

1 High-order cognition is supported by information-rich but
2 compressible brain activity patterns

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

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

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

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

March 15, 2023

Abstract

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

²⁴ Introduction

25 Large-scale networks, including the human brain, may be conceptualized as occupying one or
26 more positions along on a continuum. At one extreme, every node is fully independent from
27 every other node. At the other extreme, all nodes behave identically. Each extreme optimizes
28 key properties of how the network functions. When every node is independent, the network is

29 maximally *expressive*: if we define the network’s “state” as the activity pattern across its nodes, then
30 every state is equally reachable by a network with fully independent nodes. On the other hand, a
31 network of identically behaved nodes optimizes *robustness*: any subset of nodes may be removed
32 from the network without any loss of function or expressive power, as long as any single node
33 remains. Presumably, most natural systems tend to occupy positions between these extremes. We
34 wondered: might the human brain reconfigure itself to be more flexible or more robust according
35 to ongoing demands? In other words, might the brain reconfigure its connections or behaviors
36 under different circumstances to change its position along this continuum?

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

55 We take as a given that brain activity is highly flexible: our brains can exhibit nearly infinite
56 varieties of activity patterns. This flexibility implies that our brains activity patterns are highly in-
57 formation rich. However, brain activity patterns are also highly structured. For example, full-brain
58 correlation matrices are stable within (???) and across (????) individuals. This stability suggests

59 that our brains' activity patterns are at least partially constrained, for example by anatomical, external, or internal factors. Constraints on brain activity that limit its flexibility decrease expressiveness
60 (i.e., its information rate). However, constraints on brain activity also increase its robustness to
61 noise (e.g., "missing" or corrupted signals may be partially recovered). For example, recent work
62 has shown that full-brain activity patterns may be reliably recovered from only a relatively small
63 number of implanted electrodes (??). This robustness property suggests that the relevant signal
64 (e.g., underlying factors that have some influence over brain activity patterns) are compressible.
65

66 To the extent that brain activity patterns contain rich task-relevant information, we should
67 be able to use the activity patterns to accurately differentiate between different aspects of the
68 task (e.g., using pattern classifiers; ?). For example, prior work has shown a direct correspondence
69 between classification accuracy and the information content of a signal (?). To the extent that
70 brain activity patterns are compressible, we should be able to generate simplified (e.g., lower
71 dimensional) representations of the data while still preserving the relevant or important aspects of
72 the original signal. In general, information content and compressibility are related but are partially
73 dissociable (Fig. ??). If a given signal (e.g., a representation of brain activity patterns) contains
74 more information about ongoing cognitive processes, then the peak decoding accuracy should
75 be high. And if the signal is compressible, a low-dimensional embedding of the signal will be
76 similarly informative to the original signal (Fig. ??D).

77 Several recent studies suggest that the complexity of brain activity is task-dependent, whereby
78 simpler tasks with lower cognitive demands are reflected by simpler and more compressible
79 brain activity patterns, and more complex tasks with higher cognitive demands are reflected by
80 more complex and less compressible brain activity patterns (??). These findings complement
81 other work suggesting that functional connectivity (correlation) patterns are task-dependent (??),
82 although see ?. Higher-order cognitive processing of a common stimulus also appears to drive
83 more stereotyped task-related activity and functional connectivity across individuals (????).

84 The above studies are consistent with two potential descriptions of how cognitive processes are
85 reflected in brain activity patterns. One possibility is that the information rate of brain activity in-
86 creases during more complex or higher-level cognitive processing. If so, then the ability to reliably
87 decode cognitive states from brain activity patterns should improve with task complexity or with
88 the level (or "depth") of cognitive processing. A second possibility is that the compressibility of

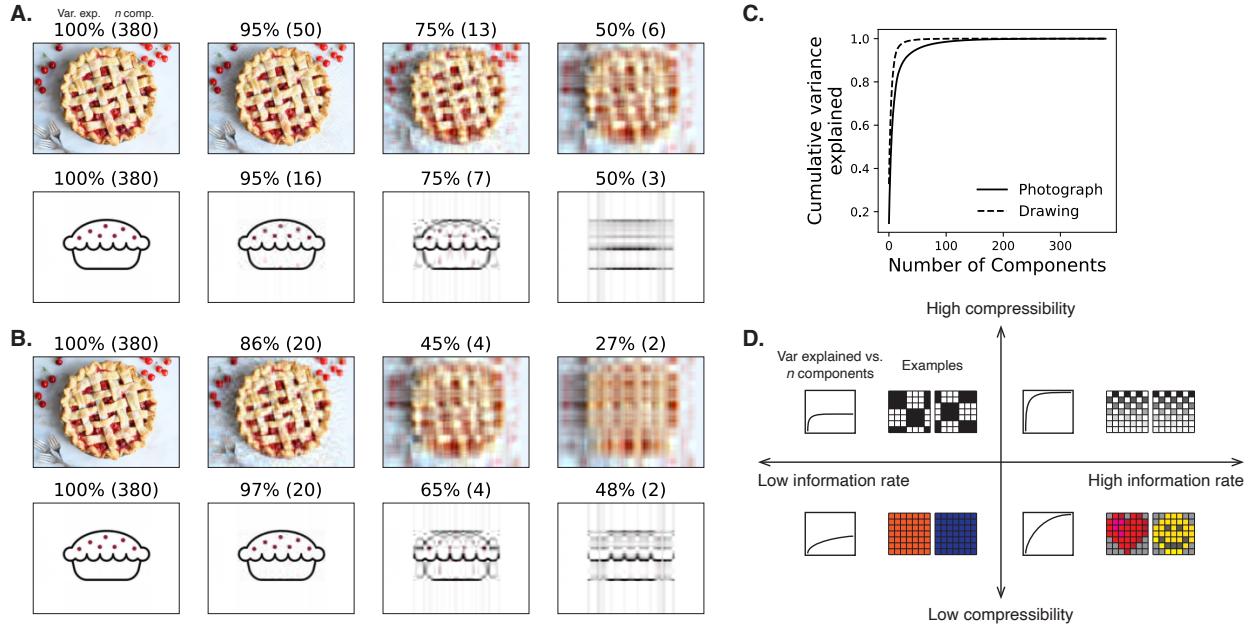


Figure 1: Information content and compressibility. **A. Variance explained for two images.** We applied principal components analysis to a photograph and drawing, treating the rows of the images as “observations.” Across columns, we identified the numbers of components required to explain 100%, 95%, 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. **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.

89 brain activity patterns increases during more complex or higher-level cognitive processing. If so,
90 then individual features of brain recordings, or compressed representations of brain recordings,
91 should carry more information during complex or high-level (versus simple or low-level) cognitive
92 tasks.

93 We used a previously collected neuroimaging dataset to estimate the extent to which each
94 of these two possibilities might hold. The dataset we examined comprised functional magnetic
95 resonance imaging (fMRI) data collected as participants listened to an audio recording of a 10-
96 minute story, temporally scrambled recordings of the story, or underwent a resting state scan (?).
97 Each of these experimental conditions evokes different depths of cognitive processing (????). We
98 used across-participant classifiers to decode listening times in each condition, as a proxy for how
99 "informative" the task-specific activity patterns were (?). We also used principle components
100 analysis to generate lower-dimensional representations of the activity patterns. We then repeated
101 the classification analyses after preserving different numbers of components and examined how
102 classification accuracy changed across the different experimental conditions.

103 Results

104 We sought to understand whether higher-level cognition is reflected by more reliable and in-
105 formative brain activity patterns, and how compressibility of brain activity patterns relates to
106 cognitive complexity. We developed a computational framework for systematically assessing the
107 informativeness and compressibility of brain activity patterns recorded under different cognitive
108 circumstances. We used across-participant decoding accuracy (see *Forward inference and decoding*
109 *accuracy*) as a proxy for informativeness. To estimate the compressibility of the brain patterns, we
110 used group principal components analysis (PCA) to project the brain patterns into k -dimensional
111 spaces, for different values of k (see *Hierarchical Topographic Factor Analysis (HTFA)* and *Principal*
112 *components analysis (PCA)*). For more compressible brain patterns, decoding accuracy should be
113 more robust to small values of k .

114 We analyzed a dataset collected by ? that comprised four experimental conditions. These
115 conditions exposed participants to stimuli that systematically varied in cognitive engagement.
116 In the *intact* experimental condition, participants listened to an audio recording of a 10-minute

117 story. In the *paragraph*-scrambled experimental condition, participants listened to a temporally
118 scrambled version of the story, where the paragraphs occurred out of order, but where the same set
119 of paragraphs was presented over the entire listening interval. All participants in this condition
120 experienced the scrambled paragraphs in the same order. In the *word*-scrambled experimental
121 condition, participants listened to a temporally scrambled version of the story, where the words
122 occurred in a random order. Again, all participants in this condition experienced the scrambled
123 words in the same order. Finally, in the *rest* experimental condition, participants lay in the scanner
124 with no overt stimulus, while keeping their eyes open and blinking as needed. This public dataset
125 provided a convenient means for testing our hypothesis that different levels of cognitive processing
126 and engagement affect how informative and compressible the associated brain patterns are.

127 To evaluate the relation between informativeness and compressibility for brain activity from
128 each experimental condition, we trained a series of across-participant temporal decoders on com-
129 pressed representations of the data. Figure ??A displays the decoding accuracy as a function
130 of the number of principal components used to represent the data (also see Fig. S1). Several
131 patterns were revealed by the analysis. First, in general (i.e., across experimental conditions),
132 decoding accuracy tends to improve as the number of components are increased. However, de-
133 coding accuracy peaked at higher levels for experimental conditions that exposed participants
134 to cognitively richer stimuli (Fig. ??D). The peak decoding accuracy was highest for the “intact”
135 condition (versus paragraph: $t(99) = 35.205, p < 0.001$; versus word: $t(99) = 43.172, p < 0.001$);
136 versus rest: $t(99) = 81.361, p < 0.001$), next highest for the “paragraph” condition (versus word:
137 $t(99) = 6.243, p < 0.001$; versus rest: $t(99) = 50.748, p < 0.001$), and next highest for the “word”
138 condition (versus rest: $t(99) = 48.791, p < 0.001$). This ordering implies that cognitively richer
139 conditions evoke more stable brain activity patterns across people.

140 The cognitively richer conditions also displayed steeper initial slopes. For example, the intact
141 condition decoders reached an arbitrarily chosen threshold of 5% accuracy using fewer components
142 than the paragraph condition decoders ($t(99) = -7.429, p < 0.001$) or word condition decoders
143 ($t(99) = -7.300, p < 0.001$), and decoding accuracy never exceeded 5% for the rest condition. This
144 suggests that brain activity patterns evoked by cognitively richer conditions are more compressible,
145 such that representing the data using the same number of principal components provides more
146 information to the temporal decoders (Figs. ??B, C). Taken together, as shown in Figure ??E, we

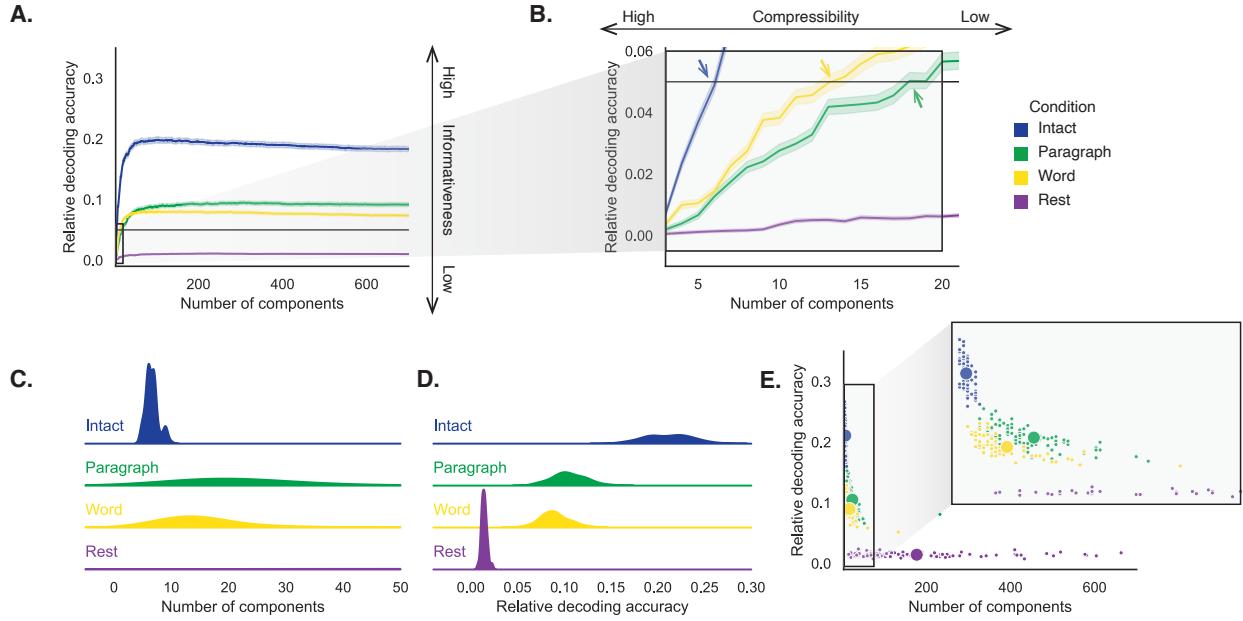


Figure 2: Decoding accuracy and compression. A. Decoding accuracy by number of components. Ribbons of each color display cross-validated decoding performance for each condition (intact, paragraph, word, and rest), as a function of the number of components (features) used to represent the data. For each 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 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 illustrated in Panel E.) **D. Estimating “informativeness” for each condition.** The density plots display the peak (corrected) decoding accuracies achieved in each condition, across all cross validation runs. **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.

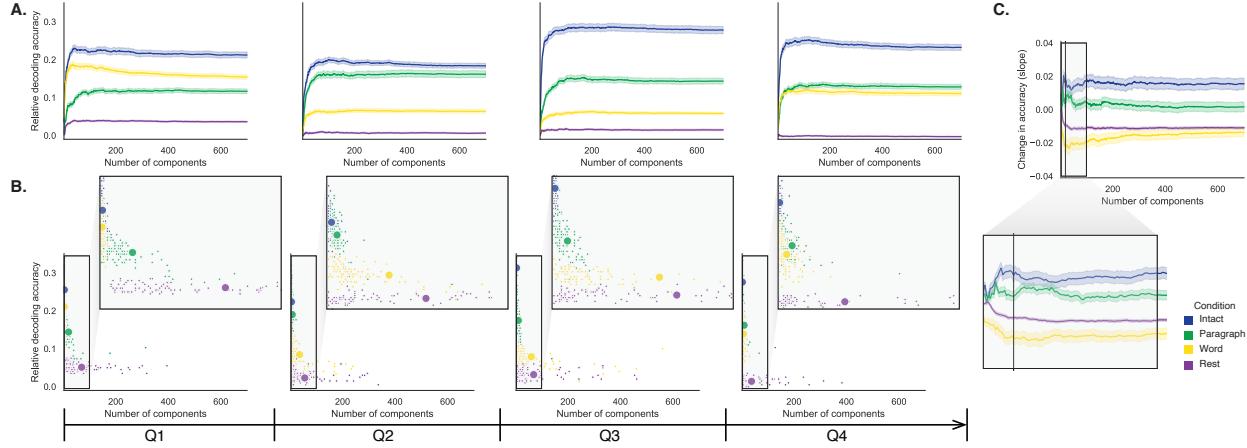


Figure 3: Changes in decoding accuracy and compression over time. **A. Decoding accuracy by number of components, by story segment.** Each family of curves is plotted in the same format as Figure ??A but reflects data only from one quarter of the dataset. **B. Informativeness versus compressibility by condition and segment.** Each scatter plot is in the same format as Figure ??E, 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. All error ribbons denote bootstrap-estimated 95% confidence intervals.

147 found that brain activity patterns evoked by cognitively richer conditions tended to be both more
 148 informative (i.e., associated with higher decoding accuracies) *and* more compressible (i.e., requiring
 149 fewer components to achieve the 5% accuracy threshold).

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

162 the decoding accuracy and compressibility *decreased* with listening time in the word condition
163 ($t(99) = -10.747, p < 0.001$) and rest condition ($t(99) = -22.081, p < 0.001$). This might reflect the
164 depletion of attentional resources in the less-engaging word and rest conditions.

165 We also wondered how informativeness and compressibility in the different experimental
166 conditions might vary across brain networks. We used a network parcellation identified by
167 ? to segment the brain into seven distinct networks. The networks can be sorted (roughly)
168 in order from lower-level to higher-level cortex as follows (Figs. ??A–C): visual, somatomotor,
169 dorsal attention, ventral attention, limbic, frontoparietal, and default mode. Next, we computed
170 decoding curves separately for the activity patterns within each network and identified each
171 network's inflection point, for each experimental condition. Moving from low-order networks
172 to higher-order networks, we found that decoding accuracy tended to increase, particularly in
173 the higer-order experimental conditions (Fig. ??D, E). This suggests that higher-order networks
174 may carry more content-relevant or stimulus-driven "information." We found no clear trends
175 in the proportions of components required to achieve 5% decoding accuracy across networks or
176 conditions (Fig. ??F).

177 In addition to examining different networks in isolation, we wondered about the general
178 structure of the full-brain (i.e., potentially multi-network) activity patterns reflected by different
179 principal components across different experimental conditions. As shown in Figure ??, we used
180 Neurosynth (?) to identify, for each component, the associations with each of 80 themes (see
181 *Reverse inference*). In general, the first principal components across all of the experimental con-
182 ditions tended to weight most heavily on themes related to cognitive control, memory, language
183 processing, attention, and perception. Other components appeared to vary more across conditions.

184 To gain further insights into which brain functions might be most closely associated with
185 the top-weighted components from each experimental condition, we manually grouped each
186 Neurosynth-derived topic into a smaller set of core cognitive functions. Separately for each com-
187 ponent, we computed the average weightings across all topics that were tagged as being associated
188 with each of these cognitive functions (Figs. ??A, S5A). To help visualize these associations, we
189 used the patterns of associations for each component to construct graphs whose nodes were ex-
190 perimental conditions and cognitive functions (Figs. ??B, S5B). We also computed correlations
191 between the sets of per-topic weightings from each of the top-weighted components from each

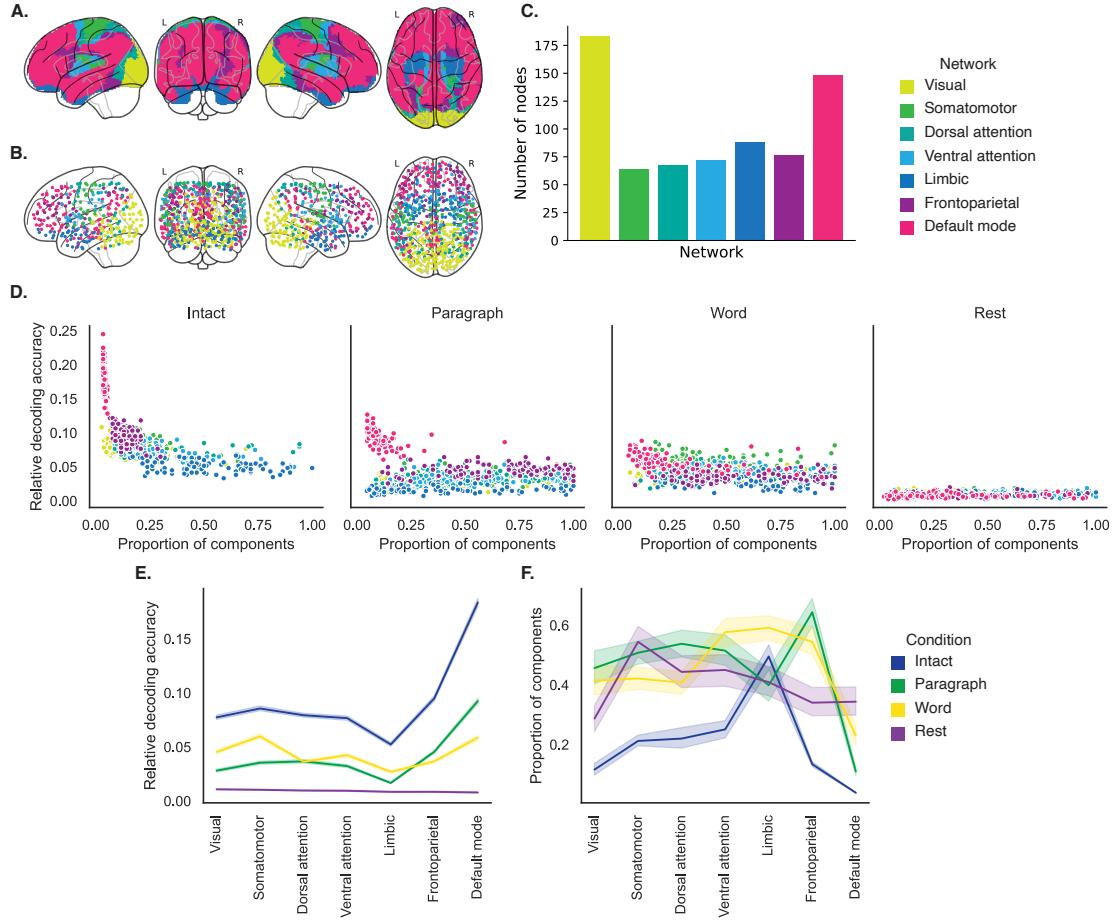


Figure 4: Network-specific decoding accuracy and compression. **A. Network parcellation map.** The glass brain displays the location of each of seven networks identified by ?. **B. Node labels.** We assigned network labels to each of 700 Hierarchical Topographic Factor Analysis-derived nodes (see *Hierarchical 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 95% bootstrap-estimated confidence intervals.

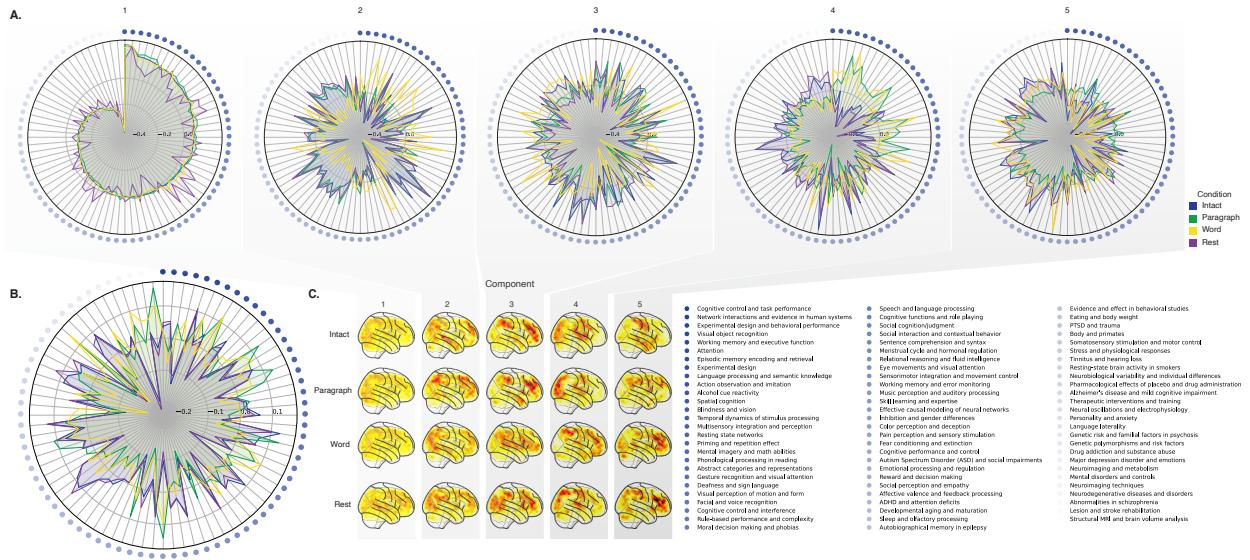


Figure 5: Neurosynth topic weightings by component. We used a reverse inference procedure (see *Reverse inference*) to estimate the correspondences between brain images and a set of 80 topics derived from the Neurosynth database of neuroimaging articles (?). **A. Topic correlations by component.** For 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 conditions (rows). Additional projections for each component may be found in Figure S3.

192 experimental condition (Fig. ??C, D) and between the brain maps for each condition's components
193 (Fig. S5C, D). Taken together, we found that each component appeared to weight on a fundamental
194 set of cognitive functions that varied by experimental condition. For example, the top principal
195 components from every condition weighted similarly on the full set of Neurosynth topics (Fig. ??A)
196 and cognitive functions (Figs. ??A, B and S5 A, B), suggesting that these components might re-
197 flect a set of functions or activity patterns that are common across all conditions. The second
198 components' weightings were similar across the intact, paragraph, and rest conditions (highest-
199 weighted functions: cognitive control, memory, social cognition, and resting state), but different
200 for the word condition (highest-weighted functions: sensory perception and cognitive control).
201 The fourth components' weighting were grouped the paragraph and word conditions (highest-
202 weighted functions: memory, language processing, and cognitive control) and the intact and rest
203 conditions (highest-weighted functions: emotion, social cognition). We also used ChatGPT (?) to
204 sort the list of manually tagged cognitive functions from lowest-level to highest-level (Tab. S2,
205 Fig. ??E; also see *Ranking cognititive processes*). We found that higher-level functions tended to
206 be weighted on more heavily by top components from the intact and paragraph conditions than
207 lower-level functions. The top components from the word condition showed the opposite ten-
208 dency, whereby *lower*-level functions tended to be weighted on more heavily than higher-level
209 functions. The components from the rest condition showed almost no differences in the weights
210 associated with high-level versus low-level functions.

211 Discussion

212 We examined fMRI data collected as participants listened to an auditory recording of a story,
213 scrambled recordings of the story, or underwent a resting state scan. We found that cognitively
214 richer stimuli evoked more reliable (i.e., consistent across people) and information rich brain activ-
215 ity patterns. The brain patterns evoked by cognitively richer stimuli were also more compressible,
216 in that each individual component provided more "signal" to temporal decoders relative to com-
217 ponents of data from less cognitively rich conditions (Fig. ??). Over time (e.g., as the experiment
218 progressed), these phenomena were strengthened. Specifically, across story segments, data from
219 more cognitively rich conditions became more informative and compressible, and data from less

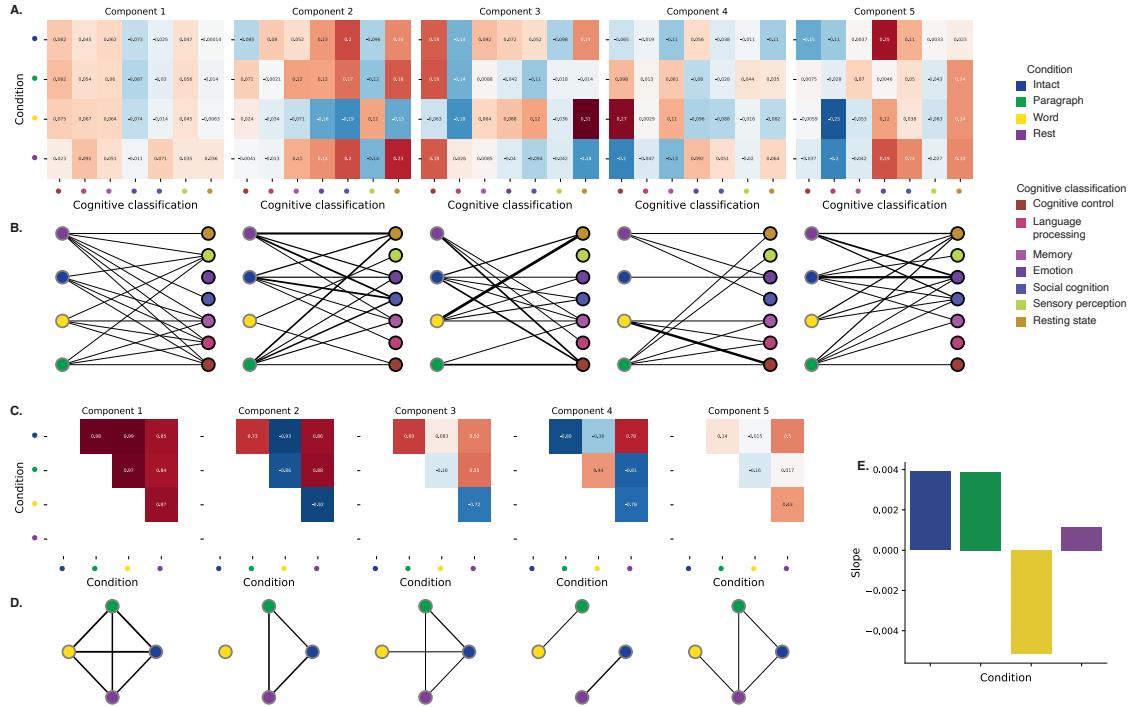


Figure 6: Summary of functions associated with top-weighted components by condition. A. Top-weighted topics by condition. Here we display per-condition (rows, indicated by colored dots) topic correlations, averaged across topics that pertain to each of several broad cognitive functions (columns within 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 heatmaps in Panel A). **C. Correlations between each principal component, by condition.** The heatmaps 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-panels network plot summarizes the pattern of correlations between the n^{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 heatmaps 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*). Also see Fig. S5 for additional information.

cognitively rich conditions became *less* informative and compressible (Fig. ??). We also repeated these analyses separately for different brain networks. We found that networks traditionally associated with higher-level cognitive functions tended to provide more informative brain patterns than networks traditionally associated with lower level cognitive functions (Fig. ??). Finally, we examined the most dominant components of the brain activity patterns from each experimental condition. We used a reverse inference approach (?) to identify the terms in the neuroimaging literature most commonly associated with the corresponding maps. As summarized in Figure ??, we found that terms associated with memory and sensory processing were associated with the strongest components in all three story listening conditions. Terms associated with sensory integration were associated with the strongest components in the intact and paragraph-scrambled conditions. Terms associated with sentence comprehension, emotion, and valence were associated with the strongest components in the intact condition. Finally, terms associated with the default mode network were associated with the strongest components in the word-scrambled and resting state conditions. Taken together, our findings indicate that the informativeness and compressibility of our brain activity patterns are task-dependent, and these properties change systematically with factors like cognitive richness and depth of processing.

Our explorations of informativeness and compressibility are related to a much broader literature on the correlational and causal structure of brain activity patterns (?????????????????????). Correlations or causal associations between different brain regions simultaneously imply that full-brain activity patterns will be compressible and also that those activity patterns will contain redundancies. For example, the extent of which activity patterns at one brain area can be inferred or predicted from activity patterns at other areas (e.g., ??), reflects overlap in the information available in or represented by those brain areas. If brain patterns in one area are recoverable using brain patterns in another area, then the “signal” used to convey the activity patterns could be compressed by removing the recoverable activity. Predictable (and therefore redundant) brain activity patterns are also more robust to signal corruption. For example, even if the activity patterns at one region are unreadable or unreliable at a given moment, that unreliability could be compensated for by other regions’ activity patterns that were predictive of the unreliable region.

Our findings that informativeness and compressibility change with task demands may follow from task-dependent changes in full-brain correlation patterns. A number of prior studies have

250 found that so-called “functional connectivity” (i.e., correlational) patterns vary reliably across tasks,
251 events, and situations (????). By examining how these task-dependent changes in correlations affect
252 informativeness and compressibility, our work suggests a potential reason why the statistical
253 structure of brain activity patterns might vary with cognitive task or with cognitive demands. For
254 lower-level tasks, or for tasks that require relatively little “deep” cognitive processing, our brains
255 may optimize activity patterns for robustness and redundancy over expressiveness, for example to
256 maximize reliability. For higher-level tasks, or for tasks that require deeper cognitive processing,
257 our brains may sacrifice some redundancy in favor of greater expressiveness.

258 In the information theory sense (?), when a signal is transmitted using a fixed alphabet of “sym-
259 bols,” the information rate decreases as the signal is compressed (e.g., fewer symbols transmitted
260 per unit time, using an alphabet with fewer symbols, etc.). Our finding that each individual brain
261 component (symbol) becomes more informative as cognitive richness increases suggests that the
262 “alphabet” of brain activity patterns is also task-dependent. Other work suggests that the rep-
263 resentations that are *reflected* by brain activity patterns may also change with task demands. For
264 example, our brains may represent the same perceptual stimulus differently depending on which
265 aspects of the stimulus or which combinations of features are task-relevant (?).

266 Different brain networks also varied in how informative and compressible their activity pat-
267 terns were across experimental conditions (e.g., Fig. ??). This might follow from evolutionary
268 optimizations that reflect the relevant constraints or demands placed on those networks. One
269 possibility is that cortex is organized in a hierarchy of networks “concerned with” or selective to
270 different levels of processing or function. To the extent that different levels of processing (e.g.,
271 low-level sensory processing versus “deeper” higher-level processing) reflect different stimulus
272 timescales (e.g., ?), the network differences we observed might also relate to the timescales at which
273 each network is maximally sensitive (????).

274 **Concluding remarks**

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

278 brain activity patterns, informativeness and compressibility, that appear to change systematically
279 with task demands and across brain networks. Our work helps to clarify how the “neural code”
280 might be structured, and how the code might vary across tasks and brain areas.

281 **Methods**

282 We measured properties of recorded neuroimaging data under different task conditions that varied
283 systematically in cognitive engagement and depth of processing. We were especially interested
284 in how *informative* and *compressible* the activity patterns were under these different conditions
285 (Fig. ??).

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

287 We examined an fMRI dataset collected by ? that the authors have made publicly available at
288 arks.princeton.edu/ark:/88435/dsp015d86p269k. The dataset comprises neuroimaging data col-
289 lected as participants listened to an audio recording of a story (intact condition; 36 partici-
290 pants), listened to temporally scrambled recordings of the same story (17 participants in the
291 paragraph-scrambled condition listened to the paragraphs in a randomized order and 36 in the
292 word-scrambled condition listened to the words in a randomized order), or lay resting with their
293 eyes open in the scanner (rest condition; 36 participants). Full neuroimaging details may be found
294 in the original paper for which the data were collected (?). Procedures were approved by the
295 Princeton University Committee on Activities Involving Human Subjects, and by the Western
296 Institutional Review Board (Puyallup, WA). All subjects were native English speakers with normal
297 hearing and provided written informed consent.

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

299 Following our prior related work, we used HTFA (?) to derive a compact representation of the
300 neuroimaging data. In brief, this approach approximates the timeseries of voxel activations (44,415
301 voxels) using a much smaller number of radial basis function (RBF) nodes (in this case, 700 nodes,
302 as determined by an optimization procedure; ??). This provides a convenient representation for
303 examining full-brain activity patterns and network dynamics. All of the analyses we carried out

304 on the neuroimaging dataset were performed in this lower-dimensional space. In other words,
305 each participant's data matrix was a number-of-timepoints (T) by 700 matrix of HTFA-derived
306 factor weights (where the row and column labels were matched across participants). Code for
307 carrying out HTFA on fMRI data may be found as part of the BrainIAK toolbox (??), which may
308 be downloaded at brainiak.org.

309 **Principal components analysis (PCA)**

310 We applied group PCA (?) separately to the HTFA-derived representations of the data (i.e., factor
311 loadings) from each experimental condition. Specifically, for each condition, we considered the
312 set of all participants' T by 700 factor weight matrices. We used group PCA to project these
313 700-dimensional matrices into a series of shared k -dimensional spaces, for $k \in \{3, 4, 5, \dots, 700\}$. This
314 yielded a set of number-of-participants matrices, each with T rows and k columns.

315 **Temporal decoding**

316 We sought to identify neural patterns that reflected participants' ongoing cognitive processing of
317 incoming stimulus information. As reviewed by ?, one way of homing in on these stimulus-driven
318 neural patterns is to compare activity patterns across individuals. In particular, neural patterns
319 will be similar across individuals to the extent that the neural patterns under consideration are
320 stimulus-driven, and to the extent that the corresponding cognitive representations are reflected
321 in similar spatial patterns across people (?). Following this logic, we used an across-participant
322 temporal decoding test developed by ? to assess the degree to which different neural patterns
323 reflected ongoing stimulus-driven cognitive processing across people. The approach entails using
324 a subset of the data to train a classifier to decode stimulus timepoints (i.e., moments in the story
325 participants listened to) from neural patterns. We use decoding (forward inference) accuracy on
326 held-out data, from held-out participants, as a proxy for the extent to which the inputted neural
327 patterns reflected stimulus-driven cognitive processing in a similar way across individuals.

328 **Forward inference and decoding accuracy**

329 We used an across-participant correlation-based classifier to decode which stimulus timepoint
330 matched each timepoint's neural pattern. For a given value of k (i.e., number of principal compo-
331 nents), we first used group PCA to project the data from each condition into a shared k -dimensional
332 space. Next, we divided the participants into two groups: a template group, $\mathcal{G}_{\text{template}}$ (i.e., training
333 data), and a to-be-decoded group, $\mathcal{G}_{\text{decode}}$ (i.e., test data). We averaged the projected data within
334 each group to obtain a single T by k matrix for each group. Next, we correlated the rows of the two
335 averaged matrices to form a T by T decoding matrix, Λ . In this way, the rows of Λ reflected time-
336 points from the template group, while the columns reflected timepoints from the to-be-decoded
337 group. We used Λ to assign temporal labels to each timepoint (row) from the test group's ma-
338 trix, using the row of the training group's matrix with which it was most highly correlated. We
339 repeated this decoding procedure, but using $\mathcal{G}_{\text{decode}}$ as the template group and $\mathcal{G}_{\text{template}}$ as the
340 to-be-decoded group. Given the true timepoint labels (for each group), we defined the decoding
341 accuracy as the average proportion of correctly decoded timepoints, across both groups (where
342 chance performance is $\frac{1}{T}$). In Figures ?? and ?? we report the decoding accuracy for each condition
343 and value of k , averaged across $n = 100$ cross validation folds.

344 **JRM NOTE: UPDATE THE NEXT TWO SUB-SECTIONS**

345 **Reverse inference**

346 To help interpret the brain activity patterns we found within the contexts of other studies, we
347 created summary maps of each principal component, for each experimental condition, by summing
348 together the 20 HTFA-derived RBF nodes (see *Hierarchical Topographic Factor Analysis*) with the
349 highest absolute value weights for each of the top 5 components (Figs. ??,). We then carried out
350 a meta analysis using Neurosynth (?) to identify the 5 terms most commonly associated with the
351 given map.

352 **Ranking cognitive processes**

353 **Data and code availability**

354 All of the code used to produce the figures and results in this manuscript, along with links to the
355 corresponding data, may be found at github.com/ContextLab/pca_paper.

356 **Acknowledgements**

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

362 **Author contributions**

363 Conceptualization: J.R.M. and L.L.W.O. Methodology: J.R.M. and L.L.W.O. Implementation: L.L.W.O.
364 Analysis: L.L.W.O. Writing, Reviewing, and Editing: J.R.M. and L.L.W.O. Supervision: J.R.M.