Article

Multi-view manifold learning of human brain-state trajectories

Erica L. Busch[1], Jessie Huang[2], Andrew Benz[3], Tom Wallenstein[2], Guillaume Lajoie[4,5], Guy Wolf[4,5], Smita Krishnaswamy [2,6,7,8,9] ✉ & Nicholas B. Turk-Browne[1,8,9]

The complexity of the human brain gives the illusion that brain activity is intrinsically high-dimensional. Nonlinear dimensionality-reduction methods such as uniform manifold approximation and t-distributed stochastic neighbor embedding have been used for high-throughput biomedical data. However, they have not been used extensively for brain activity data such as those from functional magnetic resonance imaging (fMRI), primarily due to their inability to maintain dynamic structure. Here we introduce a nonlinear manifold learning method for time-series data—including those from fMRI—called temporal potential of heat-diffusion for affinity-based transition embedding (T-PHATE). In addition to recovering a low-dimensional intrinsic manifold geometry from time-series data, T-PHATE exploits the data's autocorrelative structure to faithfully denoise and unveil dynamic trajectories. We empirically validate T-PHATE on three fMRI datasets, showing that it greatly improves data visualization, classification, and segmentation of the data relative to several other state-of-the-art dimensionality-reduction benchmarks. These improvements suggest many potential applications of T-PHATE to other high-dimensional datasets of temporally diffuse processes.

As we move through the world, the human brain performs innumerable operations to process, represent, and integrate information. How does a system composed of limited computational units represent so much while handling a constant barrage of incoming information? One possible explanation is that individual neural units operate in tandem to encode and update information in population codes, which exponentially expands the distinct ways in which information can be represented by a fixed number of units. Neural population codes better predict behavior than single neurons, particularly in how representations change over time in response to new information[1-5]. Although neural population codes exist in high ambient dimensions[6], these dimensions are redundant[7], and from them emerge dominant latent signals that code for behavior and information[8,9].

The principles of neural population codes were defined using direct neural recordings, and were extended to non-invasive, indirect measurements of brain activity using functional magnetic resonance imaging (fMRI). Multivariate pattern analysis[10,11] of fMRI activation has yielded valuable insights into the structure and content of cognitive representations, including low-[12] and high-level[11,13,14] sensory stimuli, as well as higher-order cognitive processes such as memory[15], emotion[16], narrative comprehension[17], and theory-of-mind[18]. These insights have largely come from group-level analyses, requiring aggregation over subjects and time points to overcome the

[1]Department of Psychology, Yale University, New Haven, CT, USA. [2]Department of Computer Science, Yale University, New Haven, CT, USA. [3]Department of Mathematics, Yale University, New Haven, CT, USA. [4]Department of Mathematics and Statistics, Université de Montréal, Montreal, Canada. [5]Mila—Quebec Artificial Intelligence Institute, Montreal, Canada. [6]Department of Genetics, Yale University, New Haven, CT, USA. [7]Program in Applied Mathematics, Yale University, New Haven, CT, USA. [8]Wu Tsai Institute, Yale University, New Haven, CT, USA. [9]These authors contributed equally: Smita Krishnaswamy and Nicholas B. Turk-Browne. ✉e-mail: smita.krishnaswamy@yale.edu
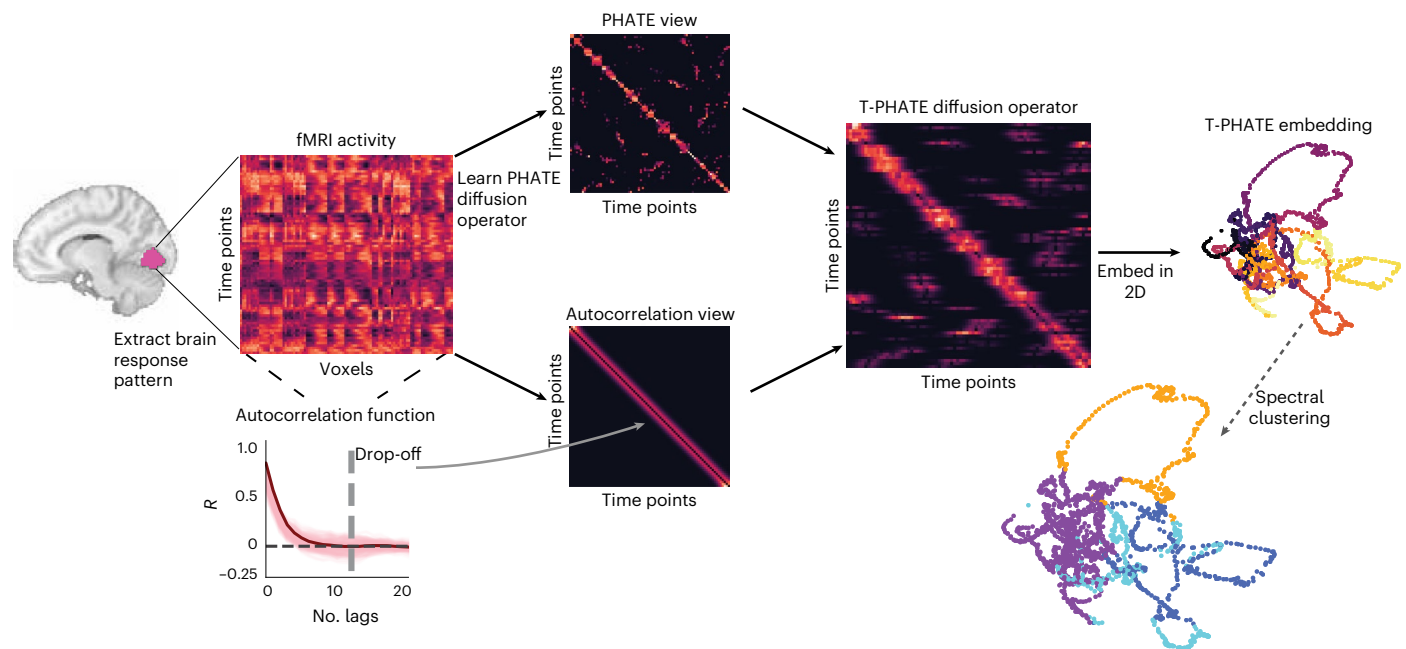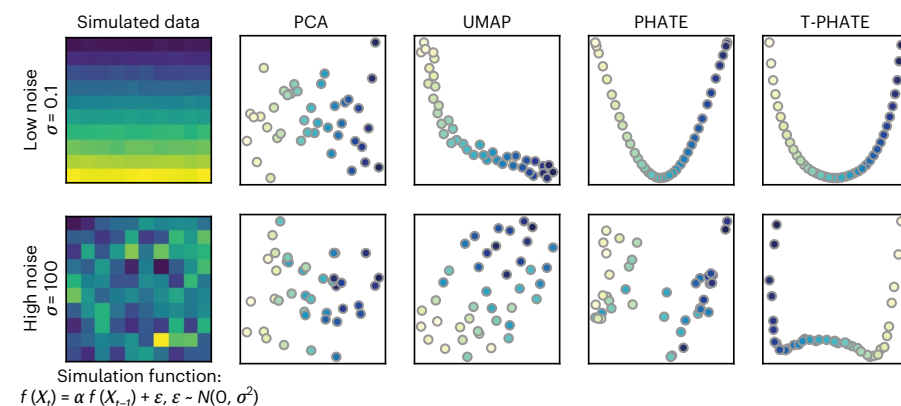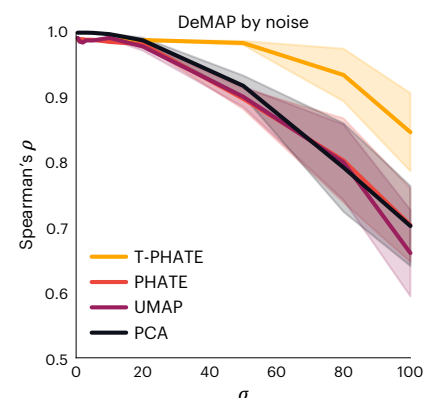
a T-PHATE procedure

b Embeddings of simulated autocorrelative data

Simulation function:
$f(X_t) = \alpha f(X_{t-1}) + \varepsilon, \; \varepsilon \sim N(0, \sigma^2)$

c DeMAP of simulated data embeddings

Fig. 1 | T-PHATE procedure. a, A multivariate pattern of fMRI activation over time (time points by voxels) is extracted from a region of interest (ROI). PHATE is then run to learn a PHATE-based affinity matrix between time points. The autocorrelation function is also estimated for the data separately to make an autocorrelation-based affinity matrix. These two views are combined into the T-PHATE diffusion operator and then embedded into lower dimensions. **b,c**, We tested this approach on simulated data with a known autocorrelative function. To do this, we took a matrix of pristine data (here, a simulated multivariate time-series in which $f(X_t) = \alpha f(X_{t-1})$) and added noise sampled from a normal distribution ($\varepsilon \sim \mathcal{N}(0, \sigma^2)$) to each sample of pristine data. We tested various values of σ from 0 to 100 (**c**, x-axis) and compared the fidelity of the manifold distances recovered by each embedding as a function of the amount of noise added to the input data using DeMAP[28]. We show two sample input matrices (**b**, far left) to the DeMAP analysis for noise sampled from a narrow distribution (top row, $\epsilon \sim \mathcal{N}(0, 0.1^2)$) and noise sampled from the widest distribution (bottom row; $\epsilon \sim \mathcal{N}(0, 100^2)$). T-PHATE both visually better reconstructs the autocorrelative signal from the noise (**b**) and quantitatively better preserves manifold distances (**c**) compared with PCA, UMAP, and PHATE. The error bands in **c** represent the 90% confidence interval of the mean DeMAP score from 1,000 bootstrap iterations. See Supplementary Fig. 1a for more benchmarks.

spatiotemporal noise inherent to fMRI. Like other high-throughput biomedical data, fMRI noise is pervasive at multiple levels, from subject movement to blood-oxygenation-level-dependent (BOLD) signal drift and physiological confounds[19]. fMRI has a lower temporal resolution than the cognitive processes many studies attempt to measure, with acquisitions every one to two seconds[20]. Finally, the BOLD signal is a slow, vascular proxy of neuronal activity, peaking approximately four to five seconds after stimulation before returning to the baseline.

In summary, many brain representations of interest are coded in high-dimensional patterns of activation[2–4,10,11], which can be characterized by low-dimensional latent signals[1,9,21]. fMRI affords unique insight into the healthy, behaving human brain, but the data are noisy, sampled slowly, and blurred in time, giving rise to high signal autocorrelation[20,22]. Furthermore, complex forms of cognition and learning unfurl and integrate over varying intrinsic timescales along processing hierarchies in the brain[23,24]. Addressing these issues requires consideration of how time and dimensionality interact with the measured BOLD signal, with the goal of interpreting fMRI activity at the single-subject level.

Here we introduce temporal potential of heat-diffusion for affinity-based transition embedding (T-PHATE) as a nonlinear manifold learning algorithm designed for high-dimensional, temporally dynamic signals. We apply T-PHATE to fMRI data measured

during cognitive tasks—a ripe testbed known for its high ambient dimensionality, multi-source noise, and temporal autocorrelation. Most previous studies exploring fMRI activity with dimensionality-reduction methods have relied on linear heuristics[25–27]. Nonlinear dimensionality-reduction algorithms have been used previously to characterize the geometry and underlying temporal dynamics of neural recordings[6,8,21] by integrating local similarities among data points into a global representation[9,28–31], but they remain underutilized in fMRI[21]. We build on PHATE[28]—a manifold learning algorithm designed for high-throughput, high-noise biomedical data—with a second, explicit model of the temporal properties of BOLD activity. This second view is learned from the data to capture the temporal autocorrelation of the signal and the dynamics specific to the stimulus (Fig. 1a).

Using the manifold preservation metric denoised manifold affinity preservation (DeMAP)[28], we benchmark T-PHATE against principal component analysis (PCA) and uniform manifold approximation (UMAP)[32]—dimensionality-reduction methods that have been commonly applied to fMRI data[25,26,33,34]—as well as PHATE[28], a reduced form of T-PHATE excluding the above-mentioned second, temporal view. In Supplementary Information, we compare T-PHATE with further benchmarks such as locally linear embeddings (LLE)[30,35], isometric mapping[29], and t-distributed stochastic neighbor embedding (t-SNE)[36]. We test T-PHATE on two movie-viewing fMRI datasets and find that it both denoises the data and affords enhanced access to brain-state trajectories relative to voxel data and other embeddings. By mitigating noise and voxel covariance, this subspace yields clearer access to regional dynamics of brain signals, which we then relate to time-dependent cognitive features. In all, T-PHATE reveals that information about brain dynamics during naturalistic stimuli lies in a low-dimensional latent space that is best modeled in nonlinear, temporal dimensions.

Results

fMRI is a safe, powerful, and ubiquitous tool for studying how the healthy human brain generates the mind and behavior. However, fMRI data are highly noisy in both space and time. The measured BOLD signal is delayed and blurred by the hemodynamic response with respect to underlying neuronal activity. The naturalistic, dynamic stimuli (for example, movies) increasingly used to probe real-world cognition extend over multiple timescales (for example, conversations, plot lines). For these reasons, fMRI data are autocorrelated in time, and this varies across brain regions according to their functional role in cognition (for example, sensory processing versus narrative comprehension). We therefore designed T-PHATE as a variant of the PHATE algorithm that combines the robust manifold-geometry-preserving properties of PHATE with an explicit model of signal autocorrelation in a dual-diffusion operator.

Evaluating manifold quality on simulated data

We validated the quality of manifolds learned from data with a simulated high signal autocorrelation using DeMAP[28]. DeMAP takes pristine, noiseless data (here, a simulated multivariate time-series in which $f(X_t) = \alpha f(X_{t-1})$) and computes the Spearman correlation between the geodesic distances on the noiseless data and the Euclidean distances of the embeddings learned from the noisy data. Noisy data were generated by adding an error term ϵ to the pristine data, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$. We tested robustness to noise by varying $\sigma = (0, 100)$. Higher DeMAP scores suggest that an embedding method performs effective denoising and preserves the geometric relationships in the original data (Fig. 1c; see Supplementary Fig. 1a for additional benchmarks). All of the methods achieve high DeMAP scores at low noise levels, although PCA visualizations do not seem to reflect much structure (Fig. 1b). With increasing noise, T-PHATE outperforms the other methods at denoising the simulated data, providing visualizations that most closely resemble lower noise conditions.

Temporal trajectories and static information in fMRI embeddings

Having validated that T-PHATE can learn meaningful manifolds from simulated data under noise, we next applied T-PHATE to real fMRI data. First, we embedded the movie-viewing data from the Sherlock and StudyForrest datasets with PCA, UMAP, PHATE, and T-PHATE in two dimensions to visually inspect the properties of the data highlighted by the manifold. Embeddings were computed separately per subject for four regions of interest (ROIs): early visual cortex (EV), high visual cortex (HV), early auditory cortex (EA), and posterior medial cortex (PMC). Visually, T-PHATE embeddings both denoise the time-series data and capture stimulus-related trajectory structure better than other methods, as shown in all brain regions for a sample subject (Fig. 2a). PCA visualizations show no apparent structure or clustering. UMAP shows slight clustering in HV and EA, but often creates small, shattered clusters with local structure placed through the latent space. PHATE yields slight improvements over UMAP, notably in HV and EA, with larger and less disjointed clusters of nearby time points. T-PHATE reveals trajectories through the latent space that are clearly reflective of temporal structure, and also shows a hub in the center of the space.

T-PHATE manifolds also better reflect non-temporal movie features, as shown by classification of these features (Fig. 2b and Supplementary Fig. 3). We trained a support vector machine to predict whether music was playing in the movie from embeddings of neural data (and as a baseline, the original voxel-resolution data). The significance of these predictions was tested by shifting the labels in time with respect to the brain data—which breaks the correspondence between the labels and the data without breaking the internal temporal structure—and recalculating the classification accuracy at each shift to obtain a null distribution of accuracies. Results are presented as the z-score of the true prediction accuracy between brain data and movie labels normalized to the mean and standard deviation of the shifted null distribution.

Alternative representations of time in the brain and tasks

The clarity of the dynamic structure recovered by T-PHATE prompted two follow-up questions. First, does the performance of T-PHATE depend on autocorrelation to represent a temporal structure during manifold learning? We tested two alternative models. PHATE + Time adds the time index label for each sample as an additional feature vector to the voxel data before embedding with PHATE. PHATE + Time manifolds show the clearest temporal trajectory through the embedding space (Fig. 3a), unraveling the brain data in time due to the disproportionate weighting of time in the input data, with less of the hub structure revealed by T-PHATE. Smooth PHATE performs temporal smoothing along each voxel across a number of time points t, where t is learned as the width of the autocorrelation function of the input data before embedding with PHATE. Smooth PHATE yields comparable structure to PHATE or UMAP, with mild clustering and shattered trajectories.

Second, are the benefits of T-PHATE for stimulus feature classification specific to fMRI tasks that are intrinsically linked to time, such as movie watching? We tested T-PHATE on an object category localizer task from StudyForrest with no temporal structure, during which static images of faces, bodies, objects, houses, scenes, and scrambled images were presented in a random order. Data were embedded with T-PHATE, PHATE + Time, and Smooth PHATE in EV and HV. Due to the lack of temporal structure in the task, a meaningful manifold of fMRI activity in visual regions would show a time-independent clustering of categories. T-PHATE manifolds showed the best clustering of stimulus categories in the embedding space (Fig. 3b) when compared with PHATE + Time and Smooth PHATE. Smooth PHATE showed some less-pronounced clustering, whereas PHATE + Time showed minimal clustering by category but retained
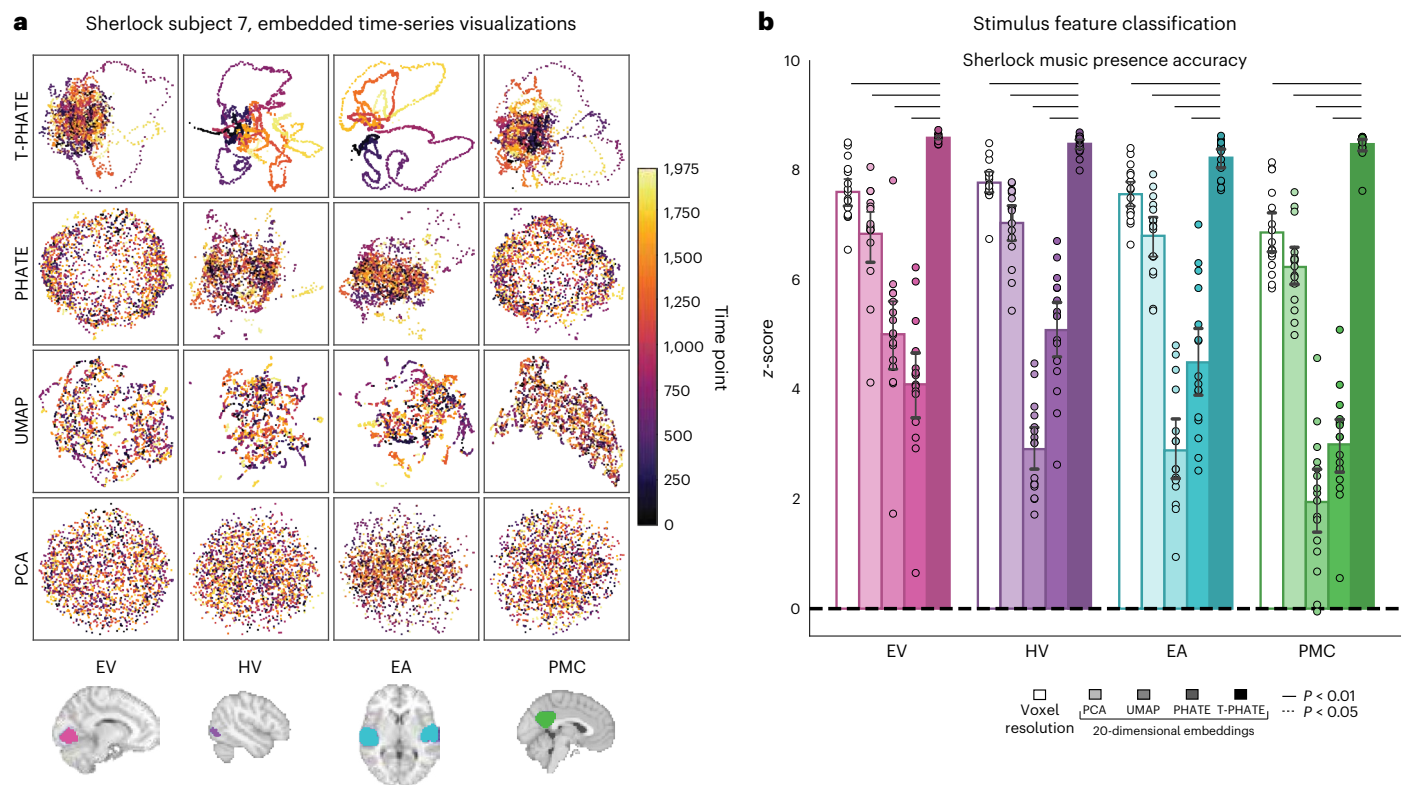
a Sherlock subject 7, embedded time-series visualizations

b Stimulus feature classification

Fig. 2 | Visual validation of T-PHATE embeddings. a, Manifold embeddings using PCA, UMAP, PHATE, and T-PHATE for the Sherlock dataset in an example subject and four ROIs. Scatter plot points are colored by time index. **b**, Support vector classification accuracy predicting the presence of music in the movie at the corresponding time point. Results are presented as the z-score of the true prediction accuracy relative to a null distribution of accuracies, obtained by shifting the music labels in time relative to the embedding data 1,000 times.

Dots represent individual subjects ($n = 16$), bars represent the average z-score across subjects, and error bars represent the 95% confidence interval of the mean, estimated with 1,000 bootstrap iterations. The significance of differences between T-PHATE and other methods is evaluated with permutation tests (10,000 iterations) and corrected for multiple comparisons. The extended results are included in Supplementary Fig. 3.

temporal structure. We used a support vector machine with leave-one-run-out cross-validation to quantify the classification of object categories in the latent space (chance = 1/6; Fig. 3c). Classification accuracy on T-PHATE embeddings surpassed other dimensionality-reduction methods in both the EV and HV ROIs (Fig. 3c; see Supplementary Fig. 1b for further benchmarks). T-PHATE therefore captures the structure in brain activity via autocorrelative denoising, whereas adding time explicitly into the manifold does little to capture data geometry aside from unrolling the signal in time. Temporal smoothing provides an advantage over time-agnostic dimensionality reduction (PCA, PHATE, UMAP), but less of an advantage than the autocorrelation kernel in T-PHATE for both the dynamic movie task and the non-temporal localizer task.

Segmentation of brain-state trajectories into events
We designed T-PHATE to incorporate temporal dynamics (namely, autocorrelation) into a manifold learning algorithm for two reasons. First, as shown above, this helps to denoise fMRI data, given its spatiotemporal noise, and improve subsequent analyses such as feature classification. Second, this may help recover cognitive processes that are represented over time in fMRI data. Many cognitive processes operate on long timescales, including our ability to segment continuous sensory experience into discrete mental events[37–39]. That is, we neither perceive the world as transient from moment to moment nor as amorphous and undifferentiated, but rather as a series of coherent, hierarchically structured epochs[40,41]. For example, consider taking a flight: traveling to the airport, passing through security, boarding, flying, deplaning, and transiting to the destination. These mental events are

reflected in stable states of the brain, which can be captured with a hidden Markov model (HMM) to identify the boundaries between events[42–44]. We hypothesized that an HMM would better segment events from fMRI data after embedding them into a manifold learned with T-PHATE. This would suggest that T-PHATE increases sensitivity to the neural dynamics associated with cognitive processing of the stimulus.

We started by learning the optimal number of events experienced by each ROI using a nested leave-one-subject-out cross-validation procedure on the voxel resolution data (Fig. 4a). We find a consistent pattern of results across datasets where the number of events K is higher for early sensory cortices (EV) than late (HV) and integrative cortices (PMC) (mean ± s.d. across subjects): Sherlock, EV = 46 ± 2, HV = 21 ± 1, EA = 18 ± 1, PMC = 24 ± 1; StudyForrest, EV = 121 ± 3, HV = 70 ± 2, EA = 69 ± 2, PMC = 110 ± 4. After fixing the K parameter for each ROI based on the voxel data, we then learned the number of manifold dimensions that would best reveal temporal structure for each region, subject, and embedding method with a leave-one-subject-out cross-validation procedure. The number of dimensions is low (between 3 and 6) with minor variance across subjects and methods (Supplementary Fig. 2), indicating substantial redundancy, covariance, and noise among voxels in the ROIs.

A goal of T-PHATE is to recover dynamic cognitive processes that may have been obscured by redundancy and noise. If the T-PHATE latent space unveils dynamical structure during movie watching, we should see better HMM performance on data embedded with T-PHATE compared with other methods, or on the voxel resolution data. This can be framed as: (1) how well can an HMM fit the data and (2) how well do boundaries learned by the model capture structure of the neural
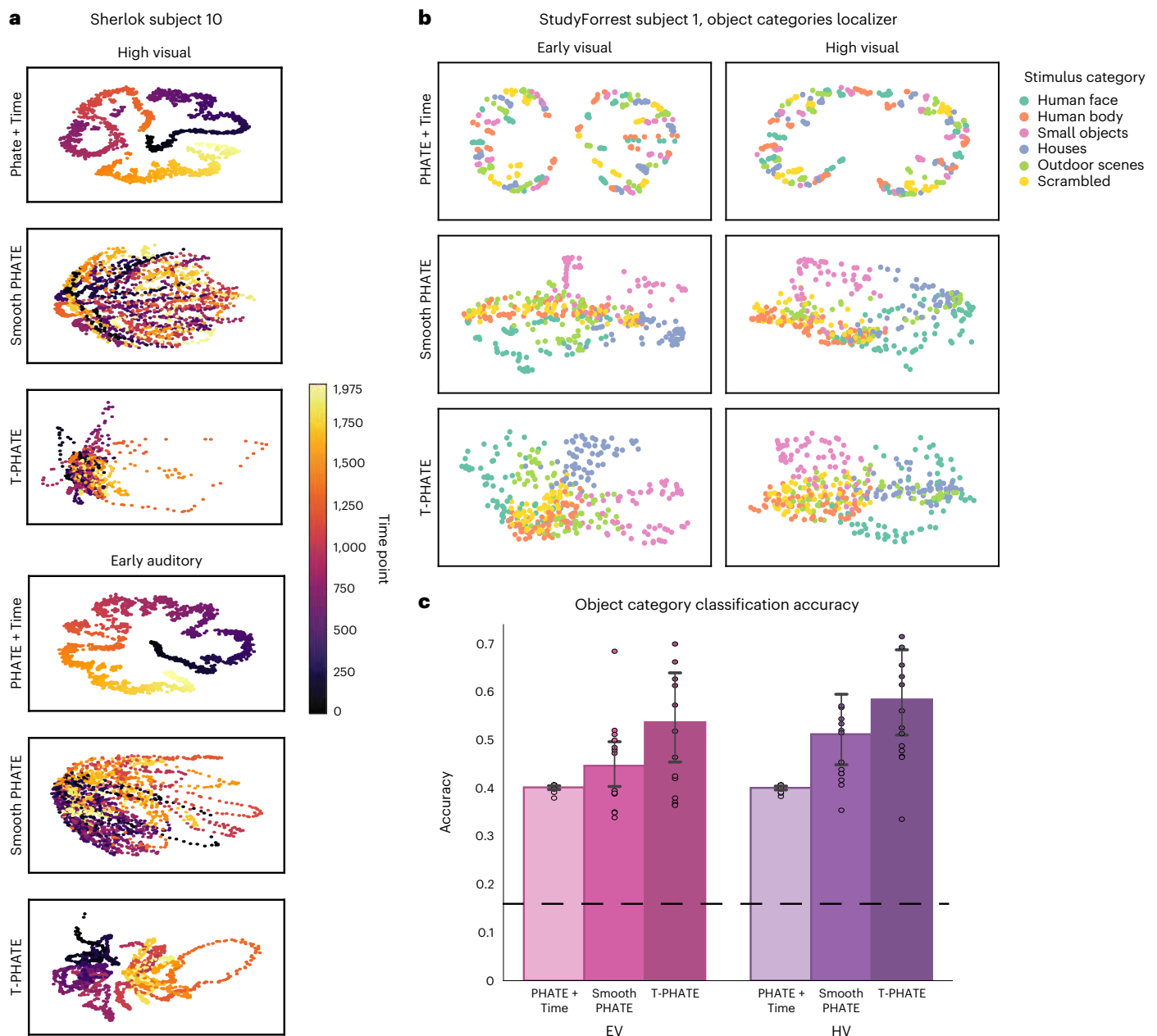
Fig. 3 | Alternative embeddings featuring time. a, Manifold embedding of Sherlock data with PHATE+Time, Smooth PHATE, and T-PHATE, in HV and EA ROIs, colored by time index. **b,** StudyForrest localizer data embeddings using PHATE+Time, Smooth PHATE, and T-PHATE in EV and HV. Points are colored by the stimulus category presented at a given time point, which are presented in a random order. T-PHATE reveals the best clustering by stimulus category, showing that the manifold successfully denois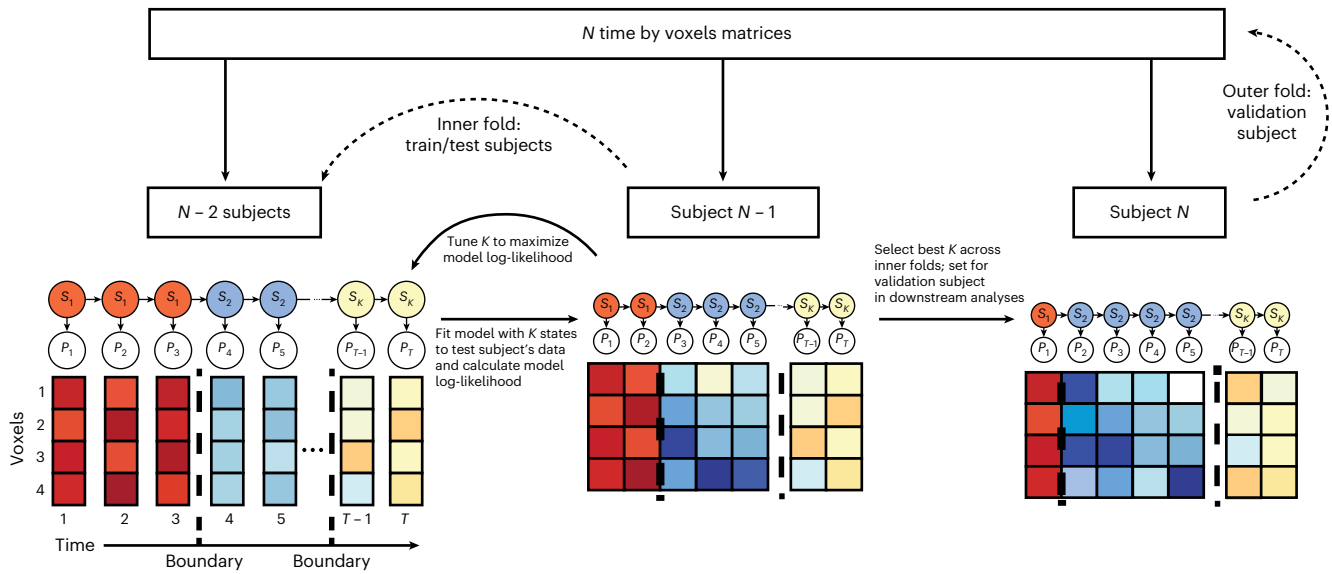es data to learn structure in the data also independent of time. **c,** Support vector classification of stimulus categories presented during the StudyForrest localizer task based on the embedding data. Dots represent individual subjects ($n = 14$); bars represent the average classification accuracy across subjects; error bars represent the 95% confidence interval of the mean, estimated with 1,000 bootstrap iterations; the dashed line represents chance classification (1/6). See Supplementary Fig. 1b for further benchmark results.

data. If an HMM can fit the data well, then there is event-related structure in the data representation for the model to capture. If the event boundaries learned by the model capture this structure, we would further expect a distinction between time points within the same event label versus those spanning across event labels. For each subject, ROI, and embedding method, we fitted an HMM with K hidden states on a M-dimensional embedding (hyperparameter K was cross-validated at the subject level and within ROI; M was cross-validated within ROI and embedding method).

The model fit was quantified on a held-out subject after hyperparameter tuning on a training sample using log-likelihood, which captures the probability that the data input to the model were generated from the learned event structure, and Akaike's information criterion (AIC[45]), which penalizes the log-likelihood of a model by the number of parameters of that model. The statistical significance of the difference in performance between T-PHATE and other methods was assessed within ROIs using a permutation test, correcting for the four multiple comparisons. The model fit for T-PHATE was much better for both the Sherlock and StudyForrest datasets in almost all regions (Supplementary Fig. 4). This indicates that an HMM trained on one representation of the data (the embedding for a given ROI and subject using one of the methods) was more

a HMM hyperparameter tuning procedure
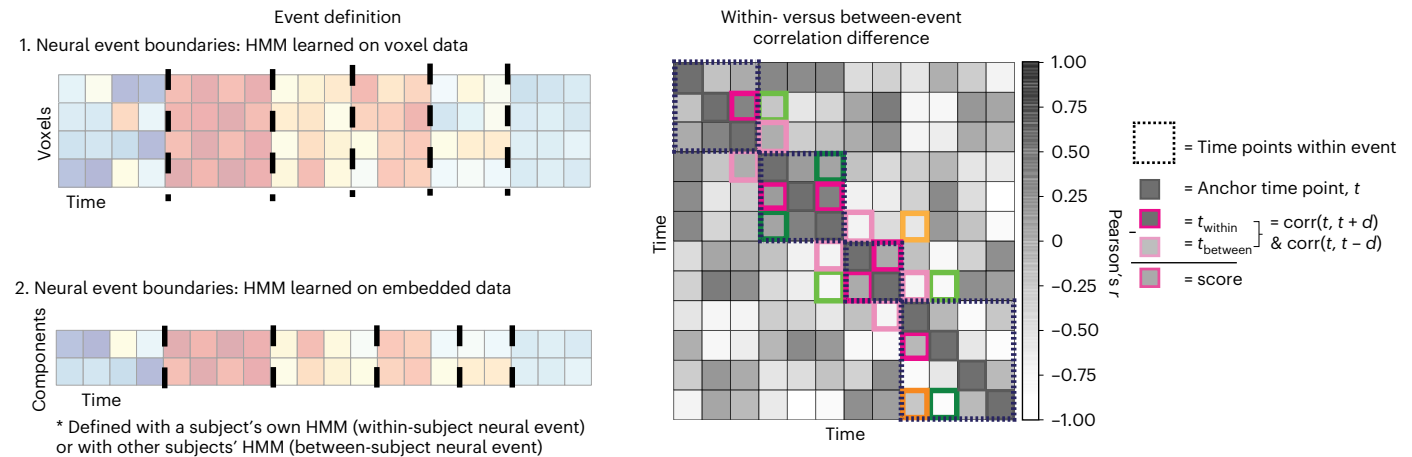

b Analysis of event-boundary fit

Fig. 4 | HMM procedure and analysis. a, Nested cross-validation procedure for hyperparameter tuning. The HMM is trained on the average data for $N - 2$ subjects to find K stable patterns of voxel activation, where the voxel patterns P for each time point belong to a given hidden state S. K is tuned according to model fit on subject $N - 1$ (folding over all of the iterations of $N - 1$ train/test splits) and then the best K value across the folds is applied to the validation subject N. This procedure is repeated for all subjects (see the 'Optimizing the number of neural events' section in Methods for more information). Event boundaries (dashed lines) are identified by the model as time points at a transition between hidden states. For dimensionality-reduced approaches, M is learned with leave-one-subject-out cross-validation, within embedding method and ROI (see the 'Optimizing the manifold dimensionality' section in Methods for more information). **b**. Left: event boundaries are learned both from the data (by applying the HMM on voxel resolution or embedded time-series data and applying those boundaries either within or across subjects)

and from independent sources such as human behavioral ratings. Right: representation of dynamic information in a latent space is quantified by applying event boundaries to embedded data. Data emphasize brain-state transitions if the boundaries maximize the difference between brain activity at time points within- versus between-events. For each time point t_0 in the time-series, we correlated the activation pattern across features at t_0 with that of t_1 and t_2 such that t_1 is in the same event as t_0 (within a dashed box) and t_2 is in a different event. The within- versus between-event correlation difference score is the subtraction of corr(t_2, t_0) from corr(t_0, t_1). To account for intrinsically higher correlations of proximal time points, we restricted comparisons to where $|t_0 - t_1| = |t_0 - t_2|$. Here, corr($t_0, t_1$) is denoted by a darker colored box and its corresponding corr(t_2, t_0) by a lighter shade, where both t_2 and t_1 are d time-steps from t_0. Pink, green, and orange boxes indicate time points when $d = 1$, 2, and 3, respectively.

confident in its learned event structure than a very similar (or, in the case of Supplementary Fig. 7, identical) HMM trained on a different data representation. This is a somewhat unorthodox approach to model fitting (comparing the fit of the same data between models that have a different number of parameters, architecture, or objective function), so we further evaluated the benefits of T-PHATE by examining event-related dynamics and additional outcomes in the brain and behavior.

Within- versus between-event stability of brain states

To validate that the HMM boundaries identified real transitions between brain states, we next tested whether patterns of fMRI activity from pairs of time points labeled as within the same event were more similar to each other than equidistant pairs of time points labeled as from two different events (Fig. 4b). We first compared the performance of voxel resolution data to T-PHATE embeddings, as this type of within- versus between-event correlation difference analysis has been applied
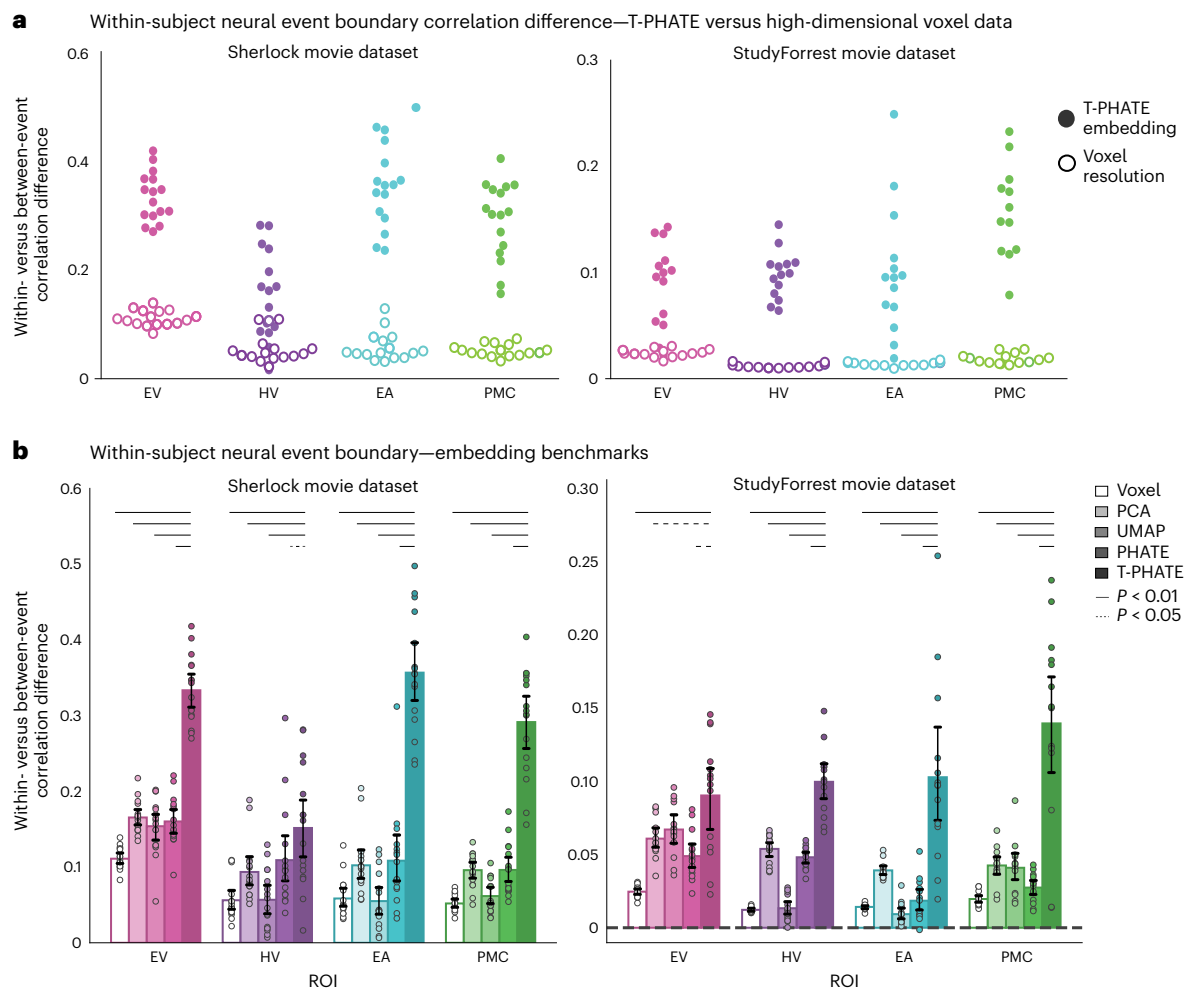
a Within-subject neural event boundary correlation difference—T-PHATE versus high-dimensional voxel data

b Within-subject neural event boundary—embedding benchmarks

Fig. 5 | Evaluating event segmentation within-subject. a, After hyper-parameters K and M were tuned with cross-validation, a new HMM with these parameters was fit to each subject's data to identify neural event boundaries and calculate the within- versus between-event score, shown as one point per subject. Embedding the data with T-PHATE greatly increased the within- versus between-event score over the voxel-resolution data. **b**, We expanded this analysis to embeddings performed with three additional dimensionality-reduction methods. Dots represent individual subjects ($n = 16$ for Sherlock, $n = 14$ for StudyForrest); bars represent the average within- versus between-event score across subjects; error bars represent the 95% confidence interval of the mean, estimated with 1,000 bootstrap iterations. The significance of the differences between T-PHATE and other methods was evaluated with permutation tests (10,000 iterations) and corrected for multiple comparisons. See Supplementary Figs. 5a and 6a for additional benchmark results.

successfully at the voxel level[42,46]. For all subjects, T-PHATE embeddings outperformed the voxel resolution data (Fig. 5a). We next asked whether this benefit was specific to T-PHATE or a generic effect of dimensionality reduction. We thus repeated this analysis on PCA, UMAP, and PHATE embeddings (Fig. 5b; see Supplementary Figs. 5a and 6a for additional benchmarks) and again found better performance for T-PHATE embeddings in all ROIs. In three out of four ROIs for both datasets, T-PHATE doubled the performance of the second-best method. Except for two instances, all of the dimensionality-reduction methods outperformed the voxel resolution data. Thus, although there is a generic advantage of lower dimensionality when revealing signal structure, T-PHATE captures dynamic signals beyond a temporally agnostic manifold.

In the previous analysis, each HMM was trained and tested within the same subject. To ensure that these results were not driven by overfitting, we performed between-subject cross-validation by applying the event boundaries learned from $N - 1$ subjects to held-out data from subject N. The premise for this analysis is that the event structure learned from voxel data can be shared across subjects[42,46]. We asked whether the same holds true in lower dimensions with embeddings performed at the subject level. For each of N subjects (per ROI and embedding method), we repeated the analysis $N - 1$ times using each

of the other subject's HMM-identified boundaries. We again found better performance for T-PHATE relative to other embedding methods overall (Fig. 6a), though with a smaller advantage than when the analysis was performed within-subject. Although T-PHATE outperformed the voxel resolution data in all cases, the voxel resolution data outperformed some of the embedding methods (Sherlock dataset: EV voxel > PCA, PHATE, UMAP; HV voxel > UMAP; PMC voxel > UMAP. StudyForrest dataset: EV voxel > PCA, PHATE; HV voxel > UMAP; EA voxel > UMAP, PHATE; PMC voxel > PCA, UMAP, PHATE). This was not true for the within-subject analysis, which suggests that embedding data in lower, subject-specific dimensions highlights idiosyncratic aspects of event structure. Voxel resolution data are assumed to be in the same anatomical template space for all subjects, which would lend itself better to generalization across subjects. T-PHATE still outperforms other methods in most cases on this task, suggesting that the event structure highlighted in subject-specific T-PHATE embeddings is at least partly generalizable to other subjects.

Relationship between neural and behavioral event boundaries
Identifying neural event boundaries with an HMM is a data-driven approach to understanding how the brain makes sense of
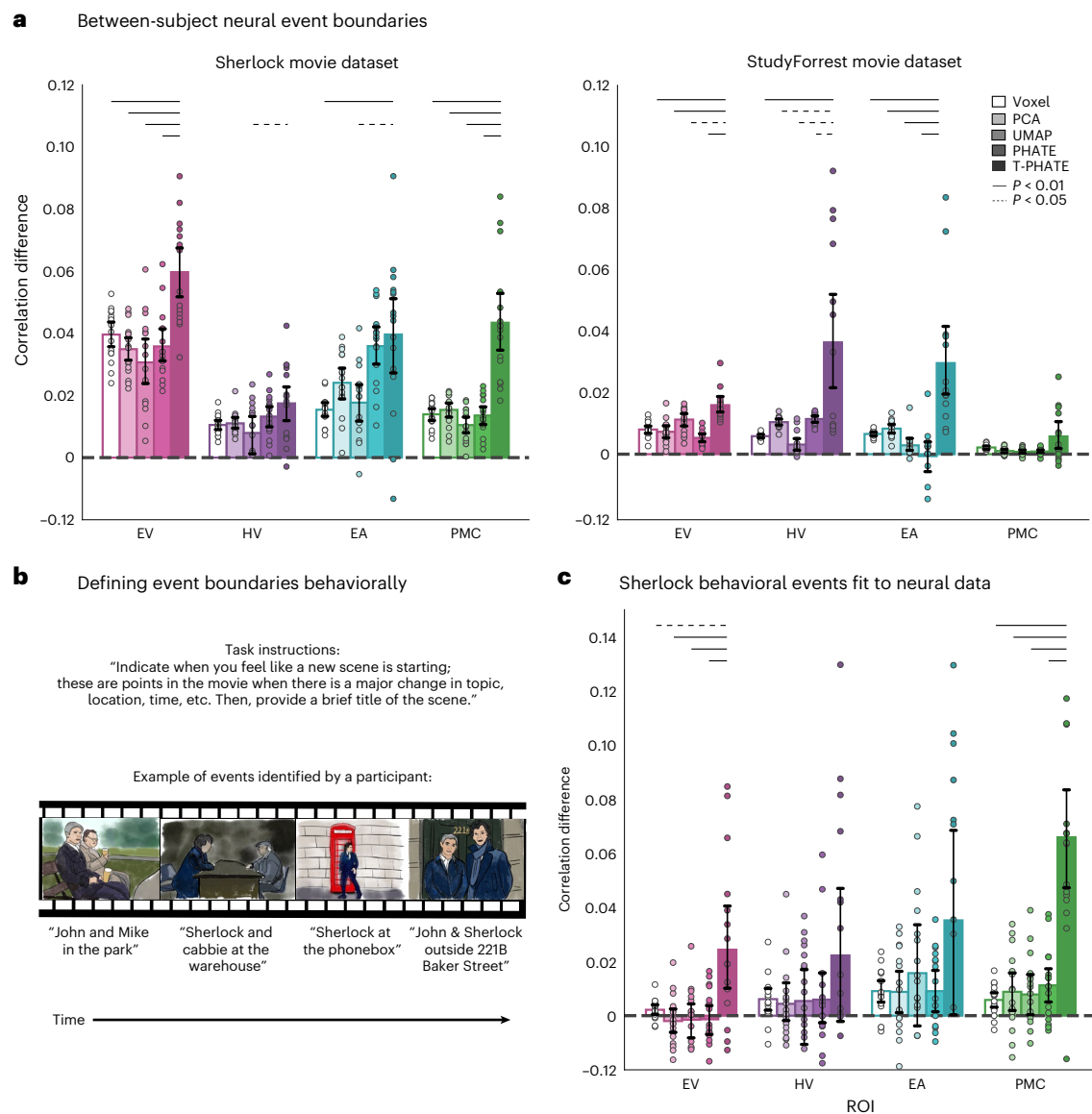
Fig. 6 | Out-of-sample event boundary fit. a, Event boundaries were rotated between subjects by testing subject *N* iteratively with the event boundaries learned from HMMs applied to each of the remaining *N* − 1 subjects (within ROI and embedding type). Within- versus between-event correlation difference was evaluated as in Fig. 4b. T-PHATE embeddings revealed robust event-related structure that generalized across subjects. See Supplementary Figs. 5b and 6b for additional benchmark results for this analysis. **b**, Event boundaries were also identified in behavioral ratings of the Sherlock movie by other subjects. These behavioral boundaries were then applied to the neural data and the within- versus between-event correlation difference was calculated as in Fig. 4b. Images in the figure are drawn for copyright reasons; movies were shown at high resolution in the experiment. **c**, T-PHATE revealed event structure in PMC (and to a weaker extent EV) that mirrored the behavioral boundaries. Dots represent individual subjects (*n* = 16 for Sherlock; *n* = 14 for StudyForrest); bars represent the average within- versus between-event correlation difference across subjects; error bars represent the 95% confidence interval of the mean, estimated with 1,000 bootstrap iterations. The significance of differences between T-PHATE and other methods was evaluated with permutation tests (10,000 iterations) and corrected for multiple comparisons.

continuous experience. This can be compared with boundaries obtained from other tasks, such as having subjects watch the same movie and indicate behaviorally when they believe "something changed"[37–39,47] (Fig. 6b). Such behavioral boundaries have been used to model event structure in the brain previously[47–50], incuding in PMC[42,51]. We used behavioral boundaries to test the generalization of neural event structure to the conscious experience of an out-of-sample cohort. We applied behavioral boundaries from this cohort to the neural embedding data and analyzed their fit as in Fig. 4b. We found much higher fMRI pattern similarity within-versus-between behavioral boundaries in the T-PHATE embedding than in the other methods, at least for EV and PMC (Fig. 6c). Critically, T-PHATE improved on voxel

performance in PMC by an order of magnitude, a particularly striking gain as the behavioral data being predicted were generated by other subjects performing a distinct behavioral task.

Discussion

Many complex aspects of cognition, including attention, memory, language, thinking, navigation, and decision-making, play out over long timescales[52]. These cognitive processes mutually drive our conscious experience in the moment, structuring continuous input into meaningful events[37–39,42,46,53]. To understand the dynamics of how this integration occurs seamlessly, it can be helpful to consider the geometry of brain states. The human fMRI literature

suggests that flexible population codes underlie complex cognitive processing[13,18,42,51,54]. Linear manifolds have so far been used to track discrete, task-related brain states[25,26], and nonlinear manifolds have been used to model dynamics across tasks[21]. These approaches aggregated whole-brain data across dozens of subjects when defining a manifold, which resulted in coarse, population-average representations that could limit utility for understanding individual brains.

With the proper choice of dimensionality-reduction algorithm[6,9], we accounted for these challenges to unveil the dynamics and contents of naturalistic cognition. We investigated the latent space of brain activity from movie-viewing fMRI data. T-PHATE manifolds revealed brain-state trajectories obscured in high-dimensional, noisy, and redundant neural population activity by modeling its latent spatio-temporal geometry. The advantages of T-PHATE representations for capturing temporal dynamics in continuous tasks did not come at the expense of decoding discrete stimuli without temporal structure. Thus, although time-considerate and time-agnostic manifolds can encode static information (Figs. 2b and 3c, and Supplementary Figs. 1b and 3), only the former additionally unveil dynamic information.

There are other multi-view variants of dimensionality-reduction methods, such as Multi-SNE[55] or Multi-LLE[56], which capture multiple views of data by summarizing different types of measurements from the same samples. Whether alternative dimensionality-reduction methods, when combined with an autocorrelation kernel, could yield similar insights remains an open question; however, given that vanilla PHATE yields better performance on most tasks here than benchmarked methods, we predict that T-PHATE benefits from the combination of superior manifold learning of fMRI by PHATE *plus* modeling temporal signals with the autocorrelation view.

The latent dimensions learned by T-PHATE with this combination of nonlinear dimensionality reduction and temporal diffusion outperforms methods lacking either of those components. This suggests that the relation between the dimensions of individual voxels and the dominant signals that emerge from them are tied by nonlinear relations. By modeling a subject's BOLD time-series in these dimensions, we can gauge the trajectory of their brain states throughout the course of the experiment, and we can relate this trajectory to theoretical frameworks such as event segmentation theory that facilitate higher-order cognitive processes in the brain. As this is performed at the level of individual subjects, future investigations could use this approach to probe individual differences, developmental trajectories, and/or clinical disorders in the native latent space.

Beyond the applications in cognitive neuroscience explored here, T-PHATE has promise for other forms of high-dimensional, noisy data. PHATE and variants have already been applied successfully to high-throughput biomedical data types including spatial transcriptomics, single-cell RNA sequencing, and blood sample data from SARS-CoV-2 data[28,57,58]. As T-PHATE converges with PHATE when there is no temporal relation among samples, we foresee T-PHATE having similar advantages in these data types. We expect T-PHATE to afford particular insight into data with temporal structure, given the multi-view approach to integrating temporal dynamics in learning the embedding. Such data types could include, for example, calcium imaging, developmental trajectories, longitudinal health data, disease progression, climate change, economic trends, language evolution, and more. In all, we present a method with promising application to various big-data challenges.

Methods

Manifold embeddings

Different dimensionality-reduction algorithms focus on different aspects of a dataset, and thus amplify distinct signals depending on the structure targeted by the algorithm. We hypothesized that a nonlinear manifold learning algorithm designed to handle noisy, high-dimensional biological data would be best suited for the signal and intrinsic noise of fMRI data. PHATE[28] is a diffusion-based manifold learning method that models local and global structures simultaneously in nonlinear dimensions[58]. Brain activity measured with fMRI is highly noisy in both space and time, with the BOLD signal canonically peaking 4–5 s after stimulus onset before slowly returning to baseline. With a temporally dependent stimulus such as a movie, where conversations and plot lines play out over different timescales, the autocorrelation of nearby time points will likely extend beyond the curve of the BOLD signal and vary by ROI along the hierarchy of temporal integration in the brain[59]. We estimate an autocorrelation kernel for each ROI by correlating each voxel time-series with lagged versions of itself until the correlation drops and stays below zero, then averaging the functions across voxels to get a regional autocorrelation function. This function is then expanded to calculate the transitional probability between all pairs of time points based solely on their estimated autocorrelation and then combined with the PHATE-based transitional probability matrix and embedded in lower dimensions (Fig. 1a).

PHATE

Given a dataset of voxel time-series data, $X = x_1, x_2, ..., x_T$, where $x_t \in \mathbb{R}^N$ at time t is a N-dimensional vector and N is the number of voxels. Construction of the PHATE[28] diffusion geometry has five main steps:

1. The Euclidean distance matrix D is computed between data pairs, where:

$$D(i,j) = ||x_i - x_j||^2 \tag{1}$$

2. D is converted from a distance matrix into a local affinity matrix K using an adaptive bandwidth Gaussian kernel, to capture local neighborhoods in the data:

$$\mathcal{K}(x,y) = \frac{\mathcal{G}(x,y)}{\| \mathcal{G}(x,\cdot)\|_1^\alpha \| \mathcal{G}(y,\cdot)\|_1^\alpha}, \quad \mathcal{G}(x,y) = e^{-\frac{\|x-y\|^\alpha}{\sigma}} \tag{2}$$

3. K is then row-normalized to define transition probabilities into the $T \times T$ row stochastic matrix, P:

$$P(x,y) = \frac{\mathcal{K}(x,y)}{\| \mathcal{K}(x,\cdot)\|_1} \tag{3}$$

4. The probabilities P are then used for the Markovian random-walk diffusion process. The PHATE diffusion timescale, t_D, is then computed, which specifies the number of steps taken in the random-walk process. This parameter provides a tradeoff between encoding local and global information in the embedding, where a larger t_D corresponds to more steps than a smaller t_D. t_D is computed automatically using the spectral or von Neumann entropy of the diffusion operator. P is raised to the power of t_D to perform the the t_D-step random walk over P. Based on the representation of P^{t_D}, PHATE then computes the diffusion potential distance P_D between the distribution at the ith row $P_i^{t_D}$ and the distribution at the jth row $P_j^{t_D}$ (both of which are distributions as P^{t_D} is Markovian):

$$P_D(i,j) = \sqrt{\sum_k \left(\log(P^{t_D}(i,k) - P^{t_D}(j,k)\right)^2} \tag{4}$$

The log scaling in the diffusion potential distance calculation acts as a damping factor which makes faraway points similarly equal to nearby points in terms of diffusion probabilities, giving PHATE the ability to maintain global context.

5. The potential distance matrix P_D is finally embedded with metric MDS (a distance embedding method, using stochastic gradient descent as a solver) as a final step to derive an M-dimensional embedding (or 2–3 dimensions for visualization). For more details about the PHATE algorithm, we refer readers to ref. [28].

T-PHATE

We designed T-PHATE as a variant of PHATE that uses a dual-view diffusion operator to embed time-series data in a low-dimensional space. The first view of the T-PHATE diffusion operator is identical to the PHATE matrix, P_D, defined above. The second view P_T is based on an affinity matrix that summarizes the autocorrelation function of the data[60]. Computing the T-PHATE diffusion operator has the following steps:

1. For each voxel time-series vector $v = v_1, v_2, ..., v_T$, where T is the number of time points, calculate its autocovariance using $T-1$ lags, resulting in a $N \times (T-1)$ matrix where N is the number of voxels. Then, average across the N voxels to obtain a single vector c of autocorrelation at each lag.

2. Smooth c with a rolling average over w time points, where w is set by the user:

$$c(i,j) = \frac{1}{w} \sum_{j=i}^{i+w-1} c(i,j) \qquad (5)$$

w serves as a damping tool to account for possible jittering around where $c = 0$. Here we use $w = 1$, the default, as we did not find much instability in this dataset where $c \to 0$. In other data types, this parameter may be more useful.

3. Find the first lag (lag_{max}) where $c = 0$. This defines the maximum width of smoothing for the temporal affinity matrix A, which is calculated as:

$$A(i,j) = \begin{cases} c(i,j), & \text{if } 0 < |i-j| \le lag_{max} \\ 0, & \text{otherwise} \end{cases} \qquad (6)$$

4. Convert the autocorrelation matrix into the transition probability matrix P_T by row-normalizing the affinity matrix A and powering it to t_D, as in equation (3).

5. Combine P_T with the result of step 4 from the PHATE algorithm via alternating diffusion:

$$P = P_D P_T \qquad (7)$$

6. Embed with metric MDS into M dimensions (where $M = 2 - 3$ for visualization or higher for downstream analysis).

This dual-view diffusion step allows T-PHATE to learn data geometry and latent signals that represent cognitive processes that play out over longer temporal windows. We compare T-PHATE's performance at learning dynamic neural manifolds with common dimensionality-reduction algorithms that are agnostic to time including PCA, UMAP[32], vanilla PHATE[58], LLE[30,35], isometric mapping[29], and t-SNE[36]. To test whether our autocorrelation kernel was the best approach to modeling temporal dynamics in this manifold, we tested two additional versions of PHATE: incorporating time labels as a feature vector in the input data (PHATE + Time) and smoothing the data temporally over the data's autocorrelation function as learned by T-PHATE (Smooth PHATE), to test whether T-PHATE's effects can be accomplished by incorporating time into the PHATE embeddings without the additional kernel. fMRI data were extracted from ROIs and z-scored before embedding.

Event segmentation modeling

Human experience of real life is continuous, yet perception and conception of this experience is typically divided into discrete moments or events. Event segmentation theory explains that humans automatically divide continuous streams of information into discrete events[37,39] to form, organize, and recollect memories, make decisions, and predict the future[61]. Participants show high consistency in explicitly segmenting continuous stimuli[38] and also in how

their brains represent these event boundaries[42,46]. Event boundaries can be represented—both behaviorally and in brain activity—at different timescales depending on the information being used to draw event boundaries[53]. In the brain, event boundaries are reflected by shifts in the stability of regional activity patterns, and we can learn the structure of events with a variant of an HMM[42]. During a continuous stimulus such as a movie, different brain regions represent events along different timescales, which reflect the dynamics of the information being represented by a given region. For example, early sensory regions represent low-level information about a stimulus. In most movie stimuli, low-level sensory features shift quickly as cameras change angles or characters speak, and so early sensory regions show more frequent event shifts[42,44]. Later sensory regions represent longer timescale events, such as conversations and physical environments that change less frequently than features such as camera angles. Regions associated with memory or narrative processing represent events on longer timescales, and these boundaries best correspond with scene changes marked by human raters[43].

We used the HMM variant presented in ref. [42] and implemented in BrainIAK[62] to learn from different representations of BOLD time-series data where a brain region experiences event boundaries, or transitions between stable states of activation. See ref. [42] for more details about the formulation of this model. Given an activation time-series of a brain region during a continuous stimulus, the HMM identifies stable activity patterns or "events" that are divided by boundaries, where the time-series transitions between two stable patterns of activation. This is done iteratively and the log-likelihood of the model fit is evaluated at each iteration, and model fitting stops when the log-likelihood begins to decrease. The first step of this process is to use the HMM to learn from the data the number of events a brain region represents for a given stimulus. Past studies have used HMMs for event segmentation on multivoxel activity patterns and have validated this approach against behavioral segmentation[42,43,46]. This shows that voxel resolution data reflect meaningful event segmentation, so we chose to estimate the optimal number of events (K) for each brain region using the voxel resolution data (Fig. 4), which also prevents overfitting to the embedding data.

To run an HMM for event segmentation on our manifold embedding data, we needed to tune two parameters for each subject: K, or the number of hidden states through which a brain region passes during the stimulus, and M, the dimensionality of the latent space that captures this signal.

Optimizing the number of neural events. We estimated a hyperparameter K for each ROI and subject from the voxel resolution data and held this constant for a subject and ROI across all embedding methods. This was performed with a nested cross-validation procedure. In the outer loop, a subject N was held out as a validation subject, and the parameter identified by the inner loop was applied to this subject. In the inner loop, another subject $N-1$ was held out as a test subject, and the BOLD data were averaged over the remaining $N-2$ training subjects to get an average time-series matrix for this region. Importantly, as voxel space is assumed to be the same across subjects (as they have all been aligned to a standard template brain based on common anatomy), we can assume correspondence between the features of these time-series matrices and can thus average them together (the same assumption cannot be made for embedding data, as those spaces are learned within-subject and are thus subject-specific). We define K as the number of hidden states in an HMM, and we optimize the K parameter by fitting a model with K states to the training data from the $N-2$ subjects and then applying that model to subject $N-1$, where the model is scored using the log-likelihood. This is repeated for all K values, testing K from 2 to 200 for Sherlock and 2 to 400 for StudyForrest; the test subject $N-1$ is then folded over for each subject in the training set, resulting in $N-1$ sets of scores for each validation subject, which are then averaged

over the $N − 1$ subjects. The K value that maximizes the log-likelihood after averaging over the inner-fold is set as subject N's K value for this ROI. After definition on the voxel resolution data, K for each subject and ROI was held constant for all embedding types. For subsequent analyses, a new HMM was fitted with K states to each subject's data.

Optimizing the manifold dimensionality. We used a leave-one-subject-out cross-validation scheme to select the dimensionality of manifold embeddings. Holding out each subject one at a time, we embedded the data for each of $N − 1$ training subjects into M dimensions, testing values of M ranging from 2 to 10. As each training subject's data were embedded into a distinct feature space, we fit an HMM to each embedding separately, and scored the model fit to each M-dimensional embedding using the within- versus between-event correlation difference (presented in Fig. 4b). We then averaged the correlation difference scores across the $N − 1$ training subjects to obtain one average score per M-value tested. We chose the M-value that maximized the within-versus-between score in the training set and fixed that as the M-value for the held-out subject N in all subsequent analyses. This procedure was performed separately for each ROI and embedding method.

To control for possible effects of embedding in different numbers of dimensions across different methods, we repeated all analyses without the M optimization step (Supplementary Fig. 7). We instead tested all embedding methods in $M = 3$ dimensions (chosen because three dimensions is the maximum embedding dimensionality for t-SNE). Our control and optimized M analyses were highly correlated (within- versus between-event correlation difference, Pearson's $r = 0.930$; log-likelihood of model fit, $r = 0.993$; AIC of model fit, $r = 0.998$).

Evaluating the fit of event boundaries. To quantify how well a learned manifold embedding amplifies dynamic structure in the data, we calculated difference in correlation across latent dimensions for pairs of time points within and across event boundaries. For this calculation, we restricted the time points that went into the calculation as follows. We calculated the length of the longest event, and only considered time points that were less than that temporal distance apart. We anchored a comparison on each time point t in the time-series, and for each temporal distance n ranging from one time point to the length of the longest event; we considered a pair of time points $t − d$ and $t + d$ if one time point fell in the same event as the anchor and the other a different event. We then took the correlation between the anchor and each time point, binning them according to whether they were a within-event comparison or a between-event comparison. This process assures that there are equal numbers of time points going into each comparison type (Fig. 4b).

One issue of defining and testing event boundaries within-subject is that the dynamics captured are subject-specific. Studies using voxel-resolution data to perform this HMM event segmentation analysis have shown that there is shared structure in voxel representations of event dynamics[42,44,46]. We therefore asked whether boundaries identified on one subject generalized across all other subjects. To test this, we cross-validated the event boundaries across subjects with a nested cross-validation procedure. In the outer loop, we held out one subject's data for testing. In the inner loop, we performed leave-one-subject-out cross-validation to learn the best number of events (K) for the HMM and the optimal dimensionality (M) for embedding methods. After learning this K parameter, we fit an HMM with K states on each of the training subjects' data to identify their event boundaries. These boundaries were then applied to the test subject one by one, scored with the within- versus between-event correlation difference, and then averaged over the training segmentations. This procedure was iterated with all subjects serving as test subject.

To assess the behavioral relevance of the neural event segmentation, we used a set of event boundaries identified by a separate cohort of Sherlock study participants. These participants watched the Sherlock episode outside of the scanner and were asked to indicate where they believed a new event began[63] (Fig. 6b). We applied these human-labeled boundaries to the neural data and measured the fit of the boundaries (as outlined above) to gauge how the embeddings not only highlight neural dynamics but how those neural dynamics relate to the conscious, real-world experience of events.

Statistical tests
All results present the following pairwise statistical testing comparing T-PHATE with each other method. For each pair of methods, within ROI and dataset, we calculated the difference in scores across the two methods, within subjects to preserve subject-wise random effects. We then generated a null distribution of difference scores across the two methods by randomly permuting the method label of a subject's score 10,000 times and recomputing the mean difference between the methods. We then tested whether T-PHATE outperformed the other method by calculating a one-tailed P-value of the true difference relative to the null distribution, and corrected for multiple comparisons using the Bonferroni method. All figures show bootstrapped 95% confidence intervals of the mean performance across subjects within method, tested with 1,000 iterations of bootstrap resampling.

Data
Sherlock dataset. See the original publication of ref. [51] for full details on the Sherlock data. Here we used data from the sixteen participants who viewed a 48 min clip of the BBC television series 'Sherlock'. Data were collected in two fMRI runs of 946 and 1,030 time points (repetition time; TRs) and was downloaded from the DataSpace public repository (http://arks.princeton.edu/ark:/88435/dsp01nz8062179). This experiment was approved by the Princeton University Institutional Review Board. The data were collected on a Siemens Skyra 3 T scanner with a 20-channel head coil. Functional images were acquired with a T2*-weighted echo-planar imaging sequence (TE, 28 ms; TR, 1.5 s; 64° flip angle, whole brain coverage with 27 slices of 4 mm thickness and $3 × 3$ mm^2 in-plane resolution, $192 × 192$ mm^2 FOV). Anatomical images were acquired with a T1-weighted MPRAGE pulse sequence with 0.89 mm^3 resolution. Slice-time correction, motion correction, linear detrending, high-pass filtering (140 s cut-off), and co-registration and affine transformation of functional volumes to the Montreal Neurological Institute template were all performed with fMRI Software Library (FSL). Functional images were then resampled from native resolution to 3 mm isotropic voxels for all analyses, z-scored across time at every voxel and smoothed with a 6 mm kernel.

StudyForrest dataset. For full details on the StudyForrest data, see the original publication of the movie dataset[64] and the localizer extension dataset[65]. Here we included data from fourteen participants who completed both tasks. All participants were native German speakers. This experiment was approved by the Ethics Committee of the Otto-von-Guericke University. In the movie-viewing task, participants watched a 2 h version of the movie 'Forrest Gump.' These data were collected in eight fMRI runs resulting in a full time-series of 3,599 TRs (451, 441, 438, 488, 462, 439, 542 and 338 per run). Movie data were collected on a 3 T Philips Achieva dStream scanner with a 32-channel headcoil. Functional images were acquired with a T2*-weighted echo-planar imaging sequence (TE, 30 ms; TR, 2 s; whole brain coverage with 35 slices of 3 mm thickness and $3 × 3$ mm^2 in-plane resolution, 240 mm FOV). In the localizer task, the same fifteen participants viewed 24 unique grayscale images from six categories (human faces, human bodies without heads, houses, small objects, outdoor scenes and scrambled images) in four (156 TR) block-design runs with two 16 s

blocks per stimulus category per run. Localizer data were collected on a 3 T Philips Achieva scanner with a 32-channel headcoil. Functional images for the localizer task were acquired with a T2*-weighted echo-planar imaging sequence (TE, 30 ms; TR, 2 s; 90° flip angle, 35 slices of 3 mm thickness and $3 \times 3\ mm^2$ in-plane resolution, 240 mm FOV). Structural images were acquired with a 3 T Philips Achieva using a 32-channel headcoil. T1-weighted anatomical images were acquired with a 3D turbo field echo sequence with 0.67 mm isotropic resolution. Slice-time correction, co-registration and affine transformation to Montreal Neurological Institute template were performed with FSL where functional images were resampled to 3 mm isotropic voxels. Additional preprocessing included linear detrending, high-pass filtering (100 s cut-off), spatial smoothing with a 6 mm kernel, and nuisance regression (including six motion parameters, global signal, white matter and cerebrospinal fluid) and z-scoring within voxel to match the Sherlock preprocessing.

ROI selection

We selected four ROIs based on the original publication of the Sherlock dataset[51], subsequent publications[42] and regions known to have reliably strong signals in response to audiovisual movie stimuli[66]. The EV, EA and HV region masks were based on a functional atlas defined with resting-state connectivity[67]. As in the original Sherlock publication, we defined a PMC ROI as the posterior medial cluster of voxels within the dorsal default mode network[51]. In the Sherlock data, the dimensionality of voxel-resolution ROIs are as follows: EV = 307, HV = 571, EA = 1,018, PMC = 481. In the StudyForrest data, the dimensionality of voxel-resolution ROIs are as follows: EV = 166, HV = 456, EA = 657, PMC = 309.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The Sherlock dataset was downloaded from the Dataspace Public Repository at the following link: http://arks.princeton.edu/ark:/88435/dsp01nz8062179. The StudyForrest dataset was accessed via DataLad[68] from: https://github.com/psychoinformatics-de/studyforrest-data. Steps to reproduce our preprocessing pipeline and ROI extraction are available here: https://github.com/ericabusch/tphate_analysis_capsule (ref. [69]). Source Data are provided with this paper.

Code availability

Data analysis code is written as custom Python scripts (v.3.6.13) based on sci-kit learn v.0.23.2 (https://scikit-learn.org/), nilearn v.0.9.2 (https://nilearn.github.io), nibabel v.4.0.1 (https://github.com/nipy/nibabel), PHATE v.1.0.7 (https://phate.readthedocs.io/en/stable/)[28] and Brainiak v.0.11 (https://brainiak.org/)[62]. T-PHATE is available as a Python package at: https://github.com/KrishnaswamyLab/TPHATE (ref. [70]). The pipeline to replicate all of the analyses presented here is available at: https://github.com/ericabusch/tphate_analysis_capsule (ref. [69]).

References

1. Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**, 358–366 (2006).
2. Laurent, G. Olfactory network dynamics and the coding of multidimensional signals. *Nat. Rev. Neurosci.* **3**, 884–895 (2002).
3. Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I. & Shenoy, K. V. Cortical preparatory activity: representation of movement or first cog in a dynamical machine? *Neuron* **68**, 387–400 (2010).
4. Chang, L. & Tsao, D. Y. The code for facial identity in the primate brain. *Cell* **169**, 1013–1028.e14 (2017).
5. Freiwald, W. A. & Tsao, D. Y. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* **330**, 845 (2010).
6. Jazayeri, M. & Ostojic, S. Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Curr. Opinion Neurobiol.* **70**, 113–120 (2021).
7. Hennig, J. A. et al. Constraints on neural redundancy. *eLife* **7**, e36774 (2018).
8. Nieh, E. H. et al. Geometry of abstract learned knowledge in the hippocampus. *Nature* **595**, 80–84 (2021).
9. Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **17**, 1500–1509 (2014).
10. Norman, K. A., Polyn, S. M., Detre, G. J. & Haxby, J. V. Beyond mind-reading: multi-voxel pattern analysis of fmri data. *Trends Cogn. Sci.* **10**, 424–430 (2006).
11. Cox, D. D. & Savoy, R. L. Functional magnetic resonance imaging (fMRI) 'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage* **19**, 261–270 (2003).
12. Kamitani, Y. & Tong, F. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* **8**, 679–685 (2005).
13. Haxby, J. V. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
14. Haynes, J.-D. & Rees, G. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* **7**, 523–534 (2006).
15. Polyn, S. M., Natu, V. S., Cohen, J. D. & Norman, K. A. Category-specific cortical activity precedes retrieval during memory search. *Science* **310**, 1963–1966 (2005).
16. Peelen, M. V., Atkinson, A. P. & Vuilleumier, P. Supramodal representations of perceived emotions in the human brain. *J. Neurosci.* **30**, 10127–10134 (2010).
17. Yeshurun, Y., Nguyen, M. & Hasson, U. Amplification of local changes along the timescale processing hierarchy. *Proc. Natl Acad. Sci. USA* **114**, 9475–9480 (2017).
18. Davatzikos, C. et al. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *NeuroImage* **28**, 663–668 (2005).
19. Birn, R. M., Smith, M. A., Jones, T. B. & Bandettini, P. A. The respiration response function: the temporal dynamics of fMRI signal fluctuations related to changes in respiration. *NeuroImage* **40**, 644–654 (2008).
20. Turk-Browne, N. B. Functional interactions as big data in the human brain. *Science* **342**, 580–584 (2013).
21. Gao, S., Mishne, G. & Scheinost, D. Nonlinear manifold learning in functional magnetic resonance imaging uncovers a low-dimensional space of brain dynamics. *Hum. Brain Mapp.* **42**, 4510–4524 (2021).
22. Olszowy, W., Aston, J., Rua, C. & Williams, G. B. Accurate autocorrelation modeling substantially improves fMRI reliability. *Nat. Commun.* **10**, 1220 (2019).
23. Golesorkhi, M. et al. The brain and its time: intrinsic neural time-scales are key for input processing. *Commun. Biol.* **4**, 1–16 (2021).
24. Ito, T., Hearne, L. J. & Cole, M. W. A cortical hierarchy of localized and distributed processes revealed via dissociation of task activations, connectivity changes, and intrinsic timescales. *NeuroImage* **221**, 117141 (2020).
25. Shine, J. M. et al. The dynamics of functional brain networks: integrated network states during cognitive task performance. *Neuron* **92**, 544–554 (2016).
26. Shine, J. M. et al. Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nat. Neurosci.* **22**, 289–296 (2019).
27. Allen, E. A. et al. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex* **24**, 663–676 (2012).

28. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).

29. Tenenbaum, J. B., De Silva, V. & Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *Science* **290**, 2319–2323 (2000).

30. Roweis, S. T. & Saul, L. K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**, 2323–2326 (2000).

31. Salhov, M., Bermanis, A., Wolf, G. & Averbuch, A. Approximately-isometric diffusion maps. *Appl. Comput. Harmon. Anal.* **38**, 399–419 (2015).

32. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnol.* **37**, 38–44 (2019).

33. Gotts, S. J., Gilmore, A. W. & Martin, A. Brain networks, dimensionality, and global signal averaging in resting-state fMRI: hierarchical network structure results in low-dimensional spatiotemporal dynamics. *NeuroImage* **205**, 116289 (2020).

34. Casanova, R. et al. Embedding functional brain networks in low dimensional spaces using manifold learning techniques. *Front. Neuroinform.* **15**, 740143 (2021).

35. Mannfolk, P., Wirestam, R., Nilsson, M., Ståhlberg, F. & Olsrud, J. Dimensionality reduction of fMRI time series data using locally linear embedding. *Magn. Res. Mater. Phy.* **23**, 327–338 (2010).

36. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

37. Radvansky, G. A. & Zacks, J. M. Event boundaries in memory and cognition. *Curr. Opin. Behav. Sci.* **17**, 133–140 (2017). Memory in time and space.

38. Zacks, J. M., Speer, N. K., Swallow, K. M. & Maley, C. J. The brain's cutting-room floor: segmentation of narrative cinema. *Front. Hum. Neurosci.* **4**, 168 (2010).

39. Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S. & Reynolds, J. R. Event perception: a mind-brain perspective. *Psychol. Bull.* **133**, 273–293 (2007).

40. Zacks, J. M., Tversky, B. & Iyer, G. Perceiving, remembering, and communicating structure in events. *J. Exp. Psychol. Gen.* **130**, 29–58 (2001).

41. Kurby, C. A. & Zacks, J. M. Segmentation in the perception and memory of events. *Trends Cogn. Sci.* **12**, 72–79 (2008).

42. Baldassano, C. et al. Discovering event structure in continuous narrative perception and memory. *Neuron* **95**, 709–721.e5 (2017).

43. Lee, C. S., Aly, M. & Baldassano, C. Anticipation of temporally structured events in the brain. *eLife* **10**, e64972 (2021).

44. Baldassano, C., Hasson, U. & Norman, K. A. Representation of real-world event schemas during narrative perception. *J. Neurosci.* **38**, 9689–9699 (2018).

45. MacDonald, I. L. & Zucchini, W. *Hidden Markov and Other Models for Discrete-Valued Time Series* Vol. 110 (CRC, 1997).

46. Yates, T. S. et al. Neural event segmentation of continuous experience in human infants. *Proc. Natl Acad. Sci. USA* **119**, e2200257119 (2022).

47. Speer, N. K., Zacks, J. M. & Reynolds, J. R. Human brain activity time-locked to narrative event boundaries. *Psychol.Sci.* **18**, 449–455 (2007).

48. DuBrow, S. & Davachi, L. The influence of context boundaries on memory for the sequential order of events. *J. Exp. Psychol. Gen.* **142**, 1277–1286 (2013).

49. DuBrow, S. & Davachi, L. Temporal binding within and across events. *Neurobiol. Learn. Mem.* **134**, 107–114 (2016).

50. Ezzyat, Y. & Davachi, L. Similarity breeds proximity: pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron* **81**, 1179–1189 (2014).

51. Chen, J. et al. Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* **20**, 115–125 (2017).

52. Hasson, U., Yang, E., Vallines, I., Heeger, D. J. & Rubin, N. A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* **28**, 2539–2550 (2008).

53. Zacks, J. M. et al. Human brain activity time-locked to perceptual event boundaries. *Nat. Neurosci.* **4**, 651–655 (2001).

54. Haxby, J. V., Connolly, A. C. & Guntupalli, J. S. Decoding neural representational spaces using multivariate pattern analysis. *Ann. Rev. Neurosci.* **37**, 435–456 (2014). PMID: 25002277.

55. Rodosthenous, T., Shahrezaei, V. & Evangelou, M. S-multi-SNE: semi-supervised classification and visualisation of multi-view data. Preprint at https://arxiv.org/abs/2111.03519 (2021).

56. Rodosthenous, T., Shahrezaei, V. & Evangelou, M. Multi-view data visualisation via manifold learning. Preprint at https://arxiv.org/abs/2101.06763 (2021).

57. Kuchroo, M. et al. Multiscale phate identifies multimodal signatures of covid-19. *Nat. Biotechnol.* **40**, 681–691 (2022).

58. Moon, K. R. et al. Manifold learning-based methods for analyzing single-cell RNA-sequencing data. *Curr. Opin. Syst. Biol.* **7**, 36–46 (2018).

59. Himberger, K. D., Chien, H.-Y. & Honey, C. J. Principles of temporal processing across the cortical hierarchy. *Neuroscience* **389**, 161–174 (2018). Sensory Sequence Processing in the Brain.

60. Brockwell, P. J. & Davis, R. A. *Introduction to Time Series and Forecasting* (Springer, 2002).

61. Shin, Y. S. & DuBrow, S. Structuring memory through inference-based event segmentation. *Topics Cogn. Sci.* **13**, 106–127 (2021).

62. Kumar, M. et al. *Brainiak: The Brain Imaging Analysis Kit* (BrainIAK, 2020).

63. Vodrahalli, K. et al. Mapping between fMRI responses to movies and their natural language annotations. *NeuroImage* **180**, 223–231 (2018).

64. Hanke, M. et al. A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. Sci. Data 3, 160092 (2016)..

65. Sengupta, A. et al. A studyforrest extension, retinotopic mapping and localization of higher visual areas. *Sci. Data* **3**, 160093 (2016).

66. Haxby, J. V., Guntupalli, J. S., Nastase, S. A. & Feilong, M. Hyperalignment: modeling shared information encoded in idiosyncratic cortical topographies. *eLife* **9**, e56601 (2021).

67. Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V. & Greicius, M. D. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb Cortex* **22**, 158–165 (2012).

68. Halchenko, Y. O. et al. Datalad: distributed system for joint management of code, data, and their relationship. *J. Open Source Softw.* **6**, 3262 (2021).

69. Busch, E. *ericabusch/tphate_analysis_capsule* Version 2 release (Zenoodo, 2023); https://doi.org/10.5281/zenodo.7626543

70. Busch, E. *Krishnaswamylab/tphate* Initial release (Zenodo, 2023); https://doi.org/10.5281/zenodo.7637523

Acknowledgements

Author contributions

E.L.B., J.H., G.W., G.L., S.K. and N.B.T-B. conceived the idea. E.L.B., J.H., A.B. and S.K. designed the T-PHATE algorithm. E.L.B., J.H. and T.W. developed the T-PHATE software. E.L.B. curated the data, and

designed and performed analyses. E.L.B., S.K. and N.B.T-B. wrote the initial paper draft. All authors revised the paper. S.K. and N.B.T-B. jointly supervised the work.

Competing interests

S.K. is a visiting professor at META Fundamental AI Research (FAIR). The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43588-023-00419-0.

Correspondence and requests for materials should be addressed to Smita Krishnaswamy.

Peer review information *Nature Computational Science* thanks the anonymous reviewers for their contribution to the peer review of

this work. Handling editor: Ananya Rastogi, in collaboration with the *Nature Computational Science* team.

Reprints and permissions information is available at www.nature.com/reprints.

nature portfolio

Corresponding author(s): Smita Krishnaswamy

Last updated by author(s): Feb 8, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We did not develop any software for data collection.
Data analysis	Data analysis code is written as custom python scripts (version 3.6.13) based on sci-kit learn version 0.23.2 (https://scikit-learn.org/), nilearn version 0.9.2 (https://nilearn.github.io), nibabel version 4.0.1 (https://nipy.org/nibabel/), PHATE version 1.0.7 (https://phate.readthedocs.io/en/stable/), and Brainiak (version 0.11) (https://brainiak.org/). The method presented here, TPHATE, is released publicly on GitHub (https://github.com/KrishnaswamyLab/TPHATE) (DOI: 10.5281/zenodo.7637523). Analysis scripts are also released publicly on GitHub (https://github.com/ericabusch/tphate_analysis_capsule) (DOI: 10.5281/zenodo.7626543).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

The "Sherlock" dataset was downloaded from the Dataspace Public Repository at the following link: http://arks.princeton.edu/ark:/88435/dsp01nz8062179. The "StudyForrest" dataset was accessed via DataLad from: https://github.com/psychoinformatics-de/studyforrest-data. Steps to reproduce our preprocessing pipeline and ROI extraction are available here: https://github.com/ericabusch/tphate_analysis_capsule.

Human research participants

Reporting on sex and gender	Both StudyForrest datasets utilized here include the same 15 participants, 6 identifying as female. We excluded 1 participant for missing data, resulting in 14 participants with full data. The Sherlock dataset recruited 22 participants (10 identifying as female). Eventually, 5 were excluded due to head movement, and 1 due to stimulus presentation malfunction, resulting in 16 participants with full data.
Population characteristics	StudyForrest dataset: mean age 29.4 years, range 21-39 years, native German speakers, self-reported normal hearing and no history of neurological disorders. Sherlock dataset: mean age 20.8, range 18-26, native English speakers, right-handed, self-reported normal or corrected-to-normal vision.
Recruitment	StudyForrest dataset: subjects volunteered to participate in the study by responding to a bulletin posting (only subjects who completed the original movie experiment participated in the localizer extension experiment). Subjects received monetary compensation for participation and signed informed consent for public sharing of all anonymized data. Sherlock dataset: subjects were recruited from the Princeton University community. Subjects received monetary compensation for participation and signed informed consent for public sharing of all anonymized data.
Ethics oversight	StudyForrest dataset: approved by the Ethics Committee of the Otto-von-Guericke University Sherlock dataset: approved by the Princeton University Institutional Review Board.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	StudyForrest dataset: 15 subjects (down to 14 after missing data), sample size reported to have been chosen as comparable to current literature. Sherlock dataset: 22 subjects (down to 16 after exclusions), sample size reported in original publication to have been chosen as comparable to the current literature. We included two datasets (resulting in 3 separate tasks), as is common in the fMRI literature, to show generalizability of our method and effects across tasks, scanner types, and participant demographics.
Data exclusions	StudyForrest dataset: 1 subject was excluded due to missing data. Sherlock dataset: 5 subjects were excluded due to excessive head movement in the scanner, and 1 was excluded as their scan ended early, meaning they missed 75 seconds of the stimulus.
Replication	As all analyses reported in the present study are performed at the single-subject level. The effects are replicated across subjects, with cross-validation, and across datasets.
Randomization	No group allocation

Blinding	No group allocation

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Magnetic resonance imaging

Experimental design

Design type	StudyForrest movie dataset: naturalistic movie-viewing task StudyForrest localizer extension dataset: object-category localizer task, block design Sherlock movie dataset: naturalistic movie-viewing task
Design specifications	StudyForrest movie dataset: collected in 8 runs of approximately 15 minutes (2h total) StudyForrest localizer extension dataset: collected in 4 runs of 156 TRs, with 2 16s blocks for each of 6 stimulus category per run Sherlock movie dataset: collected in 2 runs of approximately 24 minutes (48 minutes total)
Behavioral performance measures	N/A

Acquisition

Imaging type(s)	All imaging types were functional MRI.
Field strength	StudyForrest movie dataset: 7T StudyForrest localizer extension dataset: 3T Sherlock movie dataset: 3T
Sequence & imaging parameters	StudyForrest movie dataset: T2*-weighted echo-planar imaging sequence, TE=22ms, TR=2s, whole brain coverage with 36 slices of 1.4 mm thickness and 1.4 x 1.4 mm2 in-plane resolution, 224mm FOV. StudyForrest localizer extension dataset: T2*-weighted echo-planar imaging sequence (TE 30ms, TR 2s, 90-degree flip angle, 35 slices of 3mm thickness and 3x3 mm2 in-plane resolution, 240mm FOV. Sherlock movie dataset: T2*-weighted echo-planar imaging sequence (TE 28ms, TR 1.5s, 64-degree flip angle, whole brain coverage with 27 slices of 4mm thickness and 3x3 mm2 in-plane resolution, 192x192mm2 FOV).
Area of acquisition	Whole-brain scans

Diffusion MRI   ☐ Used   ☒ Not used

Preprocessing

Preprocessing software	Data were preprocessed prior to publication of original access papers. Preprocessing was performed with FSL (http://fsl.fmrib.ox.ac.uk/fsl) for all datasets
Normalization	Slice-time correction, co-registration and affine transformation of functional images to MNI152 template were performed with FSL.
Normalization template	MNI template, 3mm isotropic voxels.
Noise and artifact removal	Linear detrending, motion correction, nuissance regression, high-pass filtering, and spatial smoothing with a 6mm kernel were performed with FSL for the Sherlock dataset prior to publication. Additional preprocessing was performed for the StudyForrest data in Python to match that of the Sherlock data, including linear detrending, high-pass filtering, spatial smoothing with a 6mm kernel, and nuissance regression (6 motion parameters, global signal, white matter, cerebrospinal fluid).

| Volume censoring | N/A |

Statistical modeling & inference

| Model type and settings | Nonlinear manifold learning models, hidden markov models, support vector machines |
| Effect(s) tested | No specific conditions tested |

Specify type of analysis: ☐ Whole brain ☒ ROI-based ☐ Both

| Anatomical location(s) | Posterior medial cortex mask based on the original Sherlock dataset publication as the posterior medial cluster of voxels within the dorsal default mode network.
Early visual, early auditory, and high visual cortex masks were based on a resting-state-defined functional atlas (Shirer et al., 2012) |

| Statistic type for inference
(See Eklund et al. 2016) | N/A |

| Correction | Multiple comparisons correction performed for all pairwise comparisons with the Bonferroni method. |

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

| Multivariate modeling and predictive analysis | Dimensionality reduction methods evaluated at the single-subject level (PCA, PHATE, TPHATE, 2 alternate TPHATE implementations, Locally Linear Embeddings, Isometric Mapping, UMAP, t-SNE). All dimensionalities evaluated are determined based on cross-validation for each analysis. For event segmentation analyses, Hidden Markov Model hyper-parameters are cross-validated across subjects. For support vector classification, model accuracy is evaluated based on KFold cross-validation within-subject. |