

High-level cognition is supported by information-rich but compressible brain activity patterns

Lucy L. W. Owen<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>2</sup>Carney Institute for Brain Sciences,  
Brown University, Providence, RI

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

December 23, 2023

## Abstract

Brain activity patterns are highly flexible and often complex, but also highly structured. Here we examined how fundamental properties of brain activity patterns relate to ongoing cognitive processes. To this end, we applied dimensionality reduction algorithms and pattern classifiers to functional neuroimaging data collected as participants listened to a story, temporally scrambled versions of the story, or underwent a resting state scanning session. These experimental conditions were intended to require different depths of processing and inspire different levels of cognitive engagement. We considered two primary aspects of the data. First, we treated the maximum achievable decoding accuracy across participants as an indicator of the “informativeness” of the recorded patterns. Second, we treated the number of features (components) required to achieve a threshold decoding accuracy as a proxy for the “compressibility” of the neural patterns (where fewer components indicate greater compression). Overall, we found that the peak decoding accuracy (achievable without restricting the numbers of features) was highest in the intact (unscrambled) story listening condition. However, the number of features required to achieve comparable classification accuracy was also lowest in the intact story listening condition. Taken together, our work suggests that our brain networks flexibly reconfigure according to ongoing task demands, and that the activity patterns associated with higher-order cognition and high engagement are both more informative and more compressible than the activity patterns associated with lower-order tasks and lower levels of engagement.

**Keywords:** information, compression, temporal decoding, dimensionality reduction, neuroimaging

**Significance Statement:** How our brains respond to ongoing experiences depends on what we are doing and thinking about, how engaged we are, and our past experiences, among other factors. To gain insights into how brain activity patterns reflect ongoing cognitive processes, we examined two fundamental aspects of brain activity under different cognitive circumstances: informativeness and compressibility. Informativeness refers to the extent to which brain patterns are both temporally specific and consistent across different people. Compressibility refers to the extent to how robust the informativeness of brain patterns is to dimensionality reduction. We found that brain activity patterns evoked by higher-level cognitive tasks are both more informative *and* more compressible than activity evoked by lower-level tasks. Our findings suggest that our brains flexibly reconfigure themselves to optimize different aspects of how they function according to ongoing cognitive demands.

## 37 Introduction

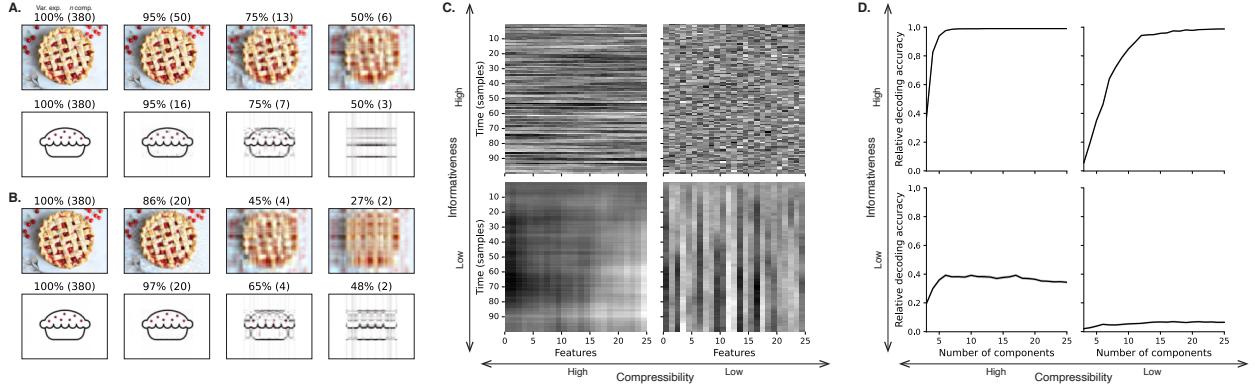
38 Large-scale networks, including the human brain, may be conceptualized as occupying one or  
39 more positions along on a continuum. At one extreme, every node is fully independent from  
40 every other node. At the other extreme, all nodes behave identically. Each extreme optimizes  
41 key properties of how the network functions. When every node is independent, the network is  
42 maximally *expressive*: if we define the network's "state" as the activity pattern across its nodes, then  
43 every state is equally reachable by a network with fully independent nodes. On the other hand, a  
44 network of identically behaved nodes optimizes *robustness*: any subset of nodes may be removed  
45 from the network without any loss of function or expressive power, as long as any single node  
46 remains. Presumably, most natural systems tend to occupy positions between these extremes. We  
47 wondered: might the human brain reconfigure itself to be more flexible or more robust according  
48 to ongoing demands? In other words, might the brain reconfigure its connections or behaviors  
49 under different circumstances to change its position along this continuum?

50 Closely related to the above notions of expressiveness versus robustness are measures of  
51 how much *information* is contained in a given signal or pattern, and how *redundant* a signal  
52 is (Shannon, 1948). Formally, information is defined as the amount of uncertainty about a given  
53 variables' outcomes (i.e., entropy), measured in *bits*, or the optimal number of yes/no questions  
54 needed to reduce uncertainty about the variable's outcomes to zero. Highly complex systems with  
55 many degrees of freedom (i.e., high flexibility and expressiveness), are more information-rich than  
56 simpler or more constrained systems. The redundancy of a signal denotes the difference between  
57 how expressive the signal *could* be (i.e., proportional to the number of unique states or symbols  
58 used to transmit the signal) and the actual information rate (i.e., the entropy of each individual  
59 state or symbol). If a brain network's nodes are fully independent, then the number of bits required  
60 to express a single activity pattern is proportional to the number of nodes. The network would  
61 also be minimally redundant, since the status of every node would be needed to fully express a  
62 single brain activity pattern. If a brain network's nodes are fully coupled and identical, then the  
63 number of bits required to express a single activity pattern is proportional to the number of unique  
64 states or values any individual node can take on. Such a network would be highly redundant,  
65 since knowing any individual node's state would be sufficient to recover the full-brain activity

66 pattern. Highly redundant systems are also robust, since there is little total information loss due  
67 to removing any given observation.

68 We take as a given that brain activity is highly flexible: our brains can exhibit nearly infinite  
69 varieties of activity patterns. This flexibility implies that our brains' activity patterns are highly  
70 information rich. However, brain activity patterns are also highly structured. For example,  
71 full-brain correlation matrices are stable within (Finn et al., 2017, 2015; Gratton et al., 2018) and  
72 across (Cole et al., 2014; Glerean et al., 2012; Gratton et al., 2018; Yeo et al., 2011) individuals. This  
73 stability suggests that our brains' activity patterns are at least partially constrained, for example  
74 by anatomical, external, or internal factors. Constraints on brain activity that limit its flexibility  
75 decrease expressiveness (i.e., its information rate). However, constraints on brain activity also  
76 increase its robustness to noise (e.g., "missing" or corrupted signals may be partially recovered).  
77 For example, recent work has shown that full-brain activity patterns may be reliably recovered  
78 from only a relatively small number of implanted electrodes (Owen et al., 2020; Scangos et al.,  
79 2021). This robustness property suggests that the relevant signal (e.g., underlying factors that  
80 have some influence over brain activity patterns) are compressible.

81 To the extent that brain activity patterns contain rich task-relevant information, we should be  
82 able to use the activity patterns to accurately differentiate between different aspects of a task (e.g.,  
83 using pattern classifiers; Norman et al., 2006). For example, prior work has shown a direct  
84 correspondence between classification accuracy and the information content of a signal (Alvarez,  
85 2002). To the extent that brain activity patterns are compressible, we should be able to generate  
86 simplified (e.g., lower dimensional) representations of the data while still preserving the relevant  
87 or important aspects of the original signal. In general, information content and compressibility  
88 are often related but are also dissociable (Fig. 1). If a given signal (e.g., a representation of brain  
89 activity patterns) contains more information about ongoing cognitive processes, then the peak  
90 decoding accuracy should be high. In the simulations shown in Figure 1C we construct synthetic  
91 datasets that have high or low levels of informativeness by varying temporal autocorrelations  
92 in the data (see *Synthetic data*). If a signal is compressible, then a low-dimensional embedding  
93 of the signal will be similarly informative as the original signal. In the simulations shown in  
94 Figure 1C we construct synthetic datasets that have high or low levels of compressibility by  
95 varying the covariance structure across features (see *Synthetic data*). As shown in Figure 1D, highly



**Figure 1: Information content and compressibility.** **A. Variance explained for two images.** Consider two example images: a photograph and simple drawing. The images have the same resolutions, but their content is very different. The photograph is more information rich: when we compare the original photograph and drawing (leftmost column), we can see that the photograph captures much more detail. We can also apply principal components analysis to the images, treating the rows of the images as “observations.” Across columns, we identified the numbers 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. This analysis reveals that the drawing is more compressible: just 16 components can explain 95% of the variance in the drawing, whereas 50 components are required to explain 95% of the variance in the photograph. **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. This analysis provides another way of characterizing the compressibility of each image by showing that the same number of components can explain more variance in the drawing than in the photograph. **C. Template data from four synthetic datasets.** We constructed four synthetic datasets, each comprising 25 features (columns) observed across 100 samples (rows) from each of 10 simulated “participants.” The datasets were constructed to contain different levels of informativeness and compressibility (see *Synthetic data*). **D. Decoding accuracy by number of components.** For each synthetic dataset, we trained across-participant classifiers to decode timepoint labels. Each panel displays the decoding accuracy as a function of the number of components used to represent the data. Error ribbons denote bootstrap-estimated 95% confidence intervals.

96 informative datasets yield higher decoding accuracies than less informative datasets (i.e., the peaks  
97 of the curves in the top panels of Fig. 1D are higher than the peaks of the curves in the bottom  
98 panels). Highly compressible datasets show steeper slopes when we plot decoding accuracy as a  
99 function of the number of components used to represent the data (i.e., the slopes of the curves in  
100 the left panels of Fig. 1D are steeper than the slopes of the curves in the right panels). Whereas  
101 characterizing the informativeness and compressibility of synthetic data can be instructive, we  
102 are ultimately interested in understanding how these properties relate to brain activity patterns  
103 recorded under different cognitive circumstances.

104 Several recent studies suggest that the complexity of brain activity is task-dependent, whereby  
105 simpler tasks with lower cognitive demands are reflected by simpler and more compressible brain  
106 activity patterns, and more complex tasks with higher cognitive demands are reflected by more  
107 complex and less compressible brain activity patterns (Mack et al., 2020; Owen et al., 2021). These  
108 patterns hold even when the stimulus itself is held constant (Mack et al., 2020). These findings  
109 complement other work suggesting that functional connectivity (correlation) patterns are task-  
110 dependent (Cole et al., 2014; Finn et al., 2017; Owen et al., 2020), although see Gratton et al. (2018).  
111 Higher-order cognitive processing of a common stimulus also appears to drive more stereotyped  
112 task-related activity and functional connectivity across individuals (Hasson et al., 2008; Lerner et  
113 al., 2011; Simony & Chang, 2020; Simony et al., 2016).

114 The above studies are consistent with two potential descriptions of how cognitive processes are  
115 reflected in brain activity patterns. One possibility is that the information rate of brain activity in-  
116 creases during more complex or higher-level cognitive processing. If so, then the ability to reliably  
117 decode cognitive states from brain activity patterns should improve with task complexity or with  
118 the level (or “depth”) of cognitive processing. A second possibility is that the compressibility of  
119 brain activity patterns decreases during more complex or higher-level cognitive processing. If so,  
120 then individual features of brain recordings should carry more information (over and above the  
121 information carried by other features) during complex or high-level (versus simple or low-level)  
122 cognitive tasks. The tradeoffs between these two aspects of brain activity may also vary across  
123 brain regions or networks, for example according to each region’s functional role.

124 We used a previously collected neuroimaging dataset to estimate the extent to which each of  
125 these two possibilities might hold. The dataset we examined comprised functional magnetic res-

126    onance imaging (fMRI) data collected as participants listened to an audio recording of a 7-minute  
127    story, temporally scrambled recordings of the story, or underwent a resting state scan (Simony  
128    et al., 2016). Each of these experimental conditions evokes different depths of cognitive process-  
129    ing (Hasson et al., 2008; Lerner et al., 2011; Owen et al., 2021; Simony et al., 2016). We used  
130    across-participant classifiers to decode listening times in each condition, as a proxy for how “infor-  
131    mative” the task-specific activity patterns were (Simony & Chang, 2020). We also used principle  
132    components analysis to generate lower-dimensional representations of the activity patterns. We  
133    then repeated the classification analyses after preserving different numbers of components and  
134    examined how classification accuracy changed across the different experimental conditions.

## 135    **Results**

136    We sought to understand whether higher-level cognition is reflected by more reliable and in-  
137    formative brain activity patterns, and how compressibility of brain activity patterns relates to  
138    cognitive complexity. We developed a computational framework for systematically assessing the  
139    informativeness and compressibility of brain activity patterns recorded under different cognitive  
140    circumstances. We used across-participant decoding accuracy (see *Forward inference and decoding*  
141    *accuracy*) as a proxy for informativeness. To estimate the compressibility of the brain patterns, we  
142    used group principal components analysis (PCA) to project the brain patterns into  $k$ -dimensional  
143    spaces, for different values of  $k$  (see *Hierarchical Topographic Factor Analysis (HTFA)* and *Principal*  
144    *components analysis (PCA)*). For more compressible brain patterns, decoding accuracy should be  
145    more robust to small values of  $k$ .

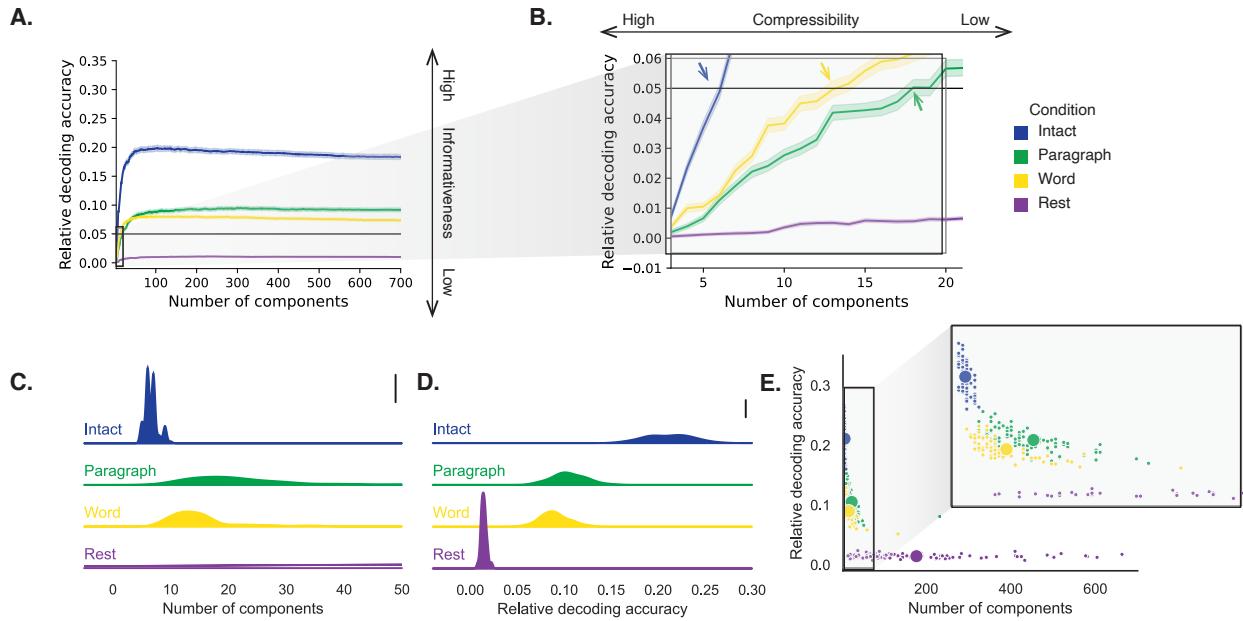
146    We analyzed a dataset collected by Simony et al. (2016) that comprised four experimental con-  
147    ditions. These conditions exposed participants to stimuli that systematically varied in cognitive  
148    engagement. In the *intact* experimental condition, participants listened to an audio recording of  
149    a 7-minute Moth Radio Hour story, *Pie Man*, by Jim O’Grady. In the *paragraph*-scrambled experi-  
150    mental condition, participants listened to a temporally scrambled version of the story, where the  
151    paragraphs occurred out of order, but where the same set of paragraphs was presented over the  
152    entire listening interval. All participants in this condition experienced the scrambled paragraphs  
153    in the same order. In the *word*-scrambled experimental condition, participants listened to a tem-

porally scrambled version of the story, where the words occurred in a random order. Again, all participants in this condition experienced the scrambled words in the same order. Finally, in the *rest* experimental condition, participants lay in the scanner with no overt stimulus, while keeping their eyes open and blinking as needed. This public dataset provided a convenient means for testing our hypothesis that different levels of cognitive processing and engagement affect how informative and compressible the associated brain patterns are.

To evaluate the relation between informativeness and compressibility for brain activity from each experimental condition, we trained a series of across-participant temporal decoders on compressed representations of the data. Figure 2A displays the decoding accuracy as a function of the number of principal components used to represent the data (also see Fig. S1). Several patterns were revealed by the analysis. First, in general (i.e., across experimental conditions), decoding accuracy tends to improve as the number of components are increased. However, decoding accuracy peaked at higher levels for experimental conditions that exposed participants to cognitively richer stimuli (Fig. 2D). The peak decoding accuracy was highest for the “intact” condition (versus paragraph:  $t(99) = 35.205, p < 0.001$ ; versus word:  $t(99) = 43.172, p < 0.001$ ); versus rest:  $t(99) = 81.361, p < 0.001$ ), next highest for the “paragraph” condition (versus word:  $t(99) = 6.243, p < 0.001$ ; versus rest:  $t(99) = 50.748, p < 0.001$ ), and next highest for the “word” condition (versus rest:  $t(99) = 48.791, p < 0.001$ ). This ordering implies that cognitively richer conditions evoke more stable brain activity patterns across people.

The cognitively richer conditions also displayed steeper initial slopes. For example, the intact condition decoders reached an arbitrarily chosen threshold of 5% accuracy using fewer components than the paragraph condition decoders ( $t(99) = -7.429, p < 0.001$ ) or word condition decoders ( $t(99) = -7.300, p < 0.001$ ), and decoding accuracy never exceeded 5% for the rest condition. This suggests that brain activity patterns evoked by cognitively richer conditions are more compressible, such that representing the data using the same number of principal components provides more information to the temporal decoders (Figs. 2B, C). Taken together, as shown in Figure 2E, we found that brain activity patterns evoked by cognitively richer conditions tended to be both more informative (i.e., associated with higher peak decoding accuracies) *and* more compressible (i.e., requiring fewer components to achieve the 5% accuracy threshold).

If informativeness (to the temporal decoders) and compressibility vary with the cognitive



**Figure 2: Decoding accuracy and compressibility.**

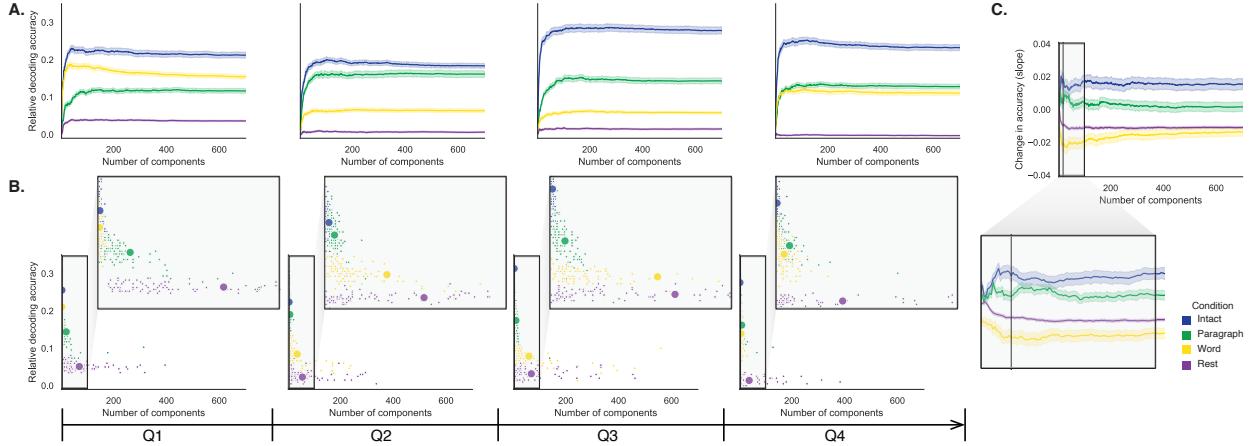
**A. Decoding accuracy by number of components.** Ribbons of each color display cross-validated decoding performance for each condition (intact, paragraph, word, and rest), as a function of the number of components (features) used to represent the data. For each condition, decoding accuracy has been normalized by subtracting expected “chance” decoding performance (estimated as  $\frac{1}{T}$ , where  $T$  is the number of timepoints in the given condition). This normalization was intended to enable fairer comparisons across conditions; a relative decoding accuracy of 0.0 denotes “chance” performance. 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. All error ribbons in Panels A and B denote bootstrap-estimated 95% confidence intervals.

**C. Estimating “compressibility” for each condition.** The probability density plots display the numbers of components required to reach at least 5% (corrected) decoding accuracy, across all cross validation runs. For the “rest” condition, where decoding accuracy never reached 5% accuracy, we display the distribution of the numbers of components required to achieve the peak decoding accuracy in each fold. (This distribution is nearly flat, as also illustrated in Panel E.) The scale bar denotes a height of 0.01.

**D. Estimating “informativeness” for each condition.** The probability density plots display the peak (corrected) decoding accuracies achieved in each condition, across all cross validation runs. The scale bar denotes a height of 0.01.

**E. Informativeness versus compressibility.** Each dot displays the minimum number of components required to achieve at least 5% corrected decoding accuracy ( $x$ -coordinate; for the rest condition the numbers of components are estimated using the peak decoding accuracies as in Panel C) and the peak (corrected) decoding accuracy ( $y$ -coordinate). The smaller dots correspond to individual cross validation runs and the larger dots display the averages across runs. The inset displays a zoomed-in view of the indicated region in the main panel.

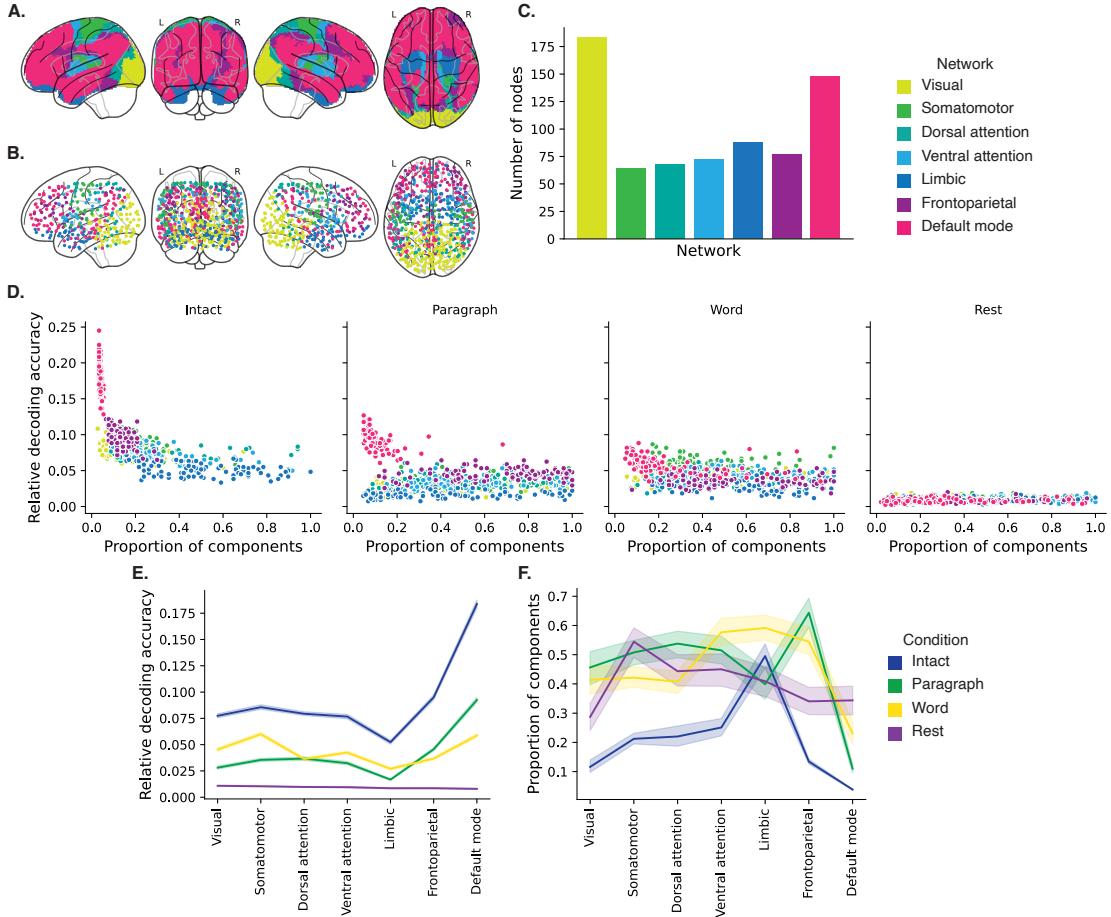


**Figure 3: Changes in decoding accuracy and compression over time. A. Decoding accuracy by number of components, by story segment.** Each family of curves is plotted in the same format as Figure 2A but reflects data only from one quarter of the dataset. **B. Informativeness versus compressibility by condition and segment.** Each scatter plot is in the same format as Figure 2E, but reflects data only from one quarter of the dataset. **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, third, and fourth quarters of the dataset (corresponding to the columns of Panels A and B). The *y*-axis denotes the slopes of the regression lines. The black vertical line marks  $k = 20$  components, as referenced in the main text. All error ribbons denote bootstrap-estimated 95% confidence intervals.

richness of the stimulus, might these measures also vary over time *within* a given condition? For example, participants in the intact condition might process the ongoing story more deeply later on in the story (compared with earlier in the story) given the additional narrative background and context they had been exposed to by that point. To examine this possibility, we divided each condition into four successive time segments. We computed decoding curves (Fig. 3A) and the numbers of components required to achieve 5% decoding accuracy (Fig. 3B) for each segment and condition. We found that, in the two most cognitively rich conditions (intact and paragraph), both decoding accuracy and compressibility, as reflected by the change in decoding curves, increased with listening time (e.g., at the annotated reference point of  $k = 20$  components in Fig. 3C: intact:  $t(99) = 7.915, p < 0.001$ ; paragraph:  $t(99) = 2.354, p = 0.021$ ). These changes may reflect an increase in comprehension or depth of processing with listening time. In contrast, the decoding accuracy and compressibility decreased with listening time in the word condition ( $t(99) = -10.747, p < 0.001$ ) and rest condition ( $t(99) = -22.081, p < 0.001$ ). This might reflect the depletion of attentional resources in the less-engaging word and rest conditions.

198 These results make some intuitive sense. As the contextual information available to participants  
199 increases (i.e., over time in the cognitively rich intact and paragraph conditions), it makes sense that  
200 this might constrain neural responses to a greater extent. While this pattern may not necessarily  
201 hold for *every* possible story or stimulus, we suspect that it is generally the case that our knowledge  
202 about what is happening in a story tends to increase as we experience more about it. In turn, this  
203 could lead to greater consistency in different people's interpretations of and neural responses to  
204 the stimulus. Similarly, as participants are left to "mind wander," or as they experience mental  
205 fatigue (i.e., over time in the less cognitively rich word and rest conditions), we suggest that this  
206 might lead to greater variability in neural responses across people, resulting in lower decoding  
207 accuracy. Again, it is not necessarily the case that every possible "unengaging" stimulus will  
208 lead to greater neural variability as time progresses, but we suspect this phenomenon is likely to  
209 hold for a variety of such stimuli. These findings replicate at least to a limited extent (e.g., across  
210 both cognitively rich conditions and across both cognitively impoverished conditions, and for the  
211 different groups of participants in each of those conditions). However, determining whether these  
212 patterns generalize to other stimuli would require additional study (with new stimuli).

213 If the informativeness and compressibility of brain activity patterns vary over time, do these  
214 properties might also vary across brain networks? We used a network parcellation identified by Yeo  
215 et al. (2011) to segment the brain into seven distinct networks. The networks can be sorted (roughly)  
216 in order from lower-level to higher-level cortex as follows (Figs. 4A–C): visual, somatomotor,  
217 dorsal attention, ventral attention, limbic, frontoparietal, and default mode. Next, we computed  
218 decoding curves separately for the activity patterns within each network and identified each  
219 network's inflection point, for each experimental condition. Moving from low-order networks  
220 to higher-order networks, we found that decoding accuracy tended to increase in the higher-  
221 level experimental conditions and decrease (slightly) in the lower-level experimental conditions  
222 (Fig. 4D, E; Spearman's rank correlation between decoding accuracy and network order: intact:  
223  $\rho = 0.362, p < 0.001$ ; paragraph:  $\rho = 0.441, p < 0.001$ ; word:  $\rho = -0.102, p = 0.007$ ; rest:  $\rho =$   
224  $-0.354, p < 0.001$ ). This suggests that higher-order networks may carry more content-relevant  
225 or stimulus-driven "information." We found no clear trends in the proportions of components  
226 required to achieve 5% decoding accuracy across networks or conditions (Fig. 4F). We note that  
227 the limbic network we considered here often overlaps with low (imaging) signal regions, and

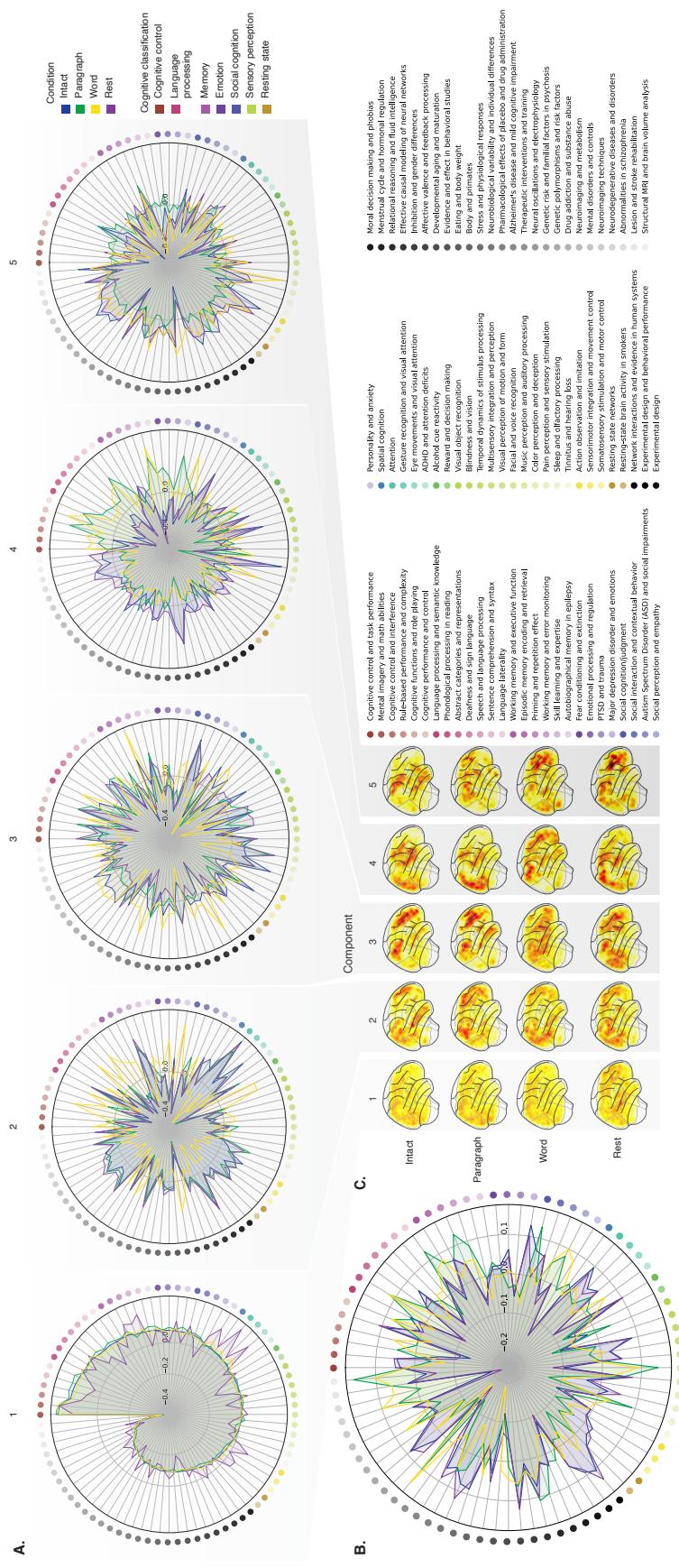


**Figure 4: Network-specific decoding accuracy and compression.** **A. Network parcellation map.** The glass brain displays the location of each of seven networks identified by Yeo et al. (2011). **B. Node labels.** We assigned network labels to each of 700 Hierarchical Topographic Factor Analysis-derived nodes (see *Hierarchical topographic factor analysis (HTFA)*). For each node, we evaluated its radial basis function (RBF) at the locations of every voxel in the parcellation map displayed in Panel A. We summed up these RBF weights separately for the voxels in each network. We defined each node's network label as its highest-weighted network across all voxels. **C. Per-network node counts.** The bar plot displays the total number of HTFA-derived nodes assigned to each network. **D. Informativeness versus compressibility by network and condition.** Each dot corresponds to a single cross validation run. The dots' coordinates denote the minimum proportion of components (relative to the number of nodes in the given network) required to achieve at least 5% corrected decoding accuracy ( $x$ -coordinate; for the rest condition these proportions are estimated using the peak decoding accuracies) and peak (corrected) decoding accuracy ( $y$ -coordinate). We repeated the analysis separately for each network (color) and experimental condition (column). **E. Informativeness by network.** For each network, sorted (roughly) from lower-order networks on the left to higher-order networks on the right, the curves display the peak (corrected) decoding accuracy for each experimental condition (color). **F. Compressibility by network.** For each network, the curves display the proportion of components (relative to the total possible number of components displayed in Panel C) required to achieve at least 5% decoding accuracy (or, for the rest condition, peak decoding accuracy, as described above). Error ribbons in Panels E and F denote bootstrap-estimated 95% confidence intervals.

228 therefore it may be difficult to draw strong conclusions about this network's informativeness or  
229 compressibility.

230 Whereas the above analyses examine different networks in isolation, how does full-brain  
231 (i.e., potentially multi-network) activity patterns reflected by different principal components vary  
232 across different experimental conditions? As shown in Figure 5, we used Neurosynth (Rubin et al.,  
233 2017) to identify, for each component, the associations with each of 80 themes (see *Reverse inference*).  
234 In general, the first principal components across all of the experimental conditions tended to weight  
235 most heavily on themes related to cognitive control, memory, language processing, attention, and  
236 perception. Other components appeared to vary more across conditions.

237 To gain further insights into which brain functions might be most closely associated with  
238 the top-weighted components from each experimental condition, we manually grouped each  
239 Neurosynth-derived topic into a smaller set of core cognitive functions. Separately for each com-  
240 ponent, we computed the average weightings across all topics that were tagged as being associated  
241 with each of these cognitive functions (Figs. 6A, S5A). To help visualize these associations, we used  
242 the patterns of associations for each component to construct graphs whose nodes were experimen-  
243 tal conditions and cognitive functions (Figs. 6B, S5B). We also computed correlations between the  
244 sets of per-topic weightings from each of the top-weighted components from each experimental  
245 condition (Fig. 6C, D) and between the brain maps for each condition's components (Fig. S5C, D).  
246 Taken together, we found that each component appeared to weight on a fundamental set of cogni-  
247 tive functions that varied by experimental condition. For example, the top principal components  
248 from every condition weighted similarly (across conditions) on the full set of Neurosynth topics  
249 (Fig. 5A) and cognitive functions (Figs. 6A, B and S5 A, B), suggesting that these components might  
250 reflect a set of functions or activity patterns that are common across all conditions. The second  
251 components' weightings were similar across the intact, paragraph, and rest conditions (highest-  
252 weighted functions: cognitive control, memory, social cognition, and resting state), but different  
253 for the word condition (highest-weighted functions: sensory perception and cognitive control).  
254 The fourth components' weighting grouped the paragraph and word conditions (highest-weighted  
255 functions: memory, language processing, and cognitive control) and the intact and rest conditions  
256 (highest-weighted functions: emotion, social cognition). We also used ChatGPT (OpenAI, 2023)  
257 to sort the list of manually tagged cognitive functions from lowest-level to highest-level (Tab. S2,

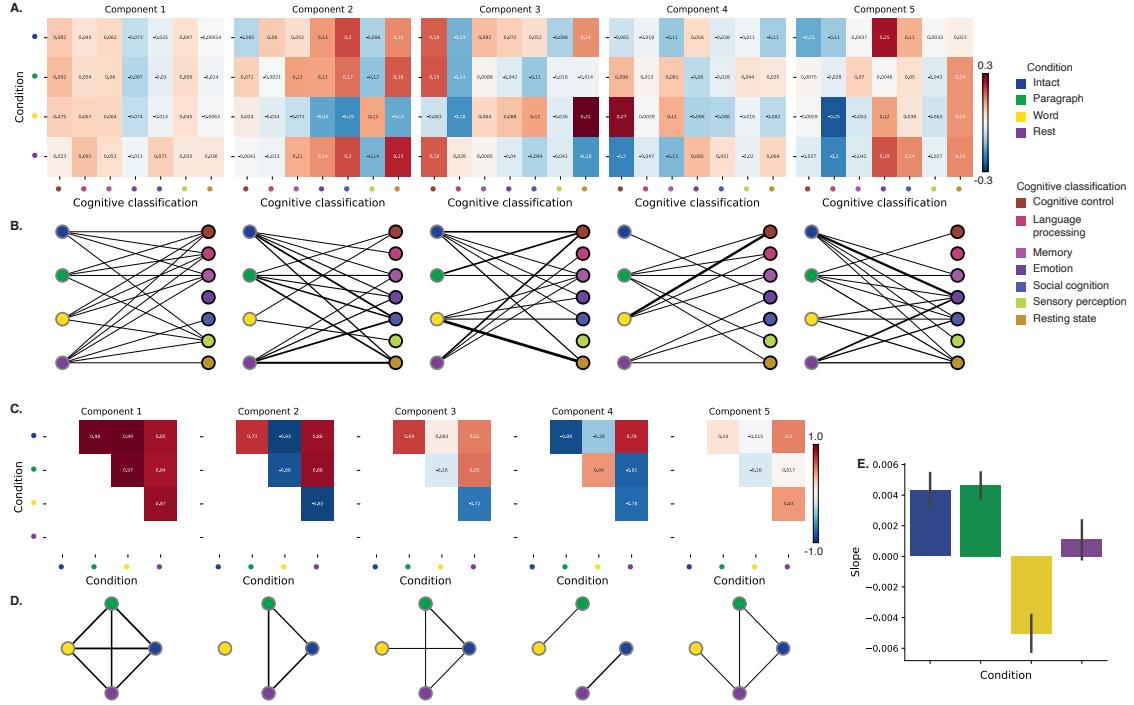


**Figure 5: Neurosynth topic weightings by component.** We used a reverse inference procedure (see *Reverse inference*) to estimate the correspondences between brain images and a set of 80 topics derived from the Neurosynth database of neuroimaging articles (Rubin et al., 2017). **A. Topic correlations by component.** For each of the top five highest-weighted principal components (columns) derived from each experimental condition (colors), the radar plots display the correlations between the component images and the per-topic images derived for each topic (topic identities are denoted by the blue dots; a legend for the topic labels is in the lower right of the figure). An annotated list of the top-weighted topics for each component and condition may be found in Figure S2 and the top-weighted terms for each topic may be found in Table S1. **B. Topic correlations averaged across components.** The radar plot is in the same format as the plots in Panel A, but here we display the per-condition average correlations across the top five components (for each condition) reflected in Panel A. **C. Component images.** Each plot displays a right sagittal view of a glass brain projection of the top five principal components (columns) for each experimental condition (rows). Additional projections for each component may be found in Figure S3.

258 Fig. 6E; also see *Ranking cognitive processes*). We found that higher-level functions tended to be  
259 weighted on more heavily by top components from the intact and paragraph conditions than lower-  
260 level functions (intact vs. word:  $t(198) = 11.059, p < 0.001$ ; intact vs. rest:  $t(198) = 3.699, p < 0.001$ ;  
261 paragraph vs. word:  $t(198) = 13.504, p < 0.001$ ; paragraph vs. rest:  $t(198) = 4.812, p < 0.001$ ; also  
262 see *Ranking cognitive processes*). The top components from the word condition showed the opposite  
263 tendency, whereby *lower*-level functions tended to be weighted on more heavily than higher-level  
264 functions (word vs. rest:  $t(198) = -7.315, p < 0.001$ ). The weighting trends for the intact and  
265 paragraph conditions were not reliably different ( $t(198) = -0.479, p = 0.633$ ). The components  
266 from the rest condition showed almost no differences in the weights associated with high-level  
267 versus low-level functions (rest vs. 0:  $t(99) = 1.836, p = 0.081$ ). These findings suggest that when  
268 participants were engaged more strongly (in the more engaging intact and paragraph conditions),  
269 their dominant neural patterns reflected higher-level cognitive functions. In contrast, when partic-  
270 ipants were engaged less strongly (in the less engaging word and rest conditions), their dominant  
271 neural patterns reflected lower-level cognitive functions. Although they were highly statistically  
272 reliable, it is also important to note that these latter effects are also relatively small (e.g., the slopes  
273 for *all* of the experimental conditions are numerically close to zero; Fig. 6E). We suggest that this  
274 phenomenon may merit further investigation in future work.

## 275 Discussion

276 We examined fMRI data collected as participants listened to an auditory recording of a story, scram-  
277 bled recordings of the story, or underwent a resting state scan. We found that cognitively richer  
278 stimuli evoked more reliable (i.e., consistent across people) and information rich brain activity pat-  
279 terns. The brain patterns evoked by cognitively richer stimuli were also more compressible, in that  
280 each individual component provided more “signal” to temporal decoders relative to components  
281 of data from less cognitively rich conditions (Fig. 2). Over time (e.g., as the experiment progressed),  
282 these phenomena were strengthened. Specifically, across story segments, data from more cogni-  
283 tively rich conditions became more informative and compressible, and data from less cognitively  
284 rich conditions became *less* informative and compressible (Fig. 3). We also repeated these analyses  
285 separately for different brain networks. We found that networks traditionally associated with



**Figure 6: Summary of functions associated with top-weighted components by condition. A. Top-weighted topics by condition.** Here we display per-condition (rows, indicated by colored dots) topic correlations, averaged across topics that pertain to each of several broad cognitive functions (columns within each sub-panel, indicated by colored dots). Each sub-panel reflects correlations for the components indicated in the panel titles. A legend for the condition and several core cognitive function classifications is displayed in the lower right of the figure. Table S1 provides a list of each topic's top-weighted terms, along with each topic's manually labeled cognitive classification. A full list of the topics most highly associated with each component may be found in Figure S2. **B. Associations between per-condition components and cognitive functions.** The network plots denote positive average correlations between the component images for each condition (gray-outlined dots on the left sides of each network; colors denote conditions) and topic-specific brain maps associated with each indicated cognitive function (black-outlined dots on the right sides of each network; colors denote cognitive functions). The line thicknesses are proportional to the correlation values (correlation coefficients are noted in the heat maps in Panel A). **C. Correlations between each principal component, by condition.** The heat maps display the correlations between the brain maps (Fig. S3) for each principal component (sub-panel), across each pair of conditions (rows and columns of each sub-panel's matrix, indicated by colored dots). **D. Associations between per-condition components, by component.** Each sub-panel's network plot summarizes the pattern of correlations between the  $n^{\text{th}}$  top-weighted principal components (sub-panel) for each experimental condition (gray-outlined dots). The line thicknesses are proportional to the correlation values (correlation coefficients are noted in the heat maps in Panel C). **E. Change in weights by condition.** The bar heights reflect the slopes of regression lines, fit separately to the top five components from each condition, between the ChatGPT-derived “rank” of each cognitive classification and the correlations between the component and topic maps associated with cognitive processes at the given rank (see *Ranking cognitive processes*) Error bars denote bootstrap-estimated 95% confidence intervals. Also see Fig. S5 for additional information.

286 higher-level cognitive functions tended to provide more informative brain patterns than networks  
287 traditionally associated with lower level cognitive functions (Fig. 4). Finally, we examined the most  
288 dominant components of the brain activity patterns from each experimental condition. We used a  
289 reverse inference approach (Rubin et al., 2017) to identify the terms in the neuroimaging literature  
290 most commonly associated with the corresponding maps. As summarized in Figure 6E, we found  
291 that the intact and paragraph conditions tended to weight on higher-level cognitive processes  
292 more than lower-level cognitive processes, whereas the word condition weighted on lower-level  
293 processes more than higher-level processes and the rest condition showed no difference in high-  
294 level versus low-level weighting. Taken together, our findings indicate that the informativeness  
295 and compressibility of our brain activity patterns are task-dependent, and these properties change  
296 systematically with factors like cognitive richness and depth of processing.

297 Our explorations of informativeness and compressibility are related to a much broader litera-  
298 ture on the correlational and causal structure of brain activity patterns and networks (Adachi et  
299 al., 2012; Bassett & Sporns, 2017; Brovelli et al., 2004; Bullmore & Sporns, 2009; Dhamala et al.,  
300 2008; Korzeniewska et al., 2008; Lynn & Bassett, 2021; Owen et al., 2021; Preti et al., 2017; Rogers  
301 et al., 2007; Rubinov & Sporns, 2010; Sizemore et al., 2018; Smith, Beckmann, et al., 2013; Smith,  
302 Vidaurre, et al., 2013; Sporns & Betzel, 2016; Sporns & Honey, 2006; Sporns & Zwi, 2004; Srinivasan  
303 et al., 2007; Tomasi & Volkow, 2011; Yeo et al., 2011). Correlations or causal associations between  
304 different brain regions simultaneously imply that full-brain activity patterns will be compressible  
305 and also that those activity patterns will contain redundancies. For example, the extent to which  
306 activity patterns at one brain area can be inferred or predicted from activity patterns at other  
307 areas (e.g., Owen et al., 2020; Scangos et al., 2021), reflects overlap in the information available  
308 in or represented by those brain areas. If brain patterns in one area are recoverable using brain  
309 patterns in another area, then a “signal” used to convey the activity patterns could be compressed  
310 by removing the recoverable activity. Predictable (and therefore redundant) brain activity patterns  
311 are also more robust to signal corruption. For example, even if the activity patterns at one region  
312 are unreadable or unreliable at a given moment, that unreliability could be compensated for by  
313 other regions’ activity patterns that were predictive of the unreliable region. Whereas compress-  
314 ible brain patterns are robust to spatial signal corruption, high versus low informativeness reflects  
315 a similar (though dissociable; e.g., Fig. 1) tradeoff between expressiveness and robustness of *tem-*

316 *poral* patterns. Highly informative brain patterns (by our measure; i.e., patterns that yield greater  
317 temporal decoding accuracy) are expressive about ongoing experiences or cognitive states, since  
318 each moment's pattern is reliably distinguishable from other moments' patterns. However, when  
319 each moment's pattern is unique, brain activity becomes less robust to temporal signal corruption.  
320 Our finding that brain activity patterns becomes more informative (i.e., less robust to temporal  
321 signal corruption) and compressible (i.e., more robust to spatial signal corruption) when cognitive  
322 engagement is higher suggests that our brain may optimize its activity patterns to prioritize either  
323 temporal or spatial robustness, according to task demands.

324 Our findings that informativeness and compressibility change with task demands may follow  
325 from task-dependent changes in full-brain correlation patterns. A number of prior studies have  
326 found that so-called “functional connectivity” (i.e., correlational) patterns vary reliably across  
327 tasks, events, and situations (Cole et al., 2014; Owen et al., 2021; Simony et al., 2016; Smith et  
328 al., 2009). By examining how these task-dependent changes in correlations affect informativeness  
329 and compressibility, our work suggests a potential reason why the statistical structure of brain  
330 activity patterns might vary with cognitive task or with cognitive demands. For lower-level tasks,  
331 or for tasks that require relatively little “deep” cognitive processing, our brains may optimize  
332 activity patterns for robustness and redundancy over expressiveness, for example to maximize  
333 reliability. For higher-level tasks, or for tasks that require deeper cognitive processing, our brains  
334 may sacrifice some redundancy in favor of greater expressiveness.

335 One potential limitation of our work concerns how our measure of informativeness might  
336 generalize across tasks or cognitive representations and processes. Our use of across-participant  
337 temporal decoding accuracy as a proxy for informativeness is motivated in part by prior work that  
338 introduced across-participant similarity (in time-varying response to a stimulus) as a means of  
339 identifying stimulus-driven brain activity patterns (Simony et al., 2016). Intuitively, only activity  
340 patterns are driven by the stimulus would be expected to synchronize (i.e., be time-locked  
341 to the stimulus) across participants. This approach implicitly removes idiosyncratic responses  
342 (e.g., neural patterns that are *not* similar across people). However, there are also some published  
343 examples, including in our own prior work, that indicate that some types of stimulus-evoked  
344 activity will be missed by across-participant comparisons. For example we have reported how  
345 brain regions like the ventromedial prefrontal cortex (vmPFC) show stimulus-driven responses

346 that are, for the most part, not similar across people. In that paper (and drawing on other work),  
347 we suggest that the vmPFC seems to represent or support highly idiosyncratic internal states, like  
348 affective responses. Although we would consider the vmPFC to be a “high-level” region (e.g., we  
349 consider affect to be a relatively high-level aspect of cognition), the measure of informativeness that  
350 we used in our current study would identify regions like the vmPFC as having *low* informativeness.  
351 This is because across-participant decoding accuracy (our proxy measure for informativeness) will  
352 only be high for representations or responses that are common across people.

353 Relatedly, even in the experimental conditions we describe as “less cognitively engaging,” we  
354 do not think it likely that high-level thought or cognitive processing is absent. Rather, we sug-  
355 gest that these high-level representations tend to be more idiosyncratic when the stimulus is less  
356 engaging, and therefore less constraining on people’s thoughts. Nonetheless, even during highly  
357 engaging tasks, people may engage in idiosyncratic stimulus-driven processes. For example, peo-  
358 ple might retrieve personal information as they listened to the story. Those retrievals could happen  
359 at different times for different people according to each individual’s prior experiences, and even  
360 when those sorts of retrievals happen to be temporally synchronized across people, the specific  
361 memories or information being retrieved might still be idiosyncratic. Our measure of informa-  
362 tiveness is insensitive to these processes. Further, even in response to an identical stimulus, task  
363 instructions or participants’ internal goals could change the relationship between compressibility  
364 and informativeness. Some work has shown that the “dimensionality” of neural representations  
365 can change systematically with task complexity, even in response to an identical stimulus (Mack et  
366 al., 2020). Taken together, we expect that the way we have defined informativeness in this paper,  
367 and the specific dataset we examined, are likely to have influenced our findings. While we see our  
368 approach as a reasonable first step, we also suggest that future work should explore alternative  
369 measures of informativeness and compressibility, and should examine how these measures vary  
370 across different tasks and datasets.

371 In the information theory sense (Shannon, 1948), when a signal is transmitted using a fixed  
372 alphabet of “symbols,” the information rate decreases as the signal is compressed (e.g., fewer sym-  
373 bols transmitted per unit time, using an alphabet with fewer symbols, etc.). Our finding that each  
374 individual brain component (symbol) becomes more informative as cognitive richness increases  
375 suggests that the “alphabet” of brain activity patterns is also task-dependent. Other work suggests

376 that the representations that are *reflected* by brain activity patterns may also change with task de-  
377 mands. For example, our brains may represent the same perceptual stimulus differently depending  
378 on which aspects of the stimulus or which combinations of features are task-relevant (Mack et al.,  
379 2020).

380 We found that different brain networks varied in how informative and compressible their  
381 activity patterns were across experimental conditions (e.g., Fig. 4). This might follow from evolu-  
382 tionary optimizations that reflect the relevant constraints or demands placed on those networks.  
383 One possibility is that cortex is organized in a hierarchy of networks “concerned with” or selective  
384 to different levels of processing or function. To the extent that different levels of processing (e.g.,  
385 low-level sensory processing versus “deeper” higher-level processing) reflect different stimulus  
386 timescales (e.g., Manning, 2020), the network differences we observed might also relate to the  
387 timescales at which each network is maximally sensitive (Baldassano et al., 2017; Hasson et al.,  
388 2008; Lerner et al., 2011; Regev et al., 2018).

389 Our reverse inference analyses (Figs. 5, 6) also provide some insights into how neural activity  
390 patterns change with cognitive engagement or task demands. Prior work has shown that the  
391 components and network “parcels” identified through covarying activity patterns can be highly  
392 similar even across different tasks (including “rest,” e.g., Laird et al., 2011; Smith et al., 2009).  
393 We replicated this basic finding in that the first principal components from all four experimental  
394 conditions were strikingly similar (e.g., see the leftmost columns of Figs. 5A and 6C, D). We also  
395 found some small, though statistically reliable, systematic changes in the weights associated with  
396 different cognitive functions across conditions (Fig. 6E). This result provided an additional way of  
397 characterizing network-level differences across conditions (Fig. 4E). Taken together, these findings  
398 suggest that although similar networks may be involved in different tasks, the ways in which those  
399 networks are engaged may vary systematically with task demands.

## 400 Concluding remarks

401 Cognitive neuroscientists are still grappling with basic questions about the fundamental “rules”  
402 describing how our brains respond, and about how brain activity patterns and the associated  
403 underlying cognitive representations and computations are linked. We identified two aspects of

404 brain activity patterns, informativeness and compressibility, that appear to change systematically  
405 with task demands and across brain networks. We speculate that these changes may reflect ongoing  
406 tradeoffs between how robust to signal corruption versus how expressive about ongoing cognitive  
407 states our brains' activity patterns are. Our work also provides a new framework for evaluating  
408 these tradeoffs in other datasets, or in future studies.

## 409 Methods

410 We measured properties of recorded neuroimaging data under different task conditions that varied  
411 systematically in cognitive engagement and depth of processing. We were especially interested in  
412 how *informative* and *compressible* the activity patterns were under these different conditions (Fig. 1).

413 Unless otherwise noted, all statistical tests are two-sided, all error bars and error ribbons  
414 denote bootstrap-estimated 95% confidence intervals across participants, and all reported *p*-values  
415 are corrected for multiple comparisons using the Benjamini-Hochberg procedure (Benjamini &  
416 Hochberg, 1995).

### 417 Functional neuroimaging data collected during story listening

418 We examined an fMRI dataset collected by Simony et al. (2016) that the authors have made publicly  
419 available at [arks.princeton.edu/ark:/88435/dsp015d86p269k](https://arks.princeton.edu/ark:/88435/dsp015d86p269k). The dataset comprises neuroimaging  
420 data collected as participants listened to an audio recording of a story (intact condition; 36 par-  
421 ticipants), listened to temporally scrambled recordings of the same story (17 participants in the  
422 paragraph-scrambled condition listened to the paragraphs in a randomized order and 36 in the  
423 word-scrambled condition listened to the words in a randomized order), or lay resting with their  
424 eyes open in the scanner (rest condition; 36 participants). Full neuroimaging details may be found  
425 in the original paper for which the data were collected (Simony et al., 2016). Procedures were  
426 approved by the Princeton University Committee on Activities Involving Human Subjects, and by  
427 the Western Institutional Review Board (Puyallup, WA). All subjects were native English speakers  
428 with normal hearing and provided written informed consent. We have excerpted the relevant  
429 portions of the dataset documentation here to provide information about the scanning parameters  
430 and preprocessing steps used to generate the data we analyzed (the original descriptions may be

431 found at the above link):

432 Subjects were scanned in a 3T full-body MRI scanner (Skyra; Siemens) with a sixteen-  
433 channel head coil. For functional scans, images were acquired using a T2\* weighted  
434 echo planer imaging (EPI) pulse sequence [repetition time (TR), 1500 ms; echo time (TE),  
435 28 ms; flip angle, 64°], each volume comprising 27 slices of 4 mm thickness with 0 mm  
436 gap; slice acquisition order was interleaved. In-plane resolution was  $3 \times 3 \text{ mm}^2$  [field of  
437 view (FOV),  $192 \times 192 \text{ mm}^2$ ]. Anatomical images were acquired using a T1-weighted  
438 magnetization-prepared rapid-acquisition gradient echo (MPRAGE) pulse sequence  
439 (TR, 2300 ms; TE, 3.08 ms; flip angle 9°;  $0.89 \text{ mm}^3$  resolution; FOV, 256 mm $^2$ ). To  
440 minimize head movement, subjects' heads were stabilized with foam padding. Stimuli  
441 were presented using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). Subjects  
442 were provided with an MRI compatible in-ear mono earbuds (Sensimetrics Model S14),  
443 which provided the same audio input to each ear. MRI-safe passive noise-canceling  
444 headphones were placed over the earbuds, for noise removal and safety.

445 Functional data were preprocessed and analyzed using FSL ([www.fmrib.ox.ac.uk/fsl](http://www.fmrib.ox.ac.uk/fsl)),  
446 including correction for head motion and slice-acquisition time, spatial smoothing (6  
447 mm FWHM Gaussian kernel), and high-pass temporal filtering (140 s period). Prepro-  
448 cessed data were aligned to coplanar and high-resolution anatomicals and the standard  
449 MNI152 brain, and interpolated to 3-mm isotropic voxels.

450 The intact and word conditions each comprised 300 TRs (7.5 minutes) per participant. The para-  
451 graph condition comprised 272 TRs (6.8 minutes) per participant. The rest condition comprised  
452 400 TRs (10 minutes) per participant.

#### 453 **Hierarchical topographic factor analysis (HTFA)**

454 Following our prior related work, we used HTFA (Manning et al., 2018) to derive a compact  
455 representation of the neuroimaging data. In brief, this approach approximates the timeseries  
456 of voxel activations (44,415 voxels) using a much smaller number of radial basis function (RBF)  
457 nodes (in this case, 700 nodes, as determined by an optimization procedure; Manning et al., 2018).  
458 This provides a convenient representation for examining full-brain activity patterns and network

459 dynamics. All of the analyses we carried out on the neuroimaging dataset were performed in  
460 this lower-dimensional space. In other words, each participant's data matrix was a number-of-  
461 timepoints ( $T$ ) by 700 matrix of HTFA-derived factor weights (where the row and column labels  
462 were matched across participants). Code for carrying out HTFA on fMRI data may be found as  
463 part of the BrainIAK toolbox (Capota et al., 2017; Kumar et al., 2021), which may be downloaded  
464 at [brainiak.org](http://brainiak.org).

465 We also considered alternative approaches to obtaining compact representations of the neu-  
466 roimaging data, including network parcellations (e.g., Gordon et al., 2016; Schaefer et al., 2018).  
467 Whereas network parcellations are typically derived from large resting state datasets, HTFA may  
468 be applied to much smaller datasets. In our prior work, we showed that HTFA applied to the same  
469 dataset used here can explain full-brain activity to within a maximum of 0.25 standard deviations  
470 of each voxel's observed activity in the original dataset, taken across all voxels, images, and par-  
471 ticipants, using the 700-node representation we also employed here (Manning et al., 2018). Some  
472 of the explanatory power of HTFA comes from the fact that each node's explanatory power falls  
473 off smoothly with distance to its center. Intuitively, the result is a representation that looks like a  
474 lightly spatially smoothed version of the original data, but where the degree of smoothing varies  
475 across the brain according to how spatially autocorrelated the local activity patterns are.

## 476 Principal components analysis (PCA)

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

## 482 Temporal decoding

483 We sought to identify neural patterns that reflected participants' ongoing cognitive processing of  
484 incoming stimulus information. As reviewed by Simony et al. (2016), one way of homing in on  
485 these stimulus-driven neural patterns is to compare activity patterns across individuals. In partic-

486 ular, neural patterns will be similar across individuals to the extent that the neural patterns under  
487 consideration are stimulus-driven, and to the extent that the corresponding cognitive representa-  
488 tions are reflected in similar spatial patterns across people (Simony & Chang, 2020). Following this  
489 logic, we used an across-participant temporal decoding test developed by Manning et al. (2018) to  
490 assess the degree to which different neural patterns reflected ongoing stimulus-driven cognitive  
491 processing across people. The approach entails using a subset of the data to train a classifier to  
492 decode stimulus timepoints (i.e., moments in the story participants listened to) from neural pat-  
493 terns. We use decoding (forward inference) accuracy on held-out data, from held-out participants,  
494 as a proxy for the extent to which the inputted neural patterns reflected stimulus-driven cognitive  
495 processing in a similar way across individuals.

#### 496 **Forward inference and decoding accuracy**

497 We used an across-participant correlation-based classifier to decode which stimulus timepoint  
498 matched each timepoint's neural pattern. For a given value of  $k$  (i.e., number of principal compo-  
499 nents), we first used group PCA to project the data from each condition into a shared  $k$ -dimensional  
500 space. Next, we divided the participants into two groups: a template group,  $\mathcal{G}_{\text{template}}$  (i.e., training  
501 data), and a to-be-decoded group,  $\mathcal{G}_{\text{decode}}$  (i.e., test data). We averaged the projected data within  
502 each group to obtain a single  $T$  by  $k$  matrix for each group. Next, we correlated the rows of the two  
503 averaged matrices to form a  $T$  by  $T$  decoding matrix,  $\Lambda$ . In this way, the rows of  $\Lambda$  reflected time-  
504 points from the template group, while the columns reflected timepoints from the to-be-decoded  
505 group. We used  $\Lambda$  to assign temporal labels to each timepoint (row) from the test group's ma-  
506 trix, using the row of the training group's matrix with which it was most highly correlated. We  
507 repeated this decoding procedure, but using  $\mathcal{G}_{\text{decode}}$  as the template group and  $\mathcal{G}_{\text{template}}$  as the  
508 to-be-decoded group. Given the true timepoint labels (for each group), we defined the decoding  
509 accuracy as the average proportion of correctly decoded timepoints, across both groups (where  
510 chance performance is  $\frac{1}{T}$ ). In Figures 2 and 3 we report the decoding accuracy for each condition  
511 and value of  $k$ , averaged across  $n = 100$  cross validation folds.

512 **Reverse inference**

513 To help interpret the brain activity patterns we found within the contexts of other studies, we  
514 created summary maps of each principal component, for each experimental condition. Each  
515 principal component comprises 700 “weights” on each of the HTFA-derived RBF nodes (see  
516 *Hierarchical Topographic Factor Analysis*). For each node, we evaluated its RBF at the locations of  
517 every voxel in the standard 2 mm MNI152 template brain and multiplied the RBF by the node’s  
518 weight. The sum of these weighted RBF activation maps provides a full-brain image, in MNI152  
519 space, of the given principal component (Fig. S3).

520 Next, we considered 80 topics estimated using Latent Dirichlet Allocation (Blei et al., 2003)  
521 applied to 9,204 functional neuroimaging articles in the Neurosynth database (Rubin et al., 2017).  
522 The topics, as well as associated brain maps identified using Neurosynth, were identified and  
523 reported in several prior studies (Chen et al., 2020; Fox et al., 2014; Sul et al., 2017). The topic  
524 labels for each topic were generated automatically with the following ChatGPT (OpenAI, 2023)  
525 prompt: “Please help me come up with intuitive labels for topics I found by fitting a topic model  
526 to thousands of neuroscience and psychology articles. I’ll paste in the top 10 highest-weighted  
527 words for each topic, and I’d like you to respond with a suggested label. For each topic, please  
528 respond with just the topic label and no other formatting or text. Here are the next topic’s top  
529 words:” followed by a comma-separated list of the given topic’s top-weighted words reflected  
530 in the Table S1. For some topics, ChatGPT responded with a longer-form response rather than a  
531 concise topic label. In these instances, on a case-by-case basis, we used a second follow-up prompt  
532 to achieve the given topic’s label: “Could you please come up with a more concise label for that  
533 topic?”. We then manually identified a set of 11 cognitive labels that were intended to encapsulate  
534 a representative range of widely studied low-level and high-level cognitive functions. In choosing  
535 the set of cognitive labels, we jointly considered each topic’s ChatGPT-derived topic label, along  
536 with the top-weighted words for the topic. We attempted to generate a concise set of labels that  
537 still spanned the full set of cognitive functions reflected across the 80 topics. Topics that appeared  
538 unrelated to specific cognitive functions (e.g., topics related to specific methods or clinical themes)  
539 are designated with dashes in Table S1.

540 Finally, following an approach used in several prior studies (Chen et al., 2020; Fox et al., 2014;

541 Sul et al., 2017) we treated the correlation between a given component’s brain map and each topic’s  
542 brain map as an approximate measure of how much the component was reflective of the given  
543 topic. This resulted in a set of 80 “weights” (correlation coefficients) for each component’s brain  
544 map, with one weight per Neurosynth-derived topic.

545 **Ranking cognitive processes**

546 We manually identified 11 cognitive labels spanning the set of 80 Neurosynth-derived topics:  
547 cognitive control, language processing, memory, emotion, social cognition, spatial cognition, at-  
548 tention, reward, sensory perception, motor control, and resting state. We then used ChatGPT  
549 to automatically “rank” the processes from high-level to low-level using the following prompt:  
550 “Please rank these cognitive processes from highest-level to lowest-level, where higher values  
551 indicate higher-order or higher-level processes. Return the result as a csv file with a header row  
552 and two columns: ‘Cognitive label’ and ‘Rank’. Here are the processes: cognitive control, lan-  
553 guage processing, memory, emotion, social cognition, spatial cognition, attention, reward, sensory  
554 perception, motor control, resting state”. Table S2 displays the output.

555 We recognize that ChatGPT is not omniscient, nor should it be treated as an expert cognitive  
556 neuroscientist. We therefore reviewed ChatGPT’s responses carefully by hand to verify that they  
557 seemed reasonable to us. Whereas prior work has often constructed such rankings by hand, we see  
558 our use of ChatGPT in this case as a small additional “sanity check” on our rankings that helped  
559 us to be slightly more objective than if we had simply created the rankings ourselves manually.

560 In the analysis presented in Figure 6E, we summarize difference in topic weightings across  
561 experimental conditions. In particular, we sought to characterize how the dominant neural patterns  
562 evoked by each experimental condition weighted on different cognitive functions. For each of the  
563 top five principal components from each experimental condition (Fig. 5), we computed the average  
564 weightings for each of the 11 manually identified (and ChatGPT-ranked) cognitive labels described  
565 above (Tab. S2). We then fit a line separately for each experiment condition ( $x$ -values: cognitive  
566 rank;  $y$ -values: weights). In carrying out this analysis, we used a bootstrap procedure to estimate  
567 the variability in the slopes of the regression lines, whereby we repeated this process for each of  
568  $n = 100$  iterations, each time resampling (with replacement) the set of observed ranks and weights.

569 This procedure yielded distributions of 100 estimated slopes for each experimental condition. We  
570 used these distributions to compare the slopes across experimental conditions and to estimate 95%  
571 confidence intervals.

572 **Synthetic data**

573 To help illustrate the relationship between informativeness and compressibility (Fig. 1), we gen-  
574 erated four synthetic datasets, varying in informativeness and compressibility. Each dataset com-  
575 prised simulated observations of  $K = 25$  features across  $N = 100$  timepoints, from each of  $S = 10$   
576 participants. To create each dataset, we first constructed a “template” matrix of  $N$  timepoints by  $K$   
577 features. We then generated participant-specific data by adding independent noise to each entry  
578 in template matrix, drawn from the unit normal distribution (i.e., with a mean of 0 and a variance  
579 of 1). We repeated this process for each participant, yielding  $S$  participant-specific matrices for  
580 each dataset.

581 Since we estimate informativeness using the temporal decoding accuracy across participants,  
582 highly informative data will tend to have observations that are highly timepoint specific. Relatively  
583 uninformative data, in contrast, will tend to have more similar observations across timepoints. To  
584 generate data with “high informativeness,” we constructed template matrices whose rows (ob-  
585 servations) were drawn independently from zero-mean multivariate normal distributions. The  
586 covariances of these distributions were determined according to the desired compressibility of  
587 the data, as described below. We used a multi-step process to generate data with “low informa-  
588 tiveness.” First we generated new template matrices using the same procedure as for the “high  
589 informativeness” datasets. We then multiplied each matrix by a constant ( $\rho = 0.1$ ) and computed  
590 the cumulative sum of each matrix’s rows. This yielded matrices whose rows were highly similar  
591 across observations.

592 Compressibility reflects the extent to which decoding accuracy is affected by reducing the num-  
593 ber of components used to represent the data. Highly compressible data will tend to exhibit more  
594 similarities across features, whereas less compressible data will tend to show greater independence  
595 across features. To generate data with “high compressibility,” we set the covariance matrix of the  
596 multivariate normal distribution to a toeplitz matrix whose first row was given by  $[K, K - 1, \dots, 1]$ .

597 To generate data with “low compressibility,” we set the covariance matrix to the identity matrix.  
598 Template matrices for datasets with high informativeness and high compressibility, high in-  
599 formativeness and low compressibility, low informativeness and high compressibility, and low  
600 informativeness and low compressibility are displayed in Figure 1C. The corresponding decoding  
601 curves are displayed in Figure 1D.

## 602 **Data and code availability**

603 All of the code used to produce the figures and results in this manuscript, along with links to the  
604 corresponding data, may be found at [github.com/ContextLab/pca\\_paper](https://github.com/ContextLab/pca_paper).

## 605 **Acknowledgements**

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

## 611 **Author contributions**

612 Conceptualization: J.R.M. and L.L.W.O. Methodology: J.R.M. and L.L.W.O. Implementation:  
613 J.R.M. and L.L.W.O. Analysis: J.R.M. and L.L.W.O. Writing, Reviewing, and Editing: J.R.M. and  
614 L.L.W.O. Funding acquisition: J.R.M. Supervision: J.R.M.

## 615 **References**

616 Adachi, Y., Osada, T., Sporns, O., Watanabe, T., Matsui, T., Miyamoto, K., & Miyashita, Y.  
617 (2012). Functional connectivity between anatomically unconnected areas is shaped by collective  
618 network-level effects in the macaque cortex. *Cerebral Cortex*, 22(7), 1586–1592.

- 619 Alvarez, S. A. (2002). *An exact analytical relation among recall, precision, and classification accuracy in*  
620 *information retrieval* (Technical Report No. BCCS-02-01). Boston College.
- 621 Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering  
622 event structure in continuous narrative perception and memory. *Neuron*, 95(3), 709–721.
- 623 Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, 20(3), 353–364.
- 624 Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: a practical and  
625 powerful approach to multiple testing. *Journal of Royal Statistical Society, Series B*, 57, 289–300.
- 626 Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning*  
627 *Research*, 3, 993–1022.
- 628 Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 443–446.
- 629 Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., & Bressler, S. (2004). Beta oscillations in  
630 a large-scale sensorimotor cortical network: directional influences revealed by Granger causality.  
631 *Proceedings of the National Academy of Sciences, USA*, 101(26), 9849–9854.
- 632 Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural  
633 and functional systems. *Nature Reviews Neuroscience*, 10(3), 186–198.
- 634 Capota, M., Turek, J., Chen, P.-H., Zhu, X., Manning, J. R., Sundaram, N., ... Shin, Y. S. (2017).  
635 *Brain imaging analysis kit*.
- 636 Chen, P.-H. A., Jolly, E., Cheong, J. H., & Chang, L. J. (2020). Intersubject representational  
637 similarity analysis reveals individual variations in affective experience when watching erotic  
638 movies. *NeuroImage*, 216(116851), doi.org/10.1016/j.neuroimage.2020.116851.
- 639 Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. (2014). Intrinsic and  
640 task-evoked network architectures of the human brain. *Neuron*, 83(1), 238–251.
- 641 Dhamala, M., Rangarajan, G., & Ding, M. (2008). Analyzing information flow in brain networks  
642 with nonparametric Granger causality. *NeuroImage*, 41(2), 354–362.

- 643 Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., & Constable, R. T. (2017).  
644 Can brain state be manipulated to emphasize individual differences in functional connectivity.  
645 *NeuroImage*, 160, 140–151.
- 646 Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., ... Constable, R. T.  
647 (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain  
648 connectivity. *Nature Neuroscience*, 18, 1664–1671.
- 649 Fox, A. S., Chang, L. J., Gorgolewski, K. J., & Yarkoni, T. (2014). Bridging psychology and  
650 genetics using large-scale spatial analysis of neuroimaging and neurogenetic data. *bioRxiv*,  
651 doi.org/10.1101/012310.
- 652 Glerean, E., Salmi, J., Lahnakoski, J. M., Jääskeläinen, I. P., & Sams, M. (2012). Functional magnetic  
653 resonance imaging phase synchronization as a measure of dynamic functional connectivity.  
654 *Brain Connectivity*, 2(2), 91–101.
- 655 Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., & Petersen, S. E. (2016).  
656 Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral  
657 Cortex*, 26, 288–303.
- 658 Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., ...  
659 Petersen, S. E. (2018). Functional brain networks are dominated by stable group and individual  
660 factors, not cognitive or daily variation. *Neuron*, 98(2), 439–452.
- 661 Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal  
662 receptive windows in human cortex. *The Journal of Neuroscience*, 28(10), 2539–2550.
- 663 Korzeniewska, A., Crainiceanu, C. M., Kus, R., Franaszczuk, P. J., & Crone, N. E. (2008). Dynamics  
664 of event-related causality in brain electrical activity. *Human Brain Mapping*, 29(10).
- 665 Kumar, M., Anderson, M. J., Antony, J. W., Baldassano, C., Brooks, P. P., Cai, M. B., ... Norman,  
666 K. A. (2021). BrainIAK: the brain image analysis kit. *Aperture*, doi.org/10.52294/31bb5b68-2184-  
667 411b-8c00-a1dacb61e1da.

- 668 Laird, A. R., Fox, P. M., Eickhoff, S. B., Turner, J. A., Ray, K. L., McKay, D. R., ... Fox, P. T. (2011).  
669 Behavioral interpretations of intrinsic connectivity networks. *Journal of Cognitive Neuroscience*,  
670 23(12), 4022–4037.
- 671 Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of  
672 temporal receptive windows using a narrated story. *The Journal of Neuroscience*, 31(8), 2906–2915.
- 673 Lynn, C. W., & Bassett, D. S. (2021). Quantifying the compressibility of complex networks.  
674 *Proceedings of the National Academy of Sciences, USA*, 118(32), e2023473118.
- 675 Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during  
676 concept learning. *Nature Communications*, 11(46), doi.org/10.1038/s41467-019-13930-8.
- 677 Manning, J. R. (2020). Context reinstatement. In M. J. Kahana & A. D. Wagner (Eds.), *Handbook of  
678 human memory*. Oxford University Press.
- 679 Manning, J. R., Zhu, X., Willke, T. L., Ranganath, R., Stachenfeld, K., Hasson, U., ... Norman,  
680 K. A. (2018). A probabilistic approach to discovering dynamic full-brain functional connectivity  
681 patterns. *NeuroImage*, 180, 243–252.
- 682 Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel  
683 pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9), 424–430.
- 684 OpenAI. (2023). ChatGPT. <https://chat.openai.com>.
- 685 Owen, L. L. W., Chang, T. H., & Manning, J. R. (2021). High-level cognition during story listening is  
686 reflected in high-order dynamic correlations in neural activity patterns. *Nature Communications*,  
687 12(5728), doi.org/10.1038/s41467-021-25876-x.
- 688 Owen, L. L. W., Muntianu, T. A., Heusser, A. C., Daly, P., Scangos, K. W., & Manning, J. R. (2020). A  
689 Gaussian process model of human electrocorticographic data. *Cerebral Cortex*, 30(10), 5333–5345.
- 690 Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers  
691 into movies. *Spatial Vision*, 10, 437–442.
- 692 Preti, M. G., Bolton, T. A. W., & Van De Ville, D. (2017). The dynamic functional connectome:  
693 state-of-the-art and perspectives. *NeuroImage*, 160, 41–54.

- 694 Regev, M., Simony, E., Lee, K., Tan, K. M., Chen, J., & Hasson, U. (2018). Propagation of information  
695 along the cortical hierarchy as a function of attention while reading and listening to stories.  
696 *Cerebral Cortex*.
- 697 Rogers, B. P., Morgan, V. L., Newton, A. T., & Gore, J. C. (2007). Assessing functional connectivity  
698 in the human brain by fMRI. *Magnetic Resonance Imaging*, 25(11), 1347–1357.
- 699 Rubin, T. N., Kyojo, O., Gorgolewski, K. J., Jones, M. N., Poldrack, R. A., & Yarkoni, T. (2017).  
700 Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human  
701 cognition. *PLoS Computational Biology*, 13(10), e1005649.
- 702 Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and  
703 interpretations. *NeuroImage*, 52, 1059–1069.
- 704 Scangos, K. W., Khambhati, A. N., Daly, P. M., Owen, L. L. W., Manning, J. R., Ambrose, J. B.,  
705 ... Chang, E. F. (2021). Biomarkers of depression symptoms defined by direct intracranial  
706 neurophysiology. *Frontiers in Human Neuroscience*, 15, doi.org/10.3389/fnhum.2021.746499.
- 707 Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Z, X.-N., Holmes, A. J., ... Yeo, B. T. T. (2018).  
708 Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI.  
709 *Cerebral Cortex*, 28, 3095–3114.
- 710 Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*,  
711 27(3), 379–423.
- 712 Simony, E., & Chang, C. (2020). Analysis of stimulus-induced brain dynamics during naturalistic  
713 paradigms. *NeuroImage*, 216, 116461.
- 714 Simony, E., Honey, C. J., Chen, J., & Hasson, U. (2016). Dynamic reconfiguration of the default  
715 mode network during narrative comprehension. *Nature Communications*, 7(12141), 1–13.
- 716 Sizemore, A. E., Giusti, C., Kahn, A., Vettel, J. M., Betzel, R. F., & Bassett, D. S. (2018). Cliques and  
717 cavities in the human connectome. *Journal of Computational Neuroscience*, 44(1), 115–145.

- 718 Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., ...  
719 Glasser, M. F. (2013). Resting-state fMRI in the Human Connectome Project. *NeuroImage*, 80,  
720 144–168.
- 721 Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., ... Beckmann,  
722 C. F. (2009). Correspondence of the brain's functional architecture during activation and rest.  
723 *Proceedings of the National Academy of Sciences, USA*, 106(31), 13040–13045.
- 724 Smith, S. M., Hyvärinen, A., Varoquaux, G., Miller, K. L., & Beckmann, C. F. (2014). Group-PCA  
725 for very large fMRI datasets. *NeuroImage*, 101, 738–749.
- 726 Smith, S. M., Vidaurre, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M., Miller, K. L., ... Van  
727 Essen, D. C. (2013). Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences*,  
728 17(12), 666–682.
- 729 Sporns, O., & Betzel, R. F. (2016). Modular brain networks. *Annual Review of Psychology*, 67,  
730 613–640.
- 731 Sporns, O., & Honey, C. J. (2006). Small worlds inside big brains. *Proceedings of the National Academy  
of Sciences, USA*, 103(51), 19219–19220.
- 733 Sporns, O., & Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics*, 2(2),  
734 145–162.
- 735 Srinivasan, R., Winter, W. R., Ding, J., & Nunez, P. L. (2007). EEG and MEG coherence: Measures of  
736 functional connectivity at distinct spatial scales of neocortical dynamics. *Journal of Neuroscience  
Methods*, 166, 41–52.
- 738 Sul, S., Güroğlu, B., Crone, E. A., & Chang, L. J. (2017). Medial prefrontal cortical thinning  
739 mediates shifts in other-regarding preferences during adolescence. *Scientific Reports*, 7(8510),  
740 doi.org/10.1038/s41598-017-08692-6.
- 741 Tomasi, D., & Volkow, N. D. (2011). Association between functional connectivity hubs and brain  
742 networks. *Cerebral Cortex*, 21, 2003–2013.

- <sup>743</sup> Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., . . . Buckner,  
<sup>744</sup> R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional  
<sup>745</sup> connectivity. *Journal of Neurophysiology*, 106(3), 1125–1165.