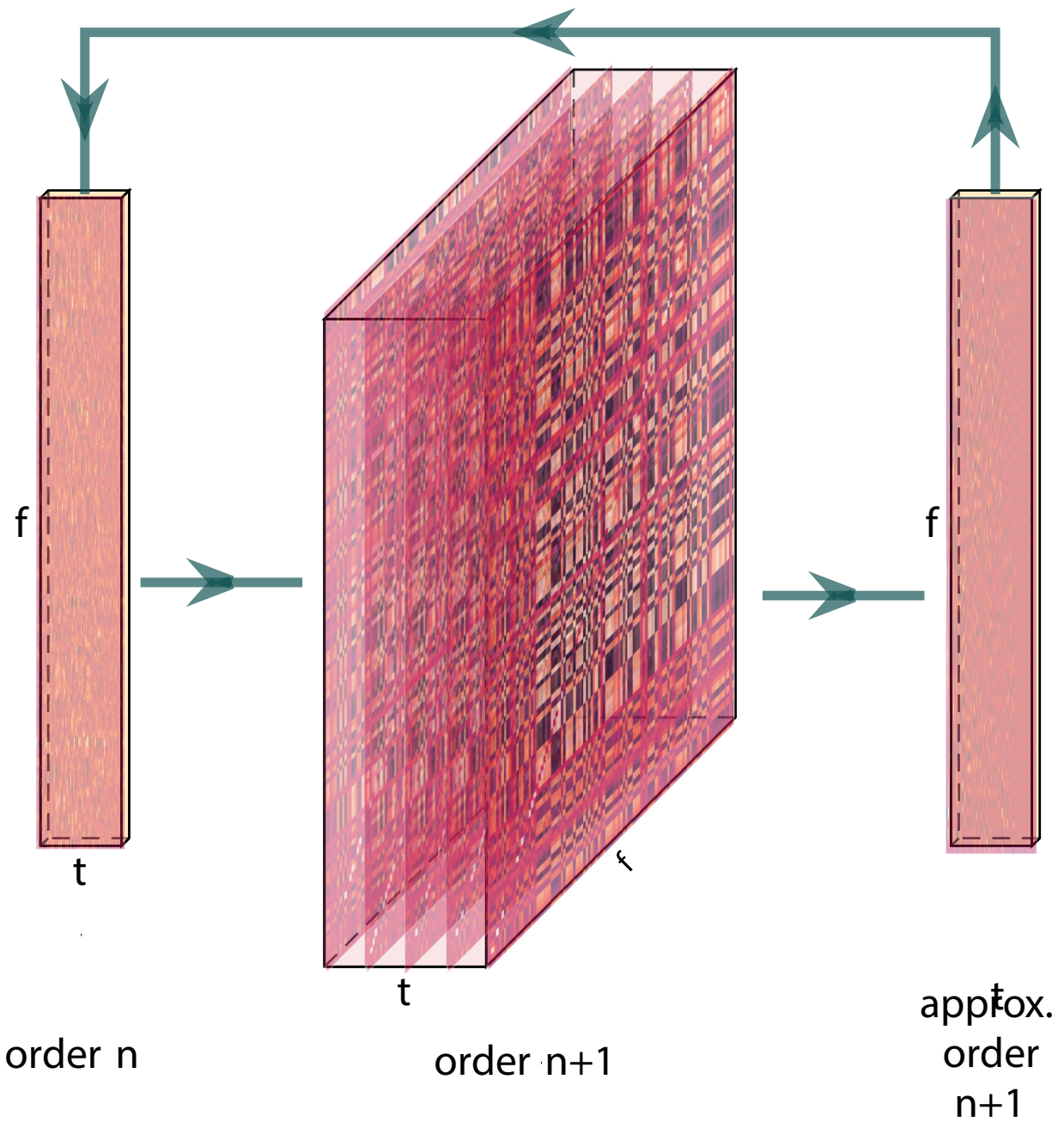


Figure 2: **Computing higher order correlations.** Correlations are computed then approximated to the same size as original data. This process is repeated to compute

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  $\hat{\mathbf{Y}}$  with  $i$  columns that is closest to the original  $\mathbf{Y}$  with  $j$  columns, and where (typically)  $i \ll j$ . The different approaches place different constraints on what properties  $\hat{\mathbf{Y}}$  must satisfy and which aspects of the data are compared (and how) to characterize the match between  $\hat{\mathbf{Y}}$  and  $\mathbf{Y}$ .

Applying dimensionality reduction algorithms to  $\mathbf{Y}$  yields a  $\hat{\mathbf{Y}}$  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  $\hat{\mathbf{Y}}_n$  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  $\hat{\mathbf{Y}}_n$  is a lossy operation. Second, whereas the columns of  $\mathbf{Y}_n$  may be mapped directly onto pairs of columns of  $\mathbf{Y}_{n-1}$ , that mapping either becomes less cleanly defined in  $\hat{\mathbf{Y}}_n$  due to the reweightings and/or nonlinear transformations.

#### **Graph theory-based approaches to computing $\hat{\mathbf{Y}}_n$**

Graph theoretic measures take as input a matrix of interactions (e.g., using the above notation, an  $F \times F$  correlation matrix or binarized correlation matrix reconstituted from a single timepoint’s row of  $\mathbf{Y}$ ) and return as output a set of  $F$  measures describing how each node (feature) sits within that interactions matrix with respect to the rest of the population. Common measures include betweenness centrality (the proportion of shortest paths between each pair of nodes in the population that involves the given node in question; e.g., Newman, 2005; Opsahl et al., 2010; Barthélemy, 2004; Geisberger et al., 2008; Freeman, 1977); diversity and dissimilarity (characterizations of how differently connected a given node is from others in the population; e.g., Rao, 1982; Lin, 2009; Ricotta & Szeidl, 2006); Eigenvector centrality and pagerank centrality (measures of how influential a given node is within the broader network; e.g., Newman, 2008; Bonacich, 2007; Lohmann et al., 2010; Halu et al., 2013); 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).

As an alternative to the above dimensionality reduction approach to embedding  $\mathbf{Y}_n$  in a lower-dimensional

space, but still allowing for scalable explorations of higher-order structure in the data, we also explore using the above graph theoretic measures as a means of obtaining  $\hat{\mathbf{Y}}_n$ . In particular: for a given graph theoretic measure,  $\eta : \mathcal{R}^{F \times F} \rightarrow \mathcal{R}^F$ , we can use  $\eta$  to transform each row of  $\mathbf{Y}_n$  in a way that characterizes the corresponding graph-theoretic properties of each column. Whereas the dimensionality reduction approach to computing  $\hat{\mathbf{Y}}_n$  is lossy, the graph-theory approach is lossless. However, whereas the dimensionality reduction approach maintains ties (direct or indirect) to the original activity patterns reflected in  $\mathbf{Y}_{n-1}$ , the graph-theory approach does not. Instead, the graph-theory characterizes the nature and timecourse of each feature's *participation* in the network.

## Evaluation metrics

We evaluate our approach to extracting dynamic correlations and higher-order correlations using several metrics detailed next. First, we generated synthetic data using known time-varying correlations, and then we evaluated the fidelity with which Equation 5 could recover those correlations (for synthetic datasets with different properties, and using different kernels to define the weights; Fig. 1). We then turned to a series of analyses on a (real) neuroimaging dataset where the ground truth correlations were *not* known. We evaluated whether the recovered correlations could be used to accurately label held-out neuroimaging data with the time at which it was collected, and also whether the history of correlations (and higher-order correlations) from time 0 to  $t - 1$  could be used to predict future activity patterns at time  $t$ . We used these latter evaluations (using timepoint decoding and predictions of held-out future data) as a proxy for gauging how much explanatory power the recovered correlations held with respect to the observed data.

## Generating synthetic data

Ramping dataset and block dataset. Create an example figure for both.

## Recovery of ground truth parameters from synthetic data

Apply timecorr with a given kernel, then correlate each recovered correlation matrix with the ground truth. Explore how recovery varies with the kernel, kernel parameters, and specific structure of the data (e.g. slow changes as in the ramping dataset, versus rapid changes as in the block dataset).

## Timepoint decoding

To explore how higher-order structure varies with stimulus structure and complexity, we used a previous neuroimaging dataset Simony et al. (2016) in which participants listened to an audio recording of a story;



36 participants listen to an intact version of the story, 17 participants listen to time-scrambled recordings of the same story where paragraphs were scrambled, 36 participants listen to word-scrambled version and 36 participants lay in rest condition.

Prior work has shown participants share similar neural responses to richly structured stimuli when compared to stimuli with less structure. To assess whether the moment-by-moment higher order correlations were reliably preserved across participants, we used inter-subject functional connectivity (ISFC) to isolate the time-varying correlational structure (functional connectivity patterns that were specifically driven by the story participants listened to. Following the analyses conducted by (HTFA) Manning et al. (2018), we first applied *hierarchical topographic factor analysis* (HTFA) to the fMRI datasets to obtain a time series of 700 node activities for every participant. We then computed the dynamic weighted ISFC using a gaussian kernel with a width of 5. We then approximated these dynamic correlation using PCA and computed the dynamic weighted ISFC on the approximations. We repeated this process up to 10th order approximated correlations.

To assess decoding accuracy, we randomly divided participants for each stimulus into training and testing groups. For the zeroth order, we computed the mean factor activity for each group. For all subsequent orders up to the tenth order, we computed the mean approximated dynamic ISFC of factor activity for each group. For each group of participants in turn, we compared these activity patterns (using Pearson correlations) to estimate the story times each pattern corresponded to.

Specifically, we asked, for each timepoint: what are the correlations between the first group's and second group's activity patterns at each order. We then took weighted mixture of the correlation matrices at each order using mixing parameter,  $\phi$ , where  $0 < \phi < 1$  reflects a weighted mixture of order based decoding. We optimized  $\phi$ , by subdividing the training group. We note that the decoding test we used is a conservative in which we count a timepoint label as incorrect if it is not an exact match.

## Results

### Synthetic data

Figure: overall timecourse of recovery, also recovery near event boundaries.

### Neuroimaging dataset (Simony et al., 2016)

Figures: decoding by level For our decoding analysis, we used HTFA-derived node activities Manning et al. (2018) from fMRI 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). We sought to demonstrate how higher-order correlations may be used to examine dynamic interactions of brain patterns in (real) multi-subject fMRI datasets. This story listening dataset was collected as part of a separate study, where the full imaging parameters, image preprocessing methods, and experimental details may be found (Simony et al., 2016). The dataset is available at <http://arks.princeton.edu/ark:/88435/dsp015d86p269k>.

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

Separately for each experimental condition, we divided participants into two groups. For the zeroth order, we computed the mean factor activity for each group. For all subsequent orders up to the tenth order, we computed the mean approximated dynamic ISFC of factor activity for each group (see *Materials and methods*). For each order, we correlated the group 1 activity patterns with group 2 activity patterns. We then subdivided the group 1 to obtain an optimal weighting parameter for each order's correlation matrix using the same cross validation method. We used the optimal weighting parameters to obtain a weighted-mixture of each order's correlation matrix. Using these correlations, we labeled the group 1 timepoints using the group 2 timepoints with which they were most highly correlated; we then computed the proportion of correctly labeled group 1 timepoints. (We also performed the symmetric analysis whereby we labeled the group 2 timepoints using the group 1 timepoints as a template.) We repeated this procedure 10 times (randomly re-assigning participants to the two groups each time) to obtain a distribution of decoding accuracies for each experimental condition. (There were 272 timepoints for paragraph condition, 300 timepoints for intact and word conditions, and 400 timepoints for rest condition, so chance performance on this decoding test is was  $\frac{1}{272}$ ,  $\frac{1}{300}$ , and  $\frac{1}{400}$  respectively.

## Discussion

- multiple timescale representations (a la Hasson group) implies first-order network interactions. higher-order interactions imply generalizations between interacting representations (e.g. mirrored

schema, a la Norman/Baldassano/Hasson). possibly cite NTB 2013 science review, using as evidence that this is where the field is going (voxels  $\rightarrow$  patterns (L0)  $\rightarrow$  interactions (L1)  $\rightarrow$  higher order patterns (L2+).

- related approaches: sliding window, phase-based correlations, within-ROI spatial correlations at each timepoint, granger causality, other explicit models (e.g. virtual brain).
- other applications: molecular interactions (protein folding?), diagnosis (e.g. psychiatric disorders as network flow problems– gratton work?), social network dynamics (e.g. financial markets, social media interactions)

## Concluding remarks

the universe is complicated and we need scalable approaches to studying how the pieces are interacting to make sense of it. one small step for mankind, and so on.

## 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 visualization of large scale networks. *arXiv*, cs/0504107v2.
- Barthélemy, M. (2004). Betweenness centrality in large complex networks. *European Physical Journal B*, 38, 163–168.

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

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

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

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

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

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

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

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

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

252 Honey, C. J., Kötter, R., Breakspear, M., & Sporns, O. (2007). Network structure of cerebral cortex shapes  
253 functional connectivity on multiple time scales. *Proceedings of the National Academy of Science USA*, 104(24),  
254 10240–10245.

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

257 Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*,  
258 401, 788–791.

259 Lin, J. (2009). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*,  
260 37(1), 145–151.

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

- 264 Mairal, J., Ponce, J., Sapiro, G., Zisserman, A., & Bach, F. R. (2009). Supervised dictionary learning. *Advances*  
265 *in Neural Information Processing Systems*, 1033–1040.
- 266 Mairal, J. B., Bach, F., Ponce, J., & Sapiro, G. (2009). Online dictionary learning for sparse coding. *Proceedings*  
267 *of the 26th annual international conference on machine learning*, 689–696.
- 268 Manning, J. R., Ranganath, R., Norman, K. A., & Blei, D. M. (2014). Topographic factor analysis: a Bayesian  
269 model for inferring brain networks from neural data. *PLoS One*, 9(5), e94914.
- 270 Manning, J. R., Zhu, X., Willke, T. L., Ranganath, R., Stachenfeld, K., Hasson, U., . . . Norman, K. A. (2018).  
271 A probabilistic approach to discovering dynamic full-brain functional connectivity patterns. *NeuroImage*,  
272 180, 243–252.
- 273 McInnes, L., & Healy, J. (2018). UMAP: Uniform manifold approximation and projection for dimension  
274 reduction. *arXiv*, 1802(03426).
- 275 Newman, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, 27,  
276 39–54.
- 277 Newman, M. E. J. (2008). The mathematics of networks. *The New Palgrave Encyclopedia of Economics*, 2, 1–12.
- 278 Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: generalizing degree  
279 and shortest paths. *Social Networks*, 32, 245–251.
- 280 Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh,*  
281 *and Dublin Philosophical Magazine and Journal of Science*, 2, 559–572.
- 282 Rao, C. R. (1982). Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology*,  
283 21(1), 24–43.
- 284 Ricotta, C., & Szeidl, L. (2006). Towards a unifying approach to diversity measures: Bridging the gap  
285 between the Shannon entropy and Rao’s quadratic index. *Theoretical Population Biology*, 70(3), 237–243.
- 286 Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpreta-  
287 tions. *NeuroImage*, 52, 1059–1069.
- 288 Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85(2), 461–464.
- 289 Simony, E., Honey, C. J., Chen, J., & Hasson, U. (2016). Uncovering stimulus-locked network dynamics  
290 during narrative comprehension. *Nature Communications*, 7(12141), 1–13.

- 291 Spearman, C. (1904). General intelligence, objectively determined and measured. *Americal Journal of*  
292 *Psychology*, 15, 201–292.
- 293 Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of Royal Statistical*  
294 *Society, Series B*, 61(3), 611–622.
- 295 van der Maaten, L. J. P., & Hinton, G. E. (2008). Visualizing high-dimensional data using t-SNE. *Journal of*  
296 *Machine Learning Research*, 9, 2579–2605.
- 297 Zar, J. H. (2010). *Biostatistical analysis*. Prentice-Hall/Pearson.