

1 **High-order cognition is supported by complex but**
2 **compressible brain activity patterns**

3 Lucy L. W. Owen^{1,2} and Jeremy R. Manning^{1,*}

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

6 ²Carney Institute for Brain Sciences,
7 Brown University, Providence, RI

8 *Address correspondence to jeremy.r.manning@dartmouth.edu

9 February 27, 2023

10 **Abstract**

11 We applied dimensionality reduction algorithms and pattern classifiers to functional neuroimaging
12 data collected as participants listened to a story, temporally scrambled versions of the story, or underwent
13 a resting state scanning session. These experimental conditions were intended to require different depths
14 of processing and inspire different levels of cognitive engagement. We considered two primary aspects of
15 the data. First, we treated the number of features (components) required to achieve a threshold decoding
16 accuracy as a proxy for the “compressibility” of the neural patterns (where fewer components indicate
17 greater compression). Second, we treated the maximum achievable decoding accuracy across participants
18 as an indicator of the “stability” of the recorded patterns. Overall, we found that neural patterns recorded
19 as participants listened to the intact story required fewer features to achieve comparable classification
20 accuracy to the other experimental conditions. However, the peak decoding accuracy (achievable with
21 more features) was also highest during intact story listening. Taken together, our work suggests that
22 our brain networks flexibly reconfigure according to ongoing task demands, and that the activity patterns
23 associated with higher-order cognition and high engagement are both more complex and more compressible
24 than the activity patterns associated with lower-order tasks and lower levels of engagement.

25 **Introduction**

26 Large-scale networks, including the human brain, may be conceptualized as occupying one or more positions
27 along on a continuum. At one extreme, every node is fully independent of every other node. At the other
28 extreme, all nodes behave identically. Each extreme optimizes key properties of how the network functions.
29 When every node is independent, the network is maximally *expressive*: if we define the network’s “state”
30 as the activity pattern across its nodes, then every state is equally reachable by a network with fully
31 independent nodes. On the other hand, a network of identically behaved nodes optimizes *robustness*: any
32 subset of nodes may be removed from the network without any loss of function or expressive power, as
33 long as any single node remains. Presumably, most natural systems tend to occupy positions between
34 these extremes. We wondered: might the human brain reconfigure itself to be more flexible or more robust

30 according to ongoing demands? In other words, might the brain reconfigure its connections or behaviors
31 under different circumstances to change its position along this continuum?

32 Closely related to the above notions of expressiveness versus robustness are measures of how much
33 *information* is contained in a given signal or pattern, and how *redundant* a signal is (Shannon, 1948). Formally,
34 information is defined as the amount of uncertainty about a given variables' outcomes (i.e., entropy),
35 measured in *bits*, or the optimal number of yes/no questions needed to reduce uncertainty about the
36 variable's outcomes to zero. Highly complex systems with many degrees of freedom (i.e., high flexibility
37 and expressiveness), are more information-rich than simpler or more constrained systems. The redundancy
38 of a signal denotes the difference how expressive the signal *could* be (i.e., proportional to the number of
39 unique states or symbols used to transmit the signal) and the actual information rate (i.e., the entropy of
40 each individual state or symbol). If a brain network's nodes are fully independent, then the number of bits
41 required to express a single activity pattern is proportional to the number of nodes. The network would
42 also be minimally redundant, since the status of every node would be needed to fully express a single brain
43 activity pattern. If a brain network's nodes are fully coupled and identical, then the number of bits required
44 to express a single activity pattern is proportional to the number of unique states or values any individual
45 node can take on. Such a network would be highly redundant, since knowing any individual node's state
46 would be sufficient to recover the full-brain activity pattern. Highly redundant systems are also robust,
47 since there is little information loss from losing any given observation.

48 We take as a given that brain activity is highly flexible: our brains can exhibit nearly infinite activity
49 patterns. This flexibility implies that our brains activity patterns are highly information rich. However,
50 brain activity patterns are also highly structured. For example, full-brain correlation matrices are stable
51 within (Finn et al., 2015, 2017; Gratton et al., 2018) and across (Yeo et al., 2011; Glerean et al., 2012; Gratton
52 et al., 2018; Cole et al., 2014) individuals. This stability suggests that our brains' activity patterns are at
53 least partially constrained, for example by anatomical, external, or internal factors. Constraints on brain
54 activity that limit its flexibility decrease expressiveness (i.e., its information rate). However, constraints on
55 brain activity also increase its robustness to noise (e.g., “missing” or corrupted signals may be partially
56 recovered). For example, recent work has shown that full-brain activity patterns may be reliably recovered
57 from only a relatively small number of implanted electrodes (Owen et al., 2020; Scangos et al., 2021). This
58 robustness property suggests that the relevant signal (e.g., underlying factors that have some influence over
59 brain activity patterns) are compressible.

60 To the extent that brain activity patterns contain rich task-relevant information, we should be able to
61 use the activity patterns to accurately differentiate between different aspects of the task (e.g., using pattern
62 classifiers; Norman et al., 2006). For example, prior work has shown a direct correspondence between

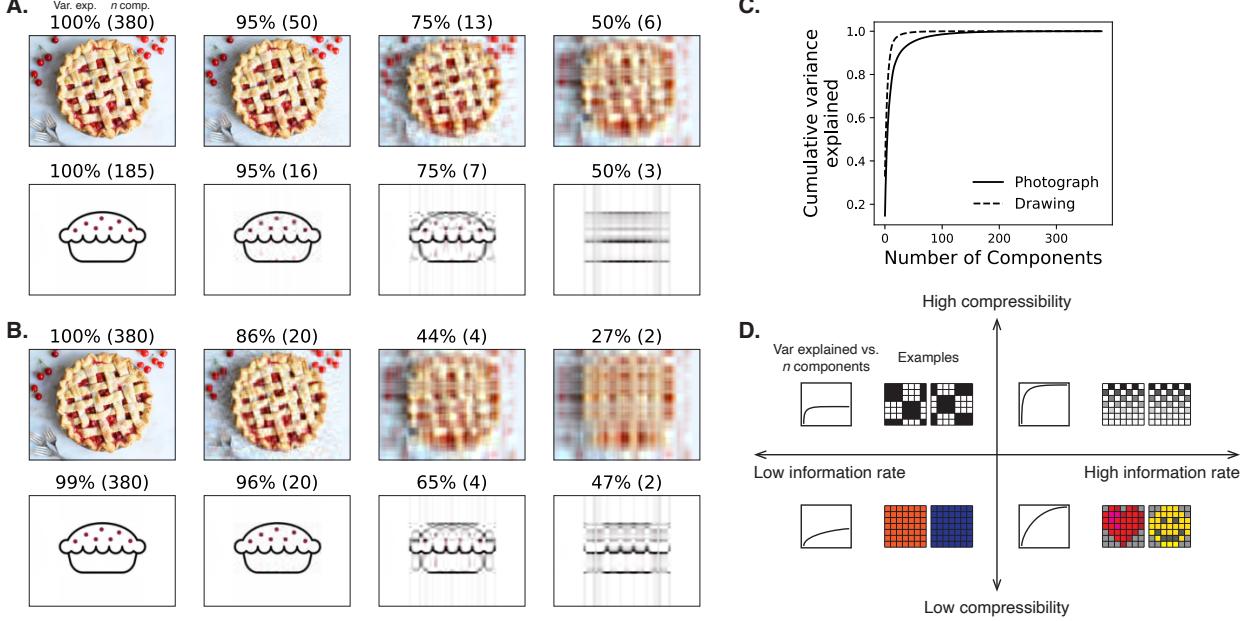


Figure 1: Information content and compressibility. **A. Variance explained for two images.** We applied principal components analysis to a photograph and drawing, treating each row of the images as “observations.” Across columns, we identified the number of components required to explain 100%, 95%, 75%, or 50% of the cumulative variance in each image (the 100% columns denote the original images). The numbers of components are indicated in parentheses, and the resulting “compressed” images are displayed. **B. Representing two images with different numbers of components.** Using the same principal component decompositions as in Panel A, we computed the cumulative proportion of variance explained with 380 (original images), 20, 4, or 2 components. **C. Cumulative variance explained versus number of components.** For the images displayed in Panels A and B, we plot the cumulative proportion of variance explained as a function of the number of components used to represent each image. **D. Information rate and compressibility.** Across multiple images, the information rate (i.e., the amount of information contained in each image; horizontal axis) is high when each individual pixel provides information that cannot be inferred from other pixels. High-information rate images tend to be high-resolution, and low-information rate images tend to be low-resolution. Compressibility is related to the difference between the information required to specify the original versus compressed images (vertical axis). Highly compressible images often contain predictable structure (redundancies) that can be leveraged to represent the images much more efficiently than in their original feature spaces.

63 classification accuracy and the information content of a signal (Alvarez, 2002). To the extent that brain
64 activity patterns are compressible, we should be able to generate simplified (e.g., lower dimensional)
65 representations of the data while still preserving the relevant or important aspects of the original signal.
66 In general, information content and compressibility are related but are partially dissociable (Fig. 1). If a
67 given signal (e.g., a representation of brain activity patterns) contains more information about ongoing
68 cognitive processes, then the peak decoding accuracy should be high. And if the signal is compressible, a
69 low-dimensional embedding of the signal will be similarly informative to the original signal (Fig. 1D).

70 Several recent studies suggest that the complexity of brain activity is task-dependent, whereby simpler
71 tasks with lower cognitive demands are reflected by simpler and more compressible brain activity patterns,
72 and more complex tasks with higher cognitive demands are reflected by more complex and less compressible
73 brain activity patterns (Mack et al., 2020; Owen et al., 2021). These findings complement other work
74 suggesting that functional connectivity (correlation) patterns are task-dependent (Finn et al., 2017; Owen
75 et al., 2020; Cole et al., 2014), although see Gratton et al. (2018). Higher-order cognitive processing of a
76 common stimulus also appears to drive more stereotyped task-related activity and functional connectivity
77 across individuals (Hasson et al., 2008; Lerner et al., 2011; Simony & Chang, 2020; Simony et al., 2016).

78 The above prior studies are consistent with two potential descriptions of how cognitive processes are
79 reflected in brain activity patterns. One possibility is that the information rate of brain activity increases during
80 more complex or higher-level cognitive processing. If so, then the ability to reliably decode cognitive states
81 from brain activity patterns should improve with task complexity or with the level (or “depth”) of cognitive
82 processing. A second possibility is that the compressibility of brain activity patterns increases during
83 more complex or higher-level cognitive processing. If so, then individual features of brain recordings, or
84 compressed representations of brain recordings, should carry more information during complex or high-
85 level (versus simple or low-level) cognitive tasks.

86 We used a previously collected neuroimaging dataset to estimate the extent to which each of these two
87 possibilities might hold. The dataset we examined comprised functional magnetic resonance imaging (fMRI)
88 data collected as participants listened to an audio recording of a 10-minute story, temporally scrambled
89 recordings of the story, or underwent a resting state scan (Simony et al., 2016). Each of these experimental
90 conditions evokes different depths of cognitive processing (Simony et al., 2016; Lerner et al., 2011; Hasson et
91 al., 2008; Owen et al., 2021). We used across-participant classifiers to decode listening times in each condition,
92 as a proxy for how “informative” the task-specific activity patterns were (Simony & Chang, 2020). We also
93 use principle components analysis to generate lower-dimensional representations of the activity patterns.
94 We then repeated the classification analyses after preserving different numbers of components and examined
95 how classification accuracy changed across the different experimental conditions.

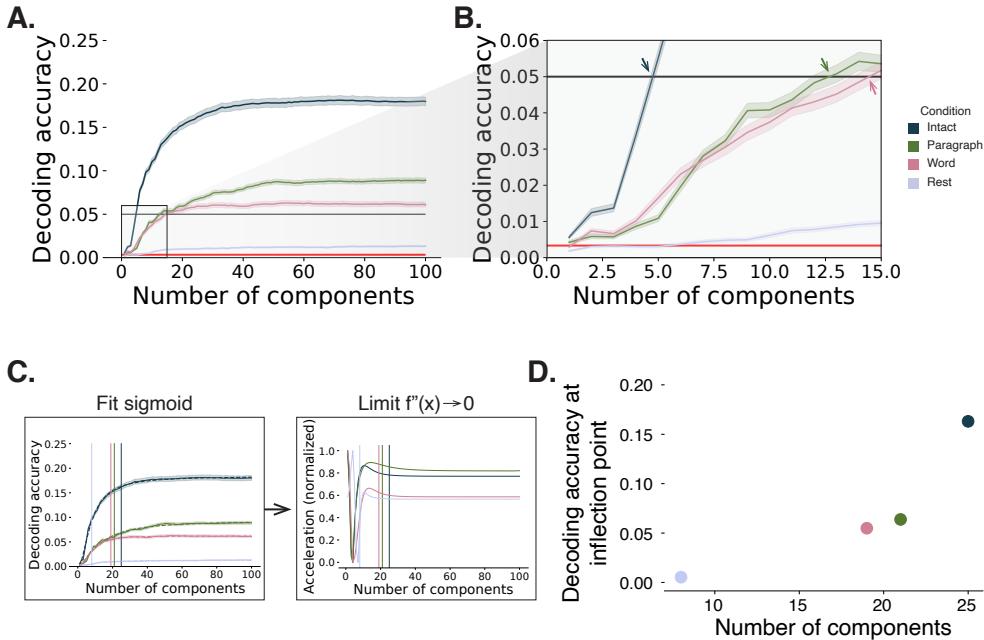


Figure 2: Decoding accuracy and compression. **A. Decoding accuracy by number of components.** Ribbons of each color display cross-validated decoding performance for each condition (intact, paragraph, word, and rest), as a function of the number of components (features) used to represent the data. The horizontal red line denotes chance performance, and the horizontal black line denotes 5% decoding accuracy (used as a reference in Panel B). **B. Numbers of components required to reach a fixed decoding accuracy threshold, by condition.** The panel displays a zoomed-in view of the inset in Panel A. Intersections between each condition's decoding accuracy curve and the 5% decoding accuracy reference line are marked by arrows. **C. Estimating inflection points.** We sought to identify an “inflection point” for each decoding curve, denoting the number of components at which the decoding curve asymptotes. We fit sigmoid functions to each decoding curve (left sub-panel) and then computed the minimum number of components where the second derivative of the sigmoid was both positive and less than a threshold value of 0.0001. **D. Inflection points by condition.** Each dot displays the number of components (x-axis) and decoding accuracy (y-axis) at one condition's inflection point. All error ribbons denote bootstrap-estimated 95% confidence intervals.

96 Results

97 Discussion

98 Methods

99 We measured properties of recorded neuroimaging data under different task conditions that varied system-
100 atically in cognitive engagement and depth of processing. We were especially interested in how *informative*
101 and *compressible* the activity patterns were under these different conditions (Fig. 1).

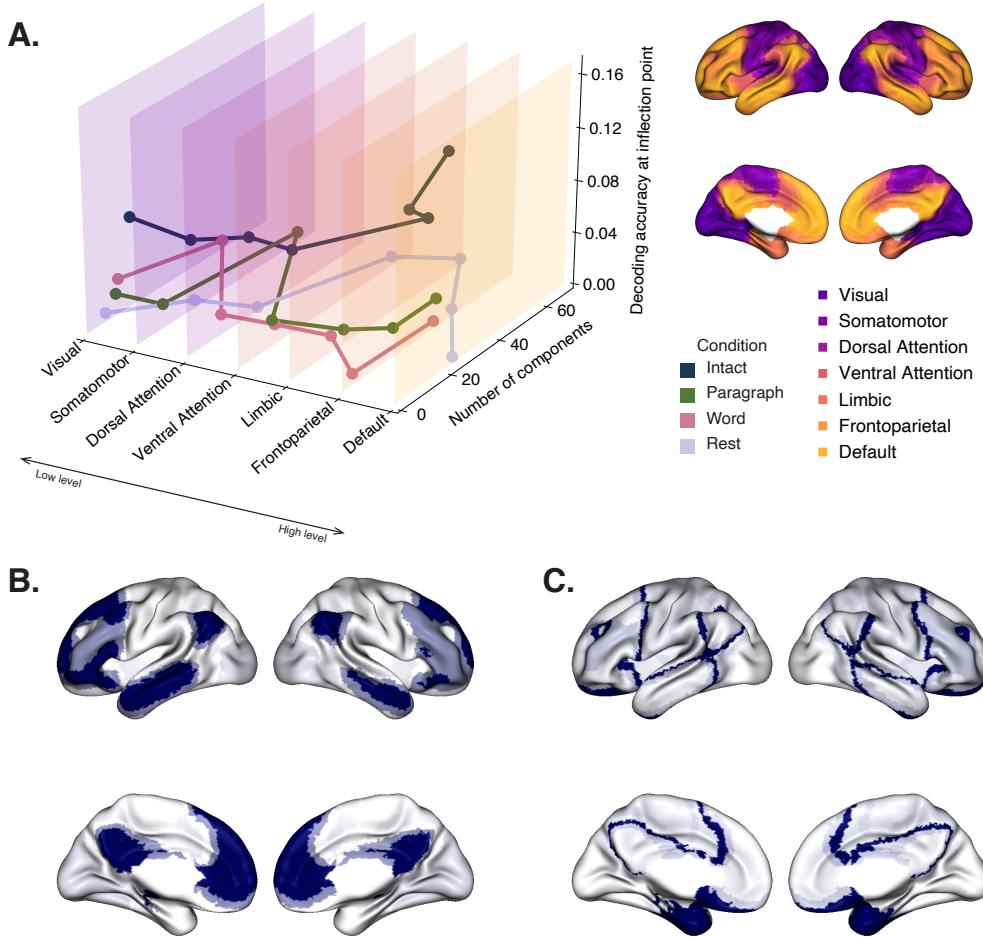


Figure 3: Network-specific decoding accuracy and compression. **A. Decoding accuracy and number of components for network-specific inflection points.** We considered the seven networks identified by Yeo et al. (2011). We computed each network's inflection point, for each experimental condition, using the procedure described in Figure 2C. **B. Network-specific decoding accuracy.** Each of the seven networks are colored according to the decoding accuracy at the network's inflection point for the "intact" experimental condition (corresponding to the dark blue curve in Panel A). **C. Network-specific compression.** Each of the seven networks are colored according to the number of components at the network's inflection point for the intact experimental condition. Larger numbers of components reflect lower compressibility.

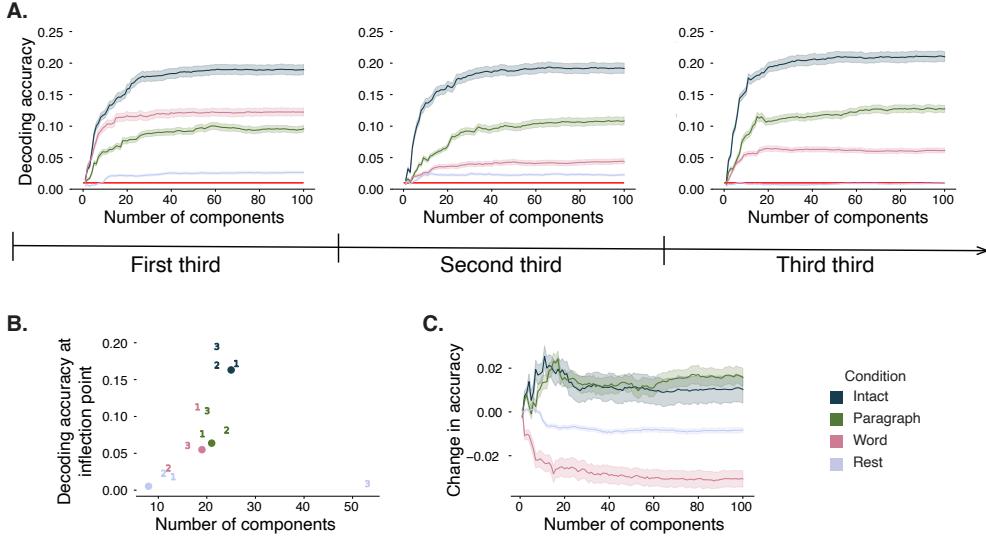


Figure 4: Changes in decoding accuracy and compression over time. A. Decoding accuracy by number of components, by story segment. Each family of curves is plotted in the same format as Figure 2A but reflects data only from one third of the dataset. **B. Inflection points by condition and segment.** The dots re-plot the inflection points from Figure 2D for reference. The numbers denote the inflection points for each third of the dataset (1: first third; 2: second third; 3: third third; colors denote experimental conditions). **C. Change in decoding accuracy over time, by number of components.** For each number of components (x -axis) and condition (color), we fit a regression line to the decoding accuracies obtained for the first, second, and third thirds of the dataset (corresponding to the left, middle, and right columns of Panel A, respectively). The y -axis denotes the slopes of the regression lines. All error ribbons denote bootstrap-estimated 95% confidence intervals.

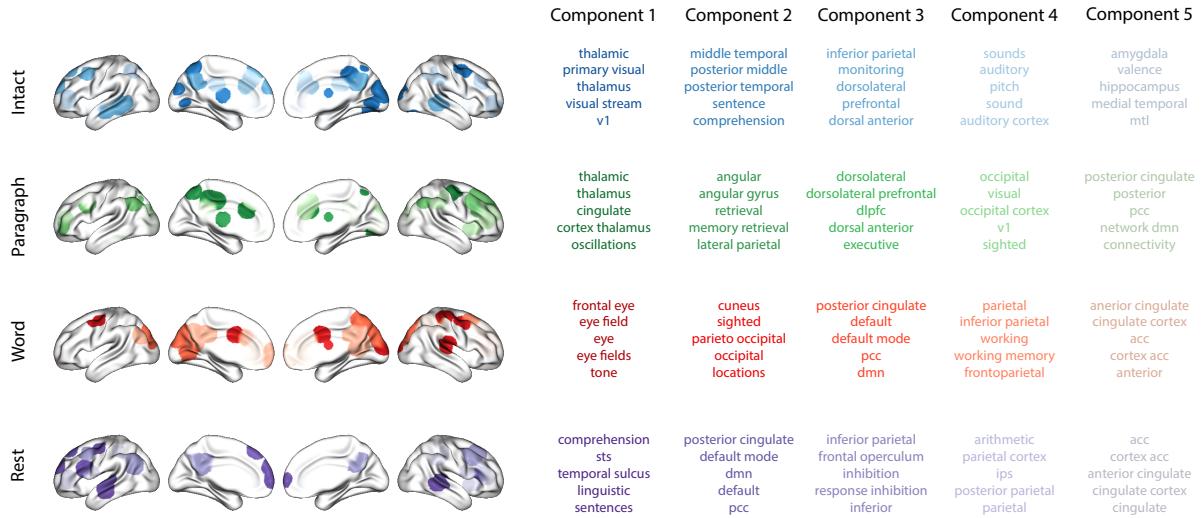


Figure 5: Top terms associated with the highest-weighted components by condition. Each row corresponds to an experimental condition, and the colors correspond to the component number (ranked by proportion of variance explained). The inflated brain plots display the top 20 highest-weighted hubs (see *Topographic Factor Analysis*) for each components'. The lists on the right display the top five Neurosynth terms (Rubin et al., 2017) decoded from each components' brain map. Analogous maps computed separately for each story segment may be found in Figure S1.

102 **Functional neuroimaging data collected during story listening**

103 We examined an fMRI dataset collected by Simony et al. (2016) that the authors have made publicly available
104 at arks.princeton.edu/ark:/88435/dsp015d86p269k. The dataset comprises neuroimaging data collected as
105 participants listened to an audio recording of a story (intact condition; 36 participants), listened to temporally
106 scrambled recordings of the same story (17 participants in the paragraph-scrambled condition listened to
107 the paragraphs in a randomized order and 36 in the word-scrambled condition listened to the words in a
108 randomized order), or lay resting with their eyes open in the scanner (rest condition; 36 participants). Full
109 neuroimaging details may be found in the original paper for which the data were collected Simony et al.
110 (2016). Procedures were approved by the Princeton University Committee on Activities Involving Human
111 Subjects, and by the Western Institutional Review Board (Puyallup, WA). All subjects were native English
112 speakers with normal hearing and provided written informed consent.

113 **Hierarchical topographic factor analysis (HTFA)**

114 Following our prior related work, we used HTFA Manning et al. (2018) to derive a compact representation
115 of the neuroimaging data. In brief, this approach approximates the timeseries of voxel activations (44,415
116 voxels) using a much smaller number of radial basis function (RBF) nodes (in this case, 700 nodes, as
117 determined by an optimization procedure; Manning et al., 2018)). This provides a convenient representation
118 for examining full-brain activity patterns and network dynamics. All of the analyses we carried out on the
119 neuroimaging dataset were performed in this lower-dimensional space. In other words, each participant's
120 data matrix, X , was a number-of-timepoints (T) by 700 matrix of HTFA-derived factor weights (where the
121 row and column labels were matched across participants). Code for carrying out HTFA on fMRI data may
122 be found as part of the BrainIAK toolbox Capota et al. (2017); Kumar et al. (2021), which may be downloaded
123 at brainiak.org.

124 **Principal components analysis (PCA)**

125 We applied group PCA (Smith et al., 2014) separately to the HTFA-derived representations of the data
126 (i.e., factor loadings) from each experimental condition. Specifically, for each condition, we considered
127 the set of all participants' T by 700 factor weight matrices. We used group PCA to project these 700-
128 dimensional matrices into a series of k -dimensional spaces, for $k \in \{3, 4, 5, \dots, 700\}$. This yielded a set of
129 number-of-participants matrices, each with T rows and k columns.

130 **Temporal decoding**

131 We sought to identify neural patterns that reflected participants' ongoing cognitive processing of incoming
132 stimulus information. As reviewed by Simony et al. (2016), one way of homing in on these stimulus-driven
133 neural patterns is to compare activity patterns across individuals. In particular, neural patterns will be
134 similar across individuals to the extent that the neural patterns under consideration are stimulus-driven, and
135 to the extent that the corresponding cognitive representations are reflected in similar spatial patterns across
136 people Simony & Chang (2020). Following this logic, we used an across-participant temporal decoding test
137 developed by Manning et al. (2018) to assess the degree to which different neural patterns reflected ongoing
138 stimulus-driven cognitive processing across people. The approach entails using a subset of the data to train
139 a classifier to decode stimulus timepoints (i.e., moments in the story participants listened to) from neural
140 patterns. We use decoding (forward inference) accuracy on held-out data, from held-out participants, as a
141 proxy for the extent to which the inputted neural patterns reflected stimulus-driven cognitive processing in
142 a similar way across individuals.

143 **Forward inference and decoding accuracy**

144 We used an across-participant correlation-based classifier to decode which stimulus timepoint matched
145 each timepoint's neural pattern. For a given value of k (i.e., number of principal components), we first
146 used group PCA to project the data from each condition into a k -dimensional space. Next, we divided the
147 participants into two groups: a template group, $\mathcal{G}_{\text{template}}$ (i.e., training data), and a to-be-decoded group,
148 $\mathcal{G}_{\text{decode}}$ (i.e., test data). We averaged the projected data within each group to obtain a single T by k matrix for
149 each group. Next, we correlated the rows of the two averaged matrices to form a T by T decoding matrix,
150 Λ . In this way, the rows of Λ reflected timepoints from the template group, while the columns reflected
151 timepoints from the to-be-decoded group. We used Λ to assign temporal labels to each timepoint (row)
152 from the test group's matrix, using the row of the training group's matrix with which it was most highly
153 correlated. We repeated this decoding procedure, but using $\mathcal{G}_{\text{decode}}$ as the template group and $\mathcal{G}_{\text{template}}$ as the
154 to-be-decoded group. Given the true timepoint labels (for each group), we defined the decoding accuracy
155 as the average proportion of correctly decoded timepoints, across both groups (where chance performance
156 is $\frac{1}{T}$). In Figures 2 and 4 we report the decoding accuracy for each condition and value of k , averaged across
157 $n = 1,000$ cross validation folds.

¹⁵⁸ **Reverse inference**

¹⁵⁹ To help interpret the brain activity patterns we found within the contexts of other studies, we created
¹⁶⁰ summary maps of each principal component, for each experimental condition, by summing together the
¹⁶¹ 20 HTFA-derived RBF nodes (see *Hierarchical Topographic Factor Analysis*) with the highest absolute value
¹⁶² weights for each of the top 5 components (Figs. 5, S1). We then carried out a meta analysis using Neu-
¹⁶³ rosynth Rubin et al. (2017) to identify the 5 terms most commonly associated with the given map.

¹⁶⁴ **Data and code availability**

¹⁶⁵ All of the code used to produce the figures and results in this manuscript, along with links to the corre-
¹⁶⁶ sponding datasets, may be found at github.com/ContextLab/pca-paper.

¹⁶⁷ **Acknowledgements**

¹⁶⁸ We acknowledge discussions with Rick Betzel, Emily Finn, and Jim Haxby. Our work was supported in part
¹⁶⁹ by NSF CAREER Award Number 2145172 to J.R.M. The content is solely the responsibility of the authors
¹⁷⁰ and does not necessarily represent the official views of our supporting organizations. The funders had no
¹⁷¹ role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

¹⁷² **Author contributions**

¹⁷³ Conceptualization: J.R.M. and L.L.W.O. Methodology: J.R.M. and L.L.W.O. Implementation: L.L.W.O. Anal-
¹⁷⁴ ysis: L.L.W.O. Writing, Reviewing, and Editing: J.R.M. and L.L.W.O. Supervision: J.R.M.

¹⁷⁵ **References**

- ¹⁷⁶ Alvarez, S. A. (2002). *An exact analytical relation among recall, precision, and classification accuracy in information*
¹⁷⁷ *retrieval* (Technical Report No. BCCS-02-01). Boston College.
- ¹⁷⁸ Capota, M., Turek, J., Chen, P.-H., Zhu, X., Manning, J. R., Sundaram, N., . . . Shin, Y. S. (2017). *Brain imaging*
¹⁷⁹ *analysis kit*.
- ¹⁸⁰ Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. (2014). Intrinsic and task-evoked
¹⁸¹ network architectures of the human brain. *Neuron*, 83(1), 238–251.

- 182 Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., & Constable, R. T. (2017). Can brain state
183 be manipulated to emphasize individual differences in functional connectivity. *NeuroImage*, 160, 140–151.
- 184 Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., . . . Constable, R. T. (2015).
185 Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature
186 Neuroscience*, 18, 1664–1671.
- 187 Gleean, E., Salmi, J., Lahnakoski, J. M., Jääskeläinen, I. P., & Sams, M. (2012). Functional magnetic resonance
188 imaging phase synchronization as a measure of dynamic functional connectivity. *Brain Connectivity*, 2(2),
189 91–101.
- 190 Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., . . . Petersen, S. E.
191 (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or
192 daily variation. *Neuron*, 98(2), 439–452.
- 193 Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive
194 windows in human cortex. *The Journal of Neuroscience*, 28(10), 2539–2550.
- 195 Kumar, M., Anderson, M. J., Antony, J. W., Baldassano, C., Brooks, P. P., Cai, M. B., . . . Norman, K. A. (2021).
196 BrainIAK: the brain image analysis kit. *Aperture*, In press.
- 197 Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal
198 receptive windows using a narrated story. *The Journal of Neuroscience*, 31(8), 2906–2915.
- 199 Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during
200 concept learning. *Nature Communications*, 11(46), doi.org/10.1038/s41467-019-13930-8.
- 201 Manning, J. R., Zhu, X., Willke, T. L., Ranganath, R., Stachenfeld, K., Hasson, U., . . . Norman, K. A. (2018).
202 A probabilistic approach to discovering dynamic full-brain functional connectivity patterns. *NeuroImage*,
203 180, 243–252.
- 204 Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern
205 analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9), 424–430.
- 206 Owen, L. L. W., Chang, T. H., & Manning, J. R. (2021). High-level cognition during story listening is
207 reflected in high-order dynamic correlations in neural activity patterns. *Nature Communications*, 12(5728),
208 doi.org/10.1038/s41467-021-25876-x.
- 209 Owen, L. L. W., Muntianu, T. A., Heusser, A. C., Daly, P., Scangos, K. W., & Manning, J. R. (2020). A Gaussian
210 process model of human electrocorticographic data. *Cerebral Cortex*, 30(10), 5333–5345.

- 211 Rubin, T. N., Kyoejo, O., Gorgolewski, K. J., Jones, M. N., Poldrack, R. A., & Yarkoni, T. (2017). Decoding
212 brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. *PLoS*
213 *Computational Biology*, 13(10), e1005649.
- 214 Scangos, K. W., Khambhati, A. N., Daly, P. M., Owen, L. L. W., Manning, J. R., Ambrose, J. B., ... Chang,
215 E. F. (2021). Biomarkers of depression symptoms defined by direct intracranial neurophysiology. *Frontiers*
216 *in Human Neuroscience*, In press.
- 217 Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3),
218 379–423.
- 219 Simony, E., & Chang, C. (2020). Analysis of stimulus-induced brain dynamics during naturalistic paradigms.
220 *NeuroImage*, 216, 116461.
- 221 Simony, E., Honey, C. J., Chen, J., & Hasson, U. (2016). Dynamic reconfiguration of the default mode
222 network during narrative comprehension. *Nature Communications*, 7(12141), 1–13.
- 223 Smith, S. M., Hyvärinen, A., Varoquaux, G., Miller, K. L., & Beckmann, C. F. (2014). Group-PCA for very
224 large fMRI datasets. *NeuroImage*, 101, 738–749.
- 225 Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... Buckner, R. L.
226 (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity.
227 *Journal of Neurophysiology*, 106(3), 1125–1165.