Personalized neural state segmentation: validating the GSBS algorithm for individual-level fMRI data

Wilford, R. E.[1], Chen, H.[2], Wharton-Shukster, E.[1], Finn, A. S.[1], Duncan, K.[1]

1. Department of Psychology, University of Toronto, Canada
2. Department of Psychology, University of Chicago, U.S.A.

Author Note

A.S.F. and K.D. are joint senior authors.

We have no conflict of interest to disclose.

Correspondence concerning this article should be addressed to R.E.W., Sydney Smith Hall, 100 St George St Room 4020, Toronto, ON M5S 2E5. Email: robyn.wilford@utoronto.ca

Abstract

Humans segment experience into a nested series of discrete events, separated by neural state transitions that can be identified in fMRI data collected during passive movie viewing. Current neural state segmentation techniques manage the noisiness of fMRI data by modelling groups of participants at once. However, the perception of event boundaries is itself idiosyncratic. As such, we developed a novel denoising pipeline to separate meaningful signal from noise and validated the Greedy State Boundary Search (GSBS) algorithm for use in individual participants. We applied the GSBS to publicly available (1) young adult (YA) and (2) developmental fMRI datasets. After extensive denoising, we confirmed that personalized YA neural state transitions exhibited a canonical temporal cortical hierarchy and were related to normative behavioural boundaries across time in key regions such as the posterior parietal cortex. Further, we used machine learning to show that the strongest neural transitions could be used to predict the timing of normative boundary judgements. Results from the developmental dataset also demonstrated important boundary conditions for estimating personalized neural state transitions. Nonetheless, some brain-behaviour relations were still apparent in individually modelled developmental data. These validations pave the way for applying personalized fMRI modelling to the study of event segmentation; what meaningful insights could we be missing when we average away what makes each of us unique?

Introduction

To manage the deluge of unfolding information in our daily lives, we carve up our continuous experience into meaningful events, a process referred to as event segmentation (Carroll & Bever, 1976; Newtson, 1973; Zacks et al., 2001). This segmentation prioritizes which information we have access to in any given moment (Carroll & Bever, 1976; Ren et al., 2021; Swallow et al., 2009), determines what we attend to (Bailey et al., 2017), and shapes the organization of our memories (Bein & Davachi, 2024; Clewett & Davachi, 2017; Davachi & DuBrow, 2015; DuBrow & Davachi, 2013; Ezzyat & Davachi, 2014; Horner et al., 2016; Shin & DuBrow, 2021; Zacks et al., 2007). Judgments about where to place these boundaries (hereafter, *behavioural event boundaries*) are consistent across individuals (statistically, e.g., Zacks et al., 2001; Zacks et al., 2007), but the emphasis on this consistency belies pervasive individual differences in event segmentation. Even in seminal work, at best, 30% of young adults agree on the timing of each event boundary (Zacks et al., 2001). These individual differences in the timing of behavioural boundaries are stable across retesting (Speer et al., 2003), and they are more variable in older adults (Zacks et al., 2006) and children (Glebkin et al., 2019; Ren et al., 2021; Zheng et al., 2020). People also differ in the duration of their perceived events – varying from many (fine granularity) to few boundaries (coarse granularity; Zacks et al., 2007). Individual differences in this granularity correlate with their memory (Bailey et al., 2017; Jafarpour et al., 2022; Sargent et al., 2013; Zacks et al., 2006). Unfortunately, this meaningful idiosyncrasy in event segmentation cannot be explored in the brain because of tool limitations. State segmentation tools (Baldassano et al., 2017; Geerligs et al., 2021) have shown that patterns of fMRI activity are relatively stable within events, punctuated by sharp shifts (hereafter, *neural state transitions*) that tend to co-occur with boundaries between events (Baldassano et al., 2017; Benear et al., 2023; Cohen et al., 2022; Geerligs et al., 2021, 2022; Lee et al., 2021). However, these tools require users to aggregate participants' fMRI data and, therefore, only consider neural state transitions that are shared across people (Baldassano et al., 2017; Geerligs et al., 2021). Here, we validate the modelling of personalized neural state transitions, providing a critical tool that bring us closer to understanding what makes each of us uniquely carve up our experience into discrete events.

Obtaining personalized estimates of neural state transitions is also important to avoid potential distortions that can arise from averaging inherently idiosyncratic signals. Individual differences in the placement of boundaries can inflate the apparent coarseness of segmentation at the group level. When there is greater mismatch between the timing of the boundaries, this decreases evidence for a group-level boundary, giving the appearance of coarser segmentation in the group. This problem could mislead our thinking about event segmentation in at least two significant ways. The first pertains to development: young children's data tend to be noisier and more variable than their older counterparts (Glebkin et al., 2019; Ren et al., 2021; Zheng et al., 2020). Greater variability in the perception of boundaries could make it seem like segmentation is coarser in younger children and becomes finer with development. Second, distortions caused by aggregating across individuals could also be problematic for the "temporal cortical hierarchy": one of the most discussed aspects of event perception. Group-modelled fMRI data consistently reveals that sensory regions have finer temporal segmentation than association cortex (Baldassano et al., 2017; Cohen et al., 2022; Geerligs et al., 2022; Yates et al., 2022). Given

that sensory cortex activation is strongly yoked to sensory experiences that are shared during movie watching, the timing of its state transitions will likely be more consistent than in the association cortex, where greater idiosyncrasy could inflate the coarseness of transitions. In short, estimating fully personalized neural state transitions is not just a methodological curiosity– a nerdy problem, as it were–it is needed to understand event cognition.

In recognition of this need, researchers have attempted to measure personalized neural state transitions, but with limitations. A notable attempt (Sava-Segal et al., 2023) leveraged the Hidden Markov Modeling (HMM) approach developed by Baldassano et al., (2017). While this method is quite powerful when characterizing groups (Baldassano et al., 2017; Cohen et al., 2022; Yates et al., 2022), it determines the optimal number of states for each brain region with a cross-validation procedure. Consequently, the granularity of segmentation is determined by the group data, once more, obscuring personalized neural state transitions. Even with this limitation, Sava-Segal et al. (2023) highlighted that inter-individual variability in the timing of transitions is pronounced, differs across brain networks, and is related to subsequent memory of movies. Nevertheless, variability in granularity cannot be estimated with this method, running the risk of distorting neural state timescales as described above.

Fortunately, the Greedy State Boundary Search (GSBS) algorithm, recently developed by Linda Geerligs and colleagues (2021), can simultaneously determine the optimal granularity of neural state transitions *and* their timing within individual participants. But, to date, it has been applied to group-averaged brains because validations cautioned against modelling individual participants' data due to noise. Specifically, this validation assessed the impact of averaging fMRI data before performing state segmentation. To do this, they ran the GSBS algorithm on individual brains along with brains averaged across groups of various sizes. They then compared the resulting neural state time courses to one obtained by modelling the average of held-out participants. The correspondence to held-out data dropped steeply as group sizes went down, with very low correspondence between each individual's time course and the group's (Figure 1a). The authors reasonably interpreted the drop-off as an unavoidable consequence of known sources of noise that limit fMRI data quality. Most notably, sudden motion wreaks havoc on fMRI time series by creating precisely the types of discontinuities in activity patterns that the GSBS algorithm is designed to detect. Since participants are likely to move at different times,

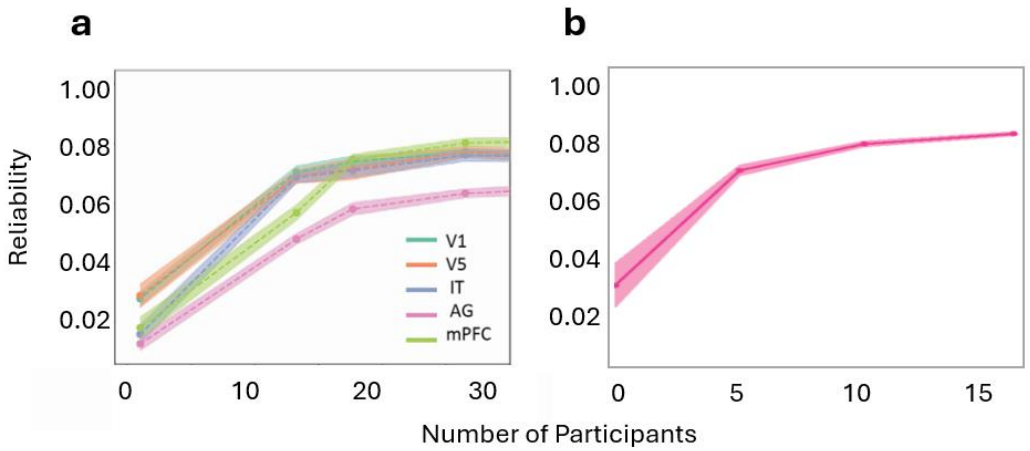

Figure 1. ***Similar reliability curves are found in the brain and behaviour.***
Reliability curve for neural state transitions **(a)** (adapted from Geerligs et al., 2021) demonstrating that, across multiple key regions, neural state transitions look more like the group-modelled transitions the more participants you average across. Reliability curve for behavioural boundary judgements **(b)** (data from Cohen et al., 2022) demonstrating that when behaviourally meaningful idiosyncrasies are averaged, a similar reliability curve is found. V1 and V5 = visual regions; IT = inferior temporal cortex; AG = angular gyrus; mPFC = medial prefrontal cortex.

averaging across enough people's data should reduce the influence of these artifacts while preserving shared state transitions related to event processing.

What this reasoning does not consider, though, is the (likely) possibility that idiosyncrasy in neural event segmentation is meaningful. Consider what happens when the same validation is performed on the time series of behavioural event boundaries (Figure 1b). Since the timing of behavioural boundaries varies, the average of the held-out participants' time series reflects only the most agreed-upon boundaries, and therefore, the more idiosyncratic boundaries are lost. Thus, each individual's time series may have relatively low correspondence with this group average, but as more individuals are added, the resulting average increasingly reflects group consensus. In other words, groups are more like other groups than their constituent individuals. Indeed, when we performed this analysis with a behavioural event segmentation dataset (N = 36), we found a correspondence curve that mirrors that observed in Geerligs et al.'s (2021) GSBS validation (Figure 1b). This raises the hopeful possibility that the apparent poor performance of the GSBS algorithm on individual brains might have, in part, reflected meaningful idiosyncrasy in how brains segment experience.

Accordingly, here, we aim to evaluate the degree to which the apparent unreliability of personalized state transitions reflects meaningful idiosyncrasy vs. fMRI artifacts like those caused by motion. We applied a multi-step approach to answer this question in two open datasets: (1) the Sherlock young adults (hereafter, *YA*) dataset (N = 16; Chen et al., 2017), which has been frequently used in the event segmentation literature (Baldassano et al., 2017; Liu et al., 2022; Song et al., 2021), and (2) a subset of the Child Mind Institute's Healthy Brain Network (hereafter, *developmental dataset*) project (N = 70; Alexander et al., 2017). We used the YA dataset to validate personalized segmentation under typical conditions, whereas we used the developmental dataset to assess boundary conditions in our application of the GSBS, as children produce notoriously high motion profiles (Figure 2).

We first developed an extensive denoising pipeline to minimize motion-related artifacts by integrating multiple recommended approaches (e.g., Ciric et al., 2018; Gratton et al., 2020). Then, we confirmed that our denoising procedure effectively mitigates GSBS's problematic placement of neural state transitions at moments of high motion. Next, we determined whether individually-modelled neural state time series are meaningful. Specifically, we assessed whether individually-modelled data contained patterns emblematic of group-modelled event segmentation, including the temporal cortical hierarchy and relationships to normative behavioural boundaries. To do so, we compared averages of personalized state transition time series to group-modelled data. We reasoned that, if individually-modelled transitions only contain noise, then their average should bear little resemblance to the group-modelled

transitions. By contrast, if the resulting time series reflect meaningful—but idiosyncratic—neural transitions, then their average should capture high-consensus transitions, like group-modelled data. We believe that this validation is a fairer test than in Geerligs et al. (2021) since we do not assume that even perfect estimates of each individual's neural segmentation should match that of the group. By validating this new approach, we pave the way to tackling previously elusive questions about individual differences in neural segmentation.

Methods

Ethics authorization

All data was deidentified and sourced from publicly available repositories. All data storage and analysis procedures fall under protocols approved by the University of Toronto Human Research Ethics Board.

fMRI datasets

Young adult dataset. We used the widely-studied Sherlock dataset (Chen et al., 2019; retrieved from https://gin.g-node.org/ljchang/Sherlock/src/master) to validate the use of GSBS in individual YA participants. We reasoned that, since this dataset was used to first demonstrate the neural segmentation of events (Baldassano et al., 2017), replicating its key findings with individually-modelled data would be a strong test of the approach.

Participants. The retrieved public dataset includes data from 16 of the 22 individuals who originally participated in Chen et al. (2017; 12 male, 10 female; ages 18–26; *M* = 20.8). The original authors excluded six participants based on excessive motion, data loss, and lack of engagement with the task. This dataset has been used frequently in the literature (e.g., Baldassano et al., 2017; Chen et al., 2017; Liu et al., 2022; Song et al., 2021) and therefore, this sample size is consistent with past literature in this area.

Stimulus & task. Participants passively viewed the first 48 minutes of the first episode of BBC's *Sherlock* television show while undergoing functional MRI scanning across two runs (23 and 25 min long). We used the first functional run for our analyses.

Image acquisition. T1 anatomical and T2*-weighted echo planar imaging (EPI) pulse functional data were collected on a Siemens Skyra 3T full-body scanner with a 20-channel head coil. Full imaging details can be found in Chen et al. (2017), but the following are parameters of note for the functional sequence: TR = 1500 ms; TE = 28 ms; flip angle = 64°; number of slices = 27; volumes = 972; slice thickness = 4 mm; in-plane resolution: 3 x 3 mm.

Preprocessing. The retrieved data had been preprocessed using fMRIPrep 20.0.6 (Esteban, Blair, et al., 2019; Esteban, Markiewicz, et al., 2019; RRID:SCR_016216), which is based on Nipype 1.4.2 (Gorgolewski et al., 2011; Gorgolewski et al. 2018; RRID:SCR_002502) using the default parameters. We created our own spike regressors in our custom extensive denoising pipeline (see below, *Methods*, *Extensively denoised pipeline*). We cropped the beginning and end of the functional data according to the specifications by Chen et al. (2019). This cropping included adjustments to account for the timing of the hemodynamic response function (HRF) by shifting the functional data back in time by 4.5 seconds. In addition to these

specifications, we cropped 30 TRs from the beginning of the run to remove strong BOLD responses to the onset of the video. Lastly, we cropped an additional 300 TRs from the end of the first run at a major scene change (final length = 617 TRs) to improve the run time of the GSBS algorithm, which increases exponentially with time series length.

Behavioural boundary ratings. We used the binary behavioural boundaries collected by Baldassano et al. (2017) as our normative measure of behavioural event segmentation for the YA dataset. Four independent (non-scanned) human raters watched the *Sherlock* episode and were asked to indicate when one scene ended, and another began (~10 sec to 3 min long). A binary boundary was placed when at least two out of the four raters agreed on a boundary. This procedure resulted in a normative boundary time series with 54 boundaries coded as "1" and the non-boundary time points coded as "0".

Developmental dataset. Our developmental dataset included a subset of the Child Mind Institute's Healthy Brain Network project (Alexander et al., 2017), retrieved when 775 fMRI datasets were available (data release 2.1). Participants in this dataset were recruited based on perceived clinical concerns; to reduce sources of inter-individual differences beyond age, our validation analyses only included participants whose only clinical diagnosis was anxiety. We chose this clinical subset instead of the typically developing one to hold out the latter for future hypothesis testing. We also required that participants had completed the relevant functional and anatomical scans to be included in our analyses.

Participants. Seventy participants (34 females, 36 males; ages 5-20, *M* = 11.25) had imaging data that met our inclusion criteria; all were included in analyses assessing the impact of motion on GSBS performance. Subsequent analyses included exclusions for high motion (criterion: >= 30% TRs flagged for scrubbing by fMRIPrep, see below), resulting in a final sample of N = 62 (29 females, 33 males; age range 5 yo to 20 yo; *M* = 11.47). This sample size is much larger than the YA dataset, the latter of which has been consistently used in the literature (e.g., Baldassano et al., 2017; Liu et al., 2022; Song et al., 2021). See Figure 2 for a plot of motion profiles across both datasets.

Stimulus & task. The participants included in our current sample watched a 10-minute clip of Despicable Me as part of a larger protocol (see https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/MRI_Protocol.html#Movies for details about the movie protocol).

Image acquisition. T1 anatomical and T2*-weighted EPI functional data were collected on Siemens Prisma and Trio Tim 3T scanners across two sites. The following are parameters of note for the EPI sequence: TR = 800ms; TE = 30ms; number of slices = 60; flip angle = 31°; number of volumes = 750; voxel size = 2.4 mm isotropic; multiband acceleration factor = 6. Full details can be found in Alexander et al. (2017).

Preprocessing. We preprocessed retrieved data using default parameters of *fMRIPrep 20.0.7* (Esteban, Markiewicz, et al., 2018; Esteban, Blair, et al., 2018; RRID:SCR_016216), which is based on Nipype 1.4.2 (Gorgolewski et al., 2011; Gorgolewski et al., 2018; RRID:SCR_002502). Frames that exceeded a threshold of 0.5 mm framewise displacement (FWD) or 1.5 standardized DVARS (stdDVARS) were annotated as motion outliers for the purposes of determining participant inclusion. Like in the YA dataset, we created our own spike regressors in our custom denoising pipeline (see below, *Methods*, *Extensively denoised*

pipeline). We then adjusted the developmental dataset to speed up computational time; since the original TR was 800ms (under half the duration of the YA data), we decided to down-sample the TRs by removing every other TR, resulting in a final time course of 355 TRs.

Behavioural boundary ratings. We used the behavioural boundaries collected by Cohen et al. (2022) as our normative measure of behavioural event segmentation (see paper for details). We took each of the N = 21 participants' boundary judgment time series, smoothed them using a Gaussian filter (R smth function; window = 9.6 seconds), and averaged across the full sample to create a normative boundary confidence time series to compare to the neural state transitions. Finally, we adjusted this normative boundary time series by 5.6 seconds in order to adjust for the timing of the HRF in fMRI data.

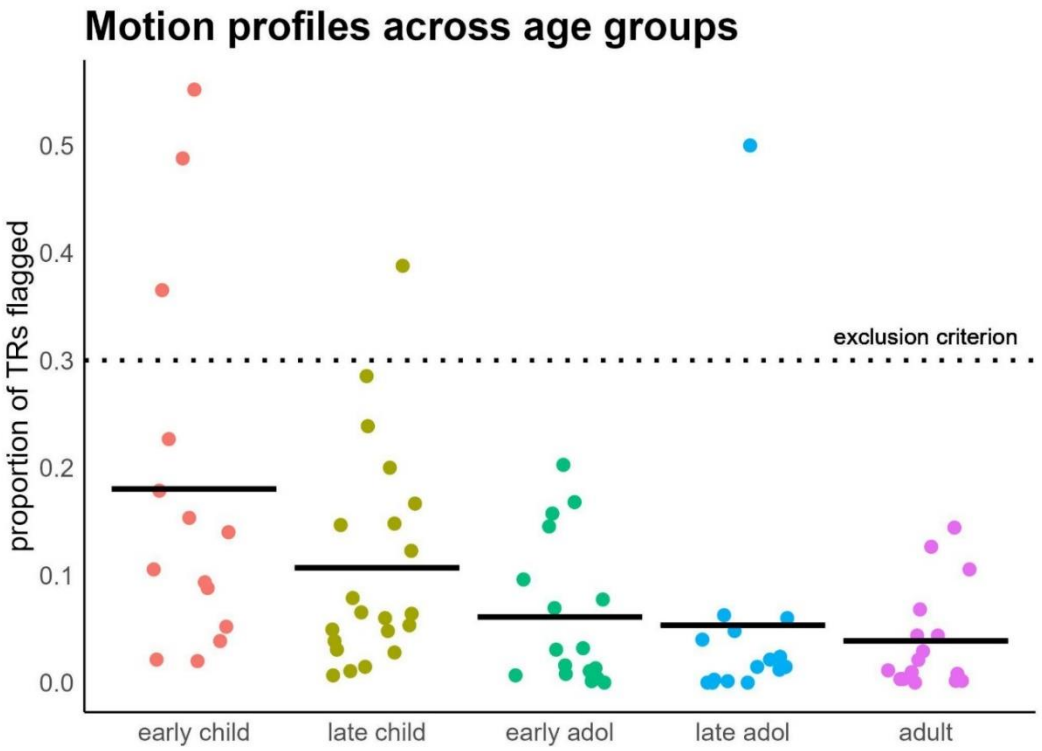

Figure 2. ***Motion Profiles across ages.***
Plot showing the motion profiles of the age groups included in the manuscript. Early child **(red)**, late child **(gold)**, early adolescent **(green)**, and late adolescent **(blue)** groups are from the developmental dataset; the adult group **(purple)** is from the YA dataset. Proportion of TRs flagged was calculated for each participant by dividing their number of motion outliers by the total number of TRs in each respective dataset. Dotted horizontal line denotes exclusion criterion cut-off; dots signify participants, colours denote different age groups.

Custom denoising pipelines

We created two custom denoising pipelines, one minimal and one extensive, to explore the effect of denoising on the GSBS algorithm performance. We used the same pipelines for both the YA and developmental datasets.

Minimally denoised pipeline. For the minimally denoised pipeline, we applied a high-pass filter (Butterworth filter; >0.006Hz) to the preprocessed BOLD data using SciPy's filtfilt function from its signal processing package. This removed scanner drift and any fluctuations in signal greater than the maximum length of events of interest (i.e., events lasting more than ~160s, which was the maximum event length in the behavioural boundaries reported in Baldassano et al., 2017). For the minimally denoised pipeline, we performed no further denoising because (a) this was the extent of denoising in Baldassano et al. (2017), and (b) we wanted to explore the GSBS algorithm's performance in the context of high noise.

Extensively denoised pipeline. For the extensively denoised pipeline, we created a custom pipeline by combining elements of field gold standards (e.g., Grattan et al., 2020; Ciric et al., 2018).

Filtering. Using AFNI's 3dBandpass function, we band-pass filtered (0.22 Hz < 0.006 Hz) both the preprocessed fMRI data and the portion of the confound regressor matrix outputted by fMRIprep that we planned to use in our confound regression (excluding the spike regressors). The high-pass component was the same as in the minimally-denoised pipeline (> 0.006 Hz). We added the low-pass component based on Grattan et al.'s (2020) recommendations for cleaning resting state fMRI data. Low-pass filtering removes high-frequency noise, like that introduced by sudden motion or mechanical issues. Grattan et al. (2020) recommended using a low-pass component with a 0.1 Hz threshold; however, we used a more lenient 0.22 Hz threshold instead because the shortest event judged by raters in the YA dataset was ~10 seconds (i.e., 0.1 Hz). Therefore, if we were to use the recommended low-pass threshold, we would risk removing important high-frequency signals associated with those short events; as such, we allowed more high-frequency signals to pass, i.e., fluctuations greater or equal to ~4.5 seconds (i.e., 0.22Hz). We then used band-pass filtered FWD and stdDVARS regressors to flag TRs as motion outliers for spike regression (FWD >= 0.1 mm, following Grattan et al. (2020); stdDVARS >= 1.5). This procedure yielded fewer flagged TRs and, therefore, resulted in less scrubbing than using unfiltered data to identify outliers, while still effectively removing the noise. Since scrubbing can disrupt the auto-correlative structure of time series by introducing discontinuities, by reducing its necessity, our pipeline is ideal for time series analyses.

Confound regression. We then regressed the band-pass filtered confound time series, including the new spike regressors, from the band-pass filtered fMRI data with AFNI's 3dTproject function. We included the following regressors from the default fMRIPrep pipeline: (a) the first 10 components of aCompcor, (b) cerebral spinal fluid, (c) white matter, (d) global signal, and (e) an 18-expansion of the 6 motion estimates (x, y, z, pitch, yaw, roll). We also included our custom spike regressors. For the developmental data, we also included stdDVARS and FWD as regressors. The residuals from this regression were used for subsequent analyses.

GSBS modelling

Overview. The output from both the minimally and extensively denoised pipelines were both group- and individually-modelled using the GSBS algorithm, leading to four separate types of neural boundary time series for each dataset: (1) individually-modelled, minimally denoised; (2) group-modelled, minimally denoised; (3) individually-modelled, extensively denoised; and (4) group-modelled, extensively denoised. This way, we were able to separately explore the effect of (a) denoising, and (b) group-averaging the data.

GSBS algorithm description. The GSBS algorithm segments multi-voxel data into discrete neural states, where a neural state is defined as a pattern of activity that is relatively stable over time (i.e., that is more similar within vs. between states). The algorithm iteratively runs through k = 2 to k = *n* states, where *n* is the maximum number of allowable states. The optimal segmentation is the one where the difference in between vs. within state similarity is maximized. To model our data, we used the GSBS function from the *statesegmentation* Python package (0.0.4; https://pypi.org/project/statesegmentation/0.0.4/). We used the following parameters for both datasets: max events = 140; finetune = 1; blocksize = 25. We parcellated all functional data using the Schaefer 2018 atlas (7 networks, 100 ROIs; Schaefer et al., 2018) and then applied the algorithm to each ROI using custom Python scripts (version 3.7.16) based on the Naturalist Data Analysis tutorial (Baldassano, Fleetwood, & Geerligs, retrieved Sept 2022).

Group vs. individual modelling. The key difference between group- and individually-modelled results is the fMRI data to which we applied the algorithm. For group-modelled results, GSBS was applied to group-averaged fMRI data for each dataset (i.e., we averaged the fMRI data across participants in each sample, for each voxel). Whereas, for the individually-modelled results, GSBS was applied to each individual participant's fMRI data, and the resulting state transition time series were averaged for comparison with the group-modelled results.
For the developmental dataset, we also group-modelled the data for 4 age groups: early childhood (5-7 yo; N = 14), late childhood (8-10 yo; N = 21), early adolescence (11-14 yo; N = 17), and late adolescence (15-20 yo; N = 15). We chose these four groups because (a) they are consistent with age ranges used in related work (e.g., Cohen et al., 2022) (b) they resulted in relatively similar sample sizes across groups.

Final model output. After determining the optimal segmentation of brain states, *statesegmentation* provides indices, listing the TRs at which transitions between states are most likely, along with the algorithm's evidence for each transition (i.e., its strength; a value between 0 and 2). We used custom Python scripts (version 3.7.16) to transform the model output from transition indices into neural transition time series, in which transitions were coded with their transition strength, and zeros marked all non-transition time points. The resulting outputs were group- and individually-modelled neural transition time series for each ROI in the Schaefer atlas, for each dataset.

Validation analyses

We analysed each combination of pipeline and dataset in R (version 4.3.1) for all subsequent analyses. Since both the YA neural transition strength and binary behavioural boundary time series were sparse and could be slightly offset in time, we applied a Gaussian smoothing kernel (R's smth function; window = 9 seconds). For the developmental dataset, we only smoothed the

neural strength time series since the normative behavioural one was composed of previously-smoothed time series (see above). All brain-wide analyses were visualized using R's ggseg and ggSchaefer packages, and all thresholded maps use false discovery rate (FDR) corrections with an alpha of p = .05.

Motion investigation. We aimed to characterise the effect of motion on GSBS performance and assess how our extensive denoising pipeline mitigates the issue. We used the participants' FWD time series to quantify their movement across time (filtered FWD to compare to the extensively denoised data; unfiltered FWD to compare to the minimally denoised data). For each individually-modelled pipeline (minimally and extensively denoised), we correlated participants' smoothed motion with (a) their smoothed neural transition strength time series and (b) the normative behavioural boundary time series using Kendall's tau rank correlation. We chose Kendall's tau because the sparseness of our time series (even after smoothing) could artifactually inflate a Pearson's correlation. We Fisher Z transformed the correlation values before testing for significant differences between the minimally and extensively denoised pipeline motion-neural transition correlations across the cortex using a repeated-measures T-test. Lastly, we assessed how motion influenced the granularity of neural state transitions, by correlating (Pearson's) the number of neural transitions (averaged across regions) and the number of TRs flagged for high motion across participants. The distributions of outlying TRs were highly skewed (Figure 2), so we transformed them to create more normal distributions (log for YA; square root for developmental data sets). We performed the same transformation on each group's distribution of neural state transitions. Note, since one of our main aims was to characterize the effect of motion on algorithm performance, we used the full developmental sample for these analyses (N = 70), including those participants with greater than 30% of TRs flagged.

Replicating patterns emblematic of event processing. To fairly determine whether individually-modelled state transitions contained meaningful information, we averaged the individually-modelled transitions. Our logic was that if individual fMRI data is too noisy to model, the results would largely reflect this noise, yielding no meaningful patterns at the group-level. On the other hand, if individually-modelled transitions are meaningful (but somewhat idiosyncratic), then their average should contain patterns that are shared across people and mirror group-modelled results.

Temporal hierarchy analyses. We assessed temporal cortical hierarchies by measuring the number of transitions detected in each ROI (i.e., segmentation granularity) for each pipeline. Note that regions with more transitions have shorter state durations. We compared the average number of individually-modelled transitions to the number of group-modelled transitions. Specifically, we used the YA group-modelled map as a template cortical hierarchy and correlated the number of transitions in this map across regions with those of the individually-modelled YA and developmental maps.

Brain-behaviour correspondence. We correlated both the group-modelled and the aggregated individually-modelled neural transitions with the normative behavioural boundaries (for each ROI, and each dataset). We defined three *a priori* ROIs using the Schaefer parcellation: early visual (Vis; formed from LH_Vis_4, LH_Vis_5, RH_Vis_4, and RH_Vis_5

regions), default temporal cortex (DTC; formed from LH_Default_Temp_1, LH_Default_Temp_2, RH_Default_Temp_1, RH_Default_Temp_2, and RH_Default_Temp_3 regions), and posterior parietal cortex (PPC; formed from LH_Default_pCunPCC_1, LH_Default_pCunPCC_2, RH_Default_pCunPCC_1, and RH_Default_pCunPCC_2 regions). We chose these regions because their neural state transitions consistently relate to behavioural boundary timing (e.g., Baldassano et al., 2017, 2022; Geerligs et al., 2021, 2023).

Machine learning analysis. Using custom Python scripts (version 3.11.0), we performed two analyses to determine how well coherent neural transitions across the brain predict normative behavioural boundaries in the YA data. The first involved a combination of data dimensionality reduction and machine learning (ML), and the second simply involved averaging evidence for neural transitions across regions. In both analyses, we defined behavioural boundaries using the unsmoothed Baldassano et al. (2017) norms as described above, and we defined evidence for neural transitions as the individually-modelled neural state transition time series, smoothed with a Gaussian kernel (scipy's gaussian_filter1d function; full-width half-maximum (FWHM) of 3 TRs).

Predictions using machine learning. We first compressed the 100 ROI neural transition time series using a temporal principal component analysis (PCA; sklearn's PCA function, 30 components). This resulted in 30 time series that were the same length as those obtained from individual ROIs to serve as features in the ML model. We further reduced these features to those with the strongest correspondence to behavioural boundaries. Specifically, we ran a logistic regression across participants that predicted behavioural boundaries using the 30 components along with their temporally shifted copies (2 TRs later, to allow for delayed neural responses). We then selected the 10 components that had the strongest relation to behaviour (absolute z-stat value). Lastly, we entered these ten independent sources of neural evidence for behavioural boundaries and their exhaustive interaction terms (generated with sklearn's PolynomialFeatures function) in a cross-validated K Nearest Neighbors (NN) classification analysis (sklearn's KNeighborsClassifier function; neighbours = 120; weights = distance; metric = euclidean; all optimized using sklean's GridSearchCV before model fitting) to derive personalized predictions of perceived event transitions. We chose NN over ridge regression and support vector classifiers because initial testing showed it had the best cross-validated performance.

For each cross-validation fold, we held out one participant while training the model. To avoid bias from having such a small proportion of TRs labelled as behavioural boundaries, we used imblearn's RandomUnderSampler to sample an equal number of behavioural boundary and non-boundary TRs during training. We then tested the model with the withheld participant's neural features to generate a personalized time series of their probability of having experienced an event transition at each TR. We assessed the statistical significance of prediction accuracy by circularly shifting each participant's boundary prediction time series by a random number of TRs 500 times to derive a null distribution of prediction-behavioural boundary correspondence (defined as occurring with 2 TRs of each other).

Predictions using averaging. To assess the added value of ML, we compared the accuracy of the above predictions to those generated by averaging smoothed neural transition evidence across regions to create a single time series of boundary evidence per participant. We

then assessed how often behavioural boundaries occurred within 2 TRs of predicted boundaries (defined as being at or above the 99th, 97.5th, 95th, 90th, 80th, or 50th percentiles of boundary evidence). Like in the machine learning analysis, we determined the statistical significance of prediction accuracy by randomly circularly shifting neural predictions for each participant 500 times. Precision was defined as the number of hits (true positives) plus the number of false alarms (false positives) divided by the total number of positives.

Results

The effects of participant motion on the GSBS algorithm performance.

Participant motion is a major source of noise in fMRI. It creates high-amplitude spikes in the signal observed across the brain, with effects that can linger for several TRs (Power et al., 2014; Satterthwaite et al., 2012; Van Dijk et al., 2012). Since the GSBS algorithm places neural transitions at moments of widespread signal change, we asked if it tends to place transitions at moments of high motion. We then determined if extensive denoising could effectively resolve the issue: a prerequisite to extracting behaviourally meaningful idiosyncrasies in neural event segmentation. Lastly, we characterized motion confounds' impact on estimating individual differences in state transition granularity.

Motion-neural state transition correspondence is mitigated with extensive denoising in adult data. To assess the impact of participant motion on the placement of neural state transitions, we calculated the correlation (Kendall's tau) between each participant's smoothed FWD (hereafter referred to as motion) time series and neural state transition time series. We calculated this correlation for both (a) minimally denoised data (to observe the possible effect of motion) and, (b) the extensively denoised data (to observe the impact of denoising on this relationship; see *Methods* for a full description of the denoising pipelines).

As expected, motion was positively correlated with neural transition placement across much of the cortex in the minimally denoised YA data (Figure 3, top panel; gold = positively correlated, white = not correlated, blue = negatively correlated). This result confirmed that, indeed, the GSBS algorithm places neural transitions at moments of high motion. Promisingly, motion was only weakly correlated with neural transition placement after extensively denoising the YA data, (Figure 3, middle panel). Indeed, extensive denoising significantly reduced the relation in many cortical regions (Figure 3, bottom panel).

However, the story is more complicated for the developmental dataset. Like in the YA dataset, we found that motion and neural transition placement were positively correlated in all four age groups' minimally denoised data (Figure 4a-d, top panels). The denoising protocol effectively reduced the influence of motion on neural transition placement in the early and late adolescent groups (Figure 4c and Figure 4d bottom panels; see middle panels for correlations in extensively denoised data), much like it did in the YA dataset. However, the denoising introduced other artifacts in early and late childhood. Specifically, while the denoising did reduce correlation values in the younger age groups (Figure 4a and 4b, bottom panels), it may have overcorrected for the problem, leaving neural transitions and motion *anti*correlated across much of the cortex (Figure 4a and 4b, middle panels). The issue may be related to younger children's

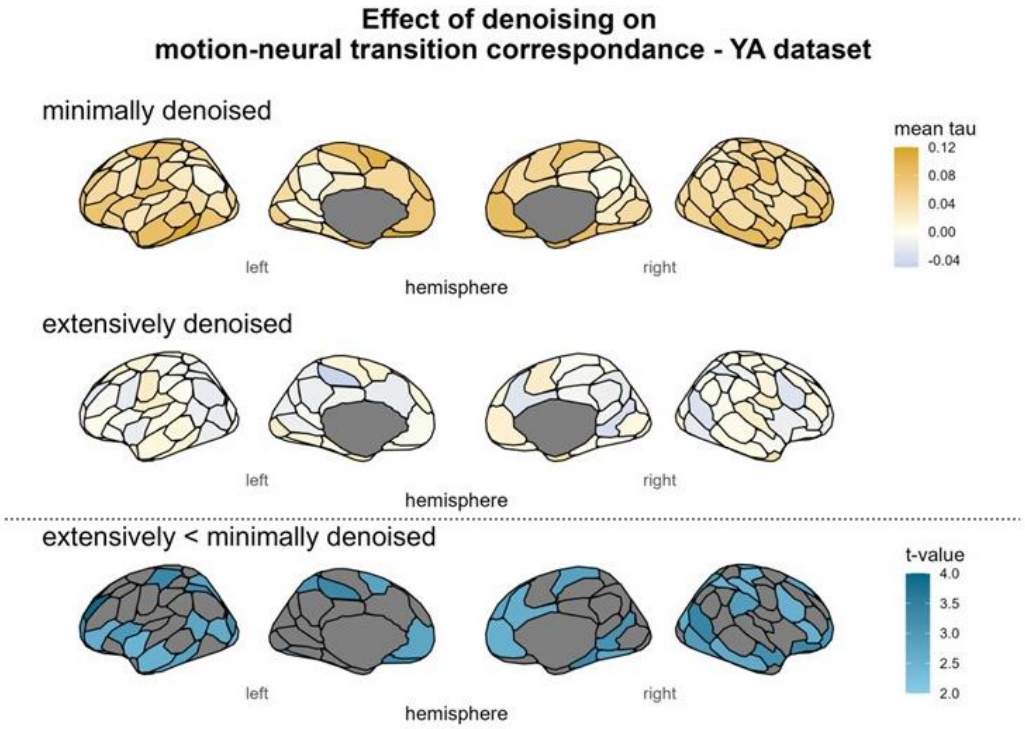

Figure 3. ***Extensive denoising reduces motion artifacts in neural transition placement in the YA dataset.***
Unthresholded average correlation between motion and neural state transitions in minimally denoised data **(top panel)**. Participant motion tended to be positively correlated in time (gold), reflecting the GSBS algorithm's tendency to place neural transitions at moments of participant motion in minimally denoised data. Average correlation between neural transition placement and motion is close to zero (white) across many cortical regions after adequate denoising of the fMRI data (**middle panel**; unthresholded). Motion-neural transition correspondence is significantly lower in extensively denoised data compared to that in the minimally denoised data across many cortical regions (**bottom panel**; FDR corrected; $p_{corrected} < .05$). Note regions that do not show a significant reduction in motion artifacts are often those that tended to have lower correlations with motion in minimally denoised data (top panel), leaving little for the denoising process to change.

motion profiles being too severe (median proportion of TRs flagged for scrubbing: early childhood = 12.27%; late childhood = 6.40%; early adolescence = 3.07%; late adolescence = 1.50%). It is an especially problematic issue because, unlike the YA sample ($\mu \neq 0$, $t(15) = -0.22$, $p = .83$), the developmental sample tended to move at behavioural boundaries ($\mu \neq 0$, $t(66) = 2.25$, $p = .03$). Therefore, our denoising could mask brain-behaviour relations in participants who move at meaningful times. In sum, our denoising pipeline was not effective in

groups with high motion profiles; accordingly, we removed participants with >30% of TRs flagged as outliers for subsequent analyses.

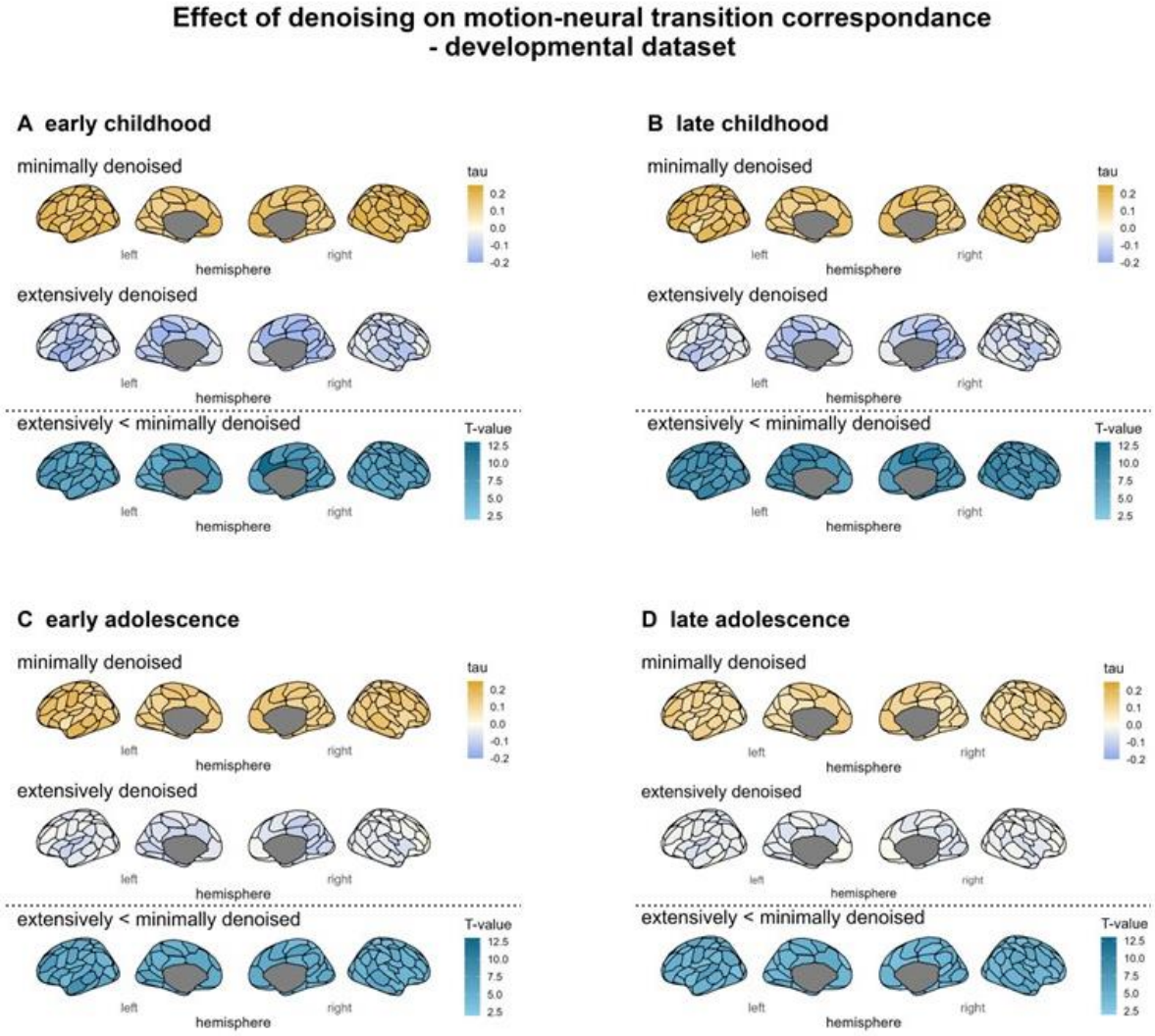

Figure 4. ***Extensive denoising effectively reduces motion's influence in adolescents but overcorrects in children.***
GSBS transition placement is positively correlated (gold) with participant motion in minimally denoised data across all four age groups (**all top panels**), but extensive denoising significantly reduces the correlation values across the cortex (**all bottom panels**). Whereas this relationship is close to zero in early **(c)** and late **(d)** adolescence, the denoising introduces negative correlations across much of the cortex in early **(a)** and late **(b)** childhood (**middle panels**). Top and middle panels are unthresholded; bottom panels are thresholded with FDR correction ($p_{corrected} < .05$).

Motion leads to artifactually coarser neural state transitions. Next, we investigated *how* motion influences neural transition placement. Since motion resulted in artifactual transitions,

we suspected that participants with more high-motion time points would have *more* neural state transitions detected. But we found the opposite pattern in minimally denoised data. GSBS detected *fewer* transitions in participants with higher motion in both the YA ($r(14) = -0.8$, $p < .001$; Figure 5a) and developmental ($r(65) = -0.36$, $p = .003$; Figure 5c) datasets. Perhaps this

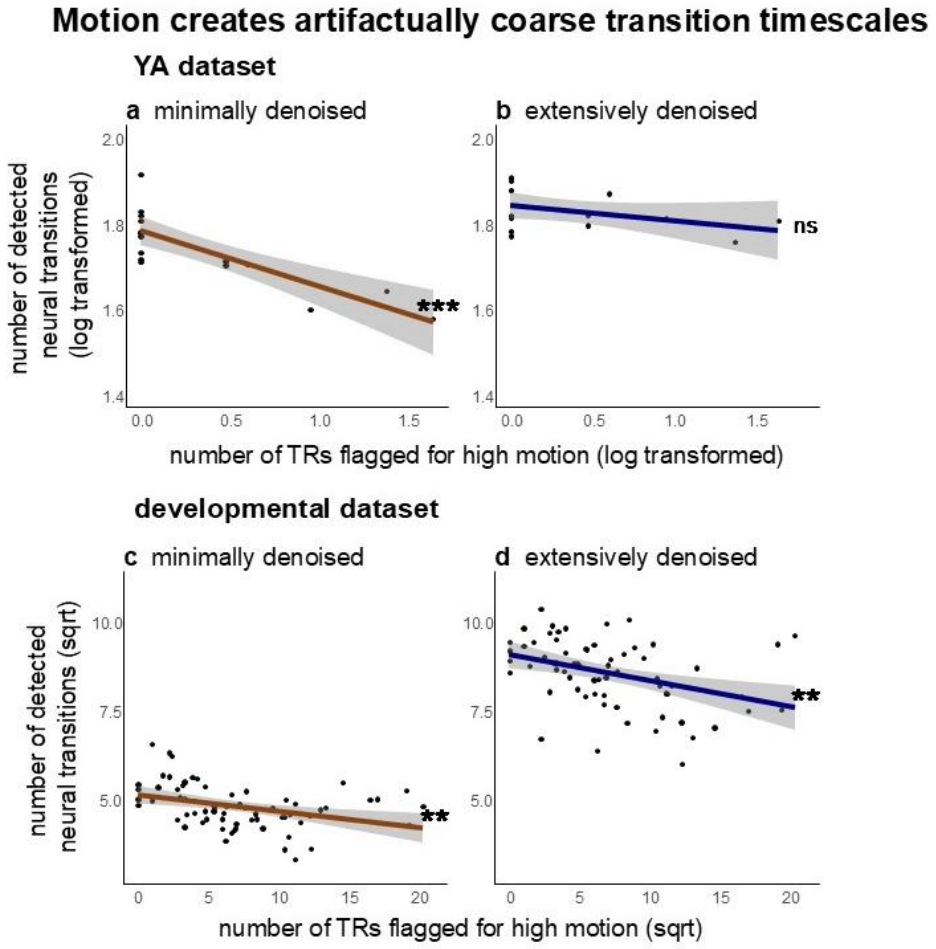

Figure 5. ***Motion drives artifactually fewer neural state transitions.***
In minimally denoised data, the number of TRs flagged for high motion (x-axis) is negatively associated with the number of neural transitions (y-axis) for both YA **(a)** and developmental **(c)** data sets. In the extensively denoised YA data **(b)**, the number of TRs flagged for motion no longer significantly relates to the number of neural state transitions. However, regardless of extensive denoising, the negative relationship between motion and number of transitions remains in the developmental dataset **(d)**. Each dot represents a participant, coloured lines depict the regression line of best fit, and ribbons reflect standard error of the mean. *** = $p < .001$; ** = $p < .01$; ns = not significant.

pattern reflects how motion artifacts overshadowed more subtle state transitions. This effect of motion on segmentation granularity was no longer significant in the YA extensively denoised pipeline ($r(14) = -0.38$, $p = .15$; Figure 5b), further building confidence in our pipeline's effectiveness in dealing with motion confounds. However, motion was still significantly negatively correlated with segmentation granularity in the developmental dataset after extensive denoising ($r(65) = -0.38$, $p = .001$; Figure 5d). We next explored whether this persistent correlation was related to participants' age, since age and motion are tightly linked. Specifically, we ran a mixed effect model (using R's lme4 package; version 4.3.1) to predict the number of neural transitions in the extensively denoised developmental data with participants' age and mean FWD, excluding participants with >30% of TRs flagged as motion outliers. The model also included a random intercept for ROI. We found that motion continued to be negatively related to granularity when controlling for participant age ($b = -8.11$, $SE = 0.53$, $t(6098) = -16.27$, $p < .001$), consistent with the limitations that we've observed in our extensive denoising for high-motion samples. Notably, age was not significantly related to granularity in the model ($b = -0.12$, $SE = 0.14$, $t(6098) = -0.80$, $p = .43$), unlike in Cohen et al. (2022). We caution against the over-interpretation of the null relation, however, since the persistent confound of motion may obscure meaningful relations with age.

In summary, we confirmed that the GSBS algorithm tends to place transitions at moments of high motion, but that our extensive denoising mitigates this artifact in groups with moderate motion profiles. However, our denoising pipeline was less effective in samples with high motion, artifactually decreasing the chance of GSBS detecting neural state transitions during periods of high motion. Further, in higher-motion samples, our denoising pipeline was not able to remove motion's effect on the granularity of neural transitions. However, it removed this effect in the YA data, potentially uncovering more subtle but behaviourally-meaningful neural state transitions.

Meaningful neural state transition structure can be preserved in individually-modelled data.

Are individually-modelled neural state transitions *meaningful*? To answer this question, we assessed whether two patterns that have become emblematic of group-modelled movie-viewing data hold for individually-modelled state transitions: (1) do they share the features of the temporal cortical hierarchy found in Baldassano et al. (2017) and replicated in Geerligs et al. (2021), and (2) are they related to behavioural event boundaries (e.g., Geerligs et al., 2021, 2023)? We contrasted individually-modelled neural state transitions in extensively denoised data (i.e., the data with minimal motion artifacts) to those detected in minimally denoised group-averaged fMRI data (i.e., what is typically done). As a reminder, to draw meaningful comparisons, we averaged the individually-modelled transitions, reasoning that they should contain shared behaviourally-meaningful information.

Temporal Cortical Hierarchies. Baldassano et al. (2017) found a hierarchy in neural transition timescales across the cortex, in which early sensory regions transitioned between states more rapidly than association areas. We observed a similar pattern in the YA dataset when visualizing the granularity of group-modelled neural state transitions (Figure 6a, top

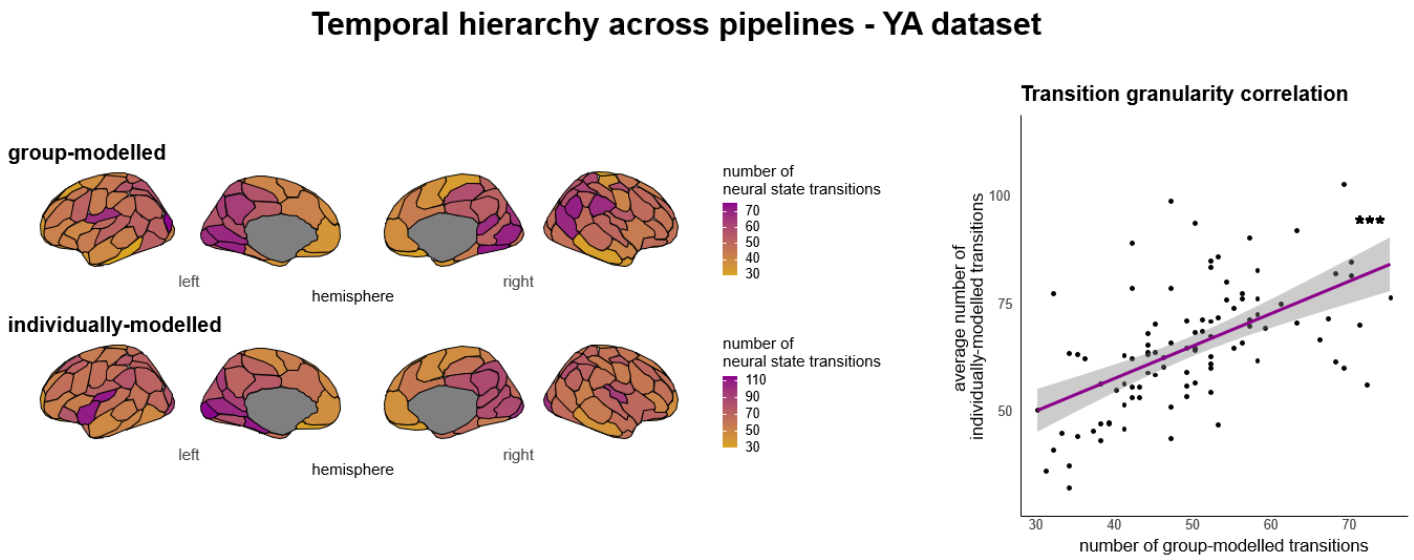

Figure 6. ***The canonical temporal hierarchy is preserved in individually-modelled YA data.***
(left) Top panel displays the map of neural transition granularity for group-modelled data, where purple signifies finer segmentation and gold signifies coarser. There is a posterior to anterior gradient of transition granularity, similar to Baldassano et al.'s (2017) temporal cortical hierarchy. The bottom panel shows individually-modelled transition granularity for extensively denoised data, demonstrating a similar pattern as the group-modelled results. **(right)** Group-modelled transition granularity is strongly positively correlated with the average granularity of individually-modelled data across ROIs, $p < .0001$. The purple line depicts the regression line of best fit, the ribbon standard error of the mean, and dots represent brain regions.

panel). Note how colours shift from purple (fine segmentation) in more posterior regions to more orange (coarse segmentation) in more anterior regions. Critically, we also found this quintessential temporal hierarchy in the YA individually-modelled neural transitions (Figure 6, bottom left panel). The similarity of these two granularity maps was reflected in their strong positive correlation across ROIs ($r(98) = 0.49$, $p < .001$; Figure 6, right). This correspondence demonstrates that individually-modelled neural state transitions also follow a temporal hierarchy structure during movie viewing.

Assessing the temporal cortical hierarchy in the developmental dataset was more challenging. We did not observe a clear temporal cortical hierarchy in group-modelled data when averaging across all ages (Figure 7, top panel) nor when separately averaging within each age group (Figure 7, 3rd panel for late adolescence; other groups not shown). We were, however, able to observe some signatures of the hierarchy when considering individually-modelled transitions across all ages (Figure 7, second panel), and especially when focussing on the late adolescent group (Figure 7, bottom panel; we see numerically more transitions in late adolescents' early visual region (83) and fewer (72) in their default temporal cortex). However, when we correlated individually-modelled maps with the YA group-modelled hierarchy map (our template of the temporal cortical hierarchy), we did not find significant relationships (all-ages correlation: $r(98) = -0.05$, $p = .59$; late adolescent correlation: $r(98) = -0.05$, $p = .59$).

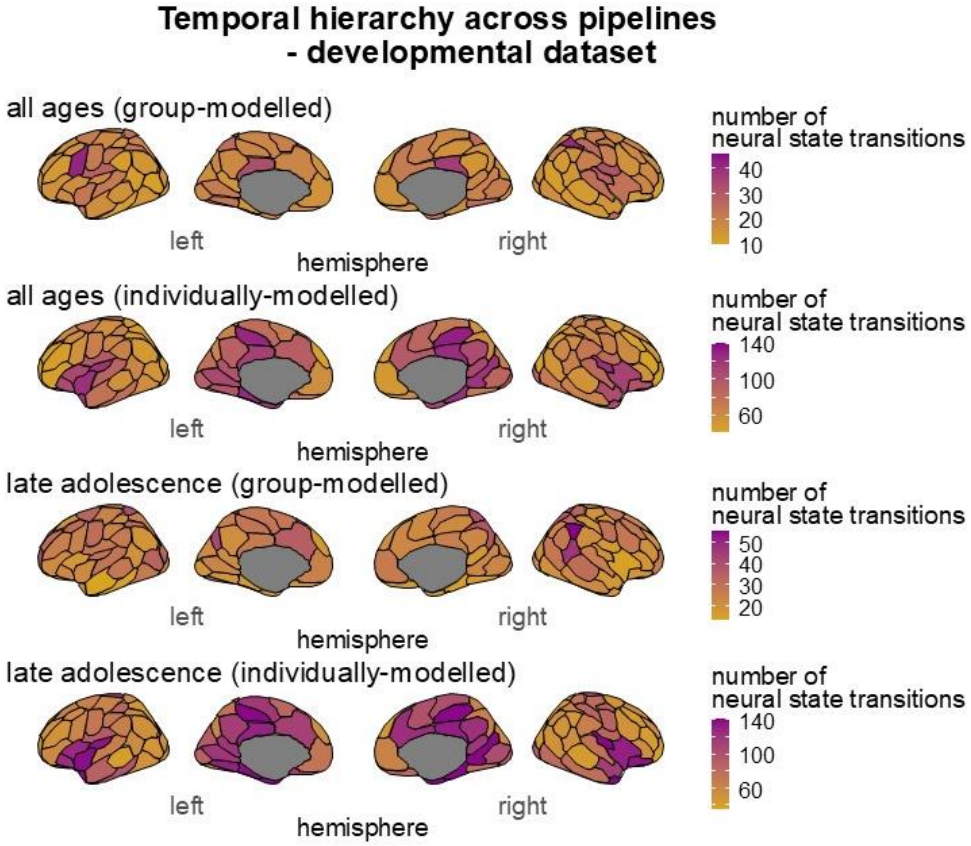

Figure 7. ***Assessing the temporal cortical hierarchy in developmental data.***
There is no evidence for the temporal cortical hierarchy in the group-modelled all-ages sample (i.e., no visible posterior-anterior gradient in transition granularity; purple-gold) **(top panel)**. The granularity of individually-modelled data **(2nd panel)** across all ages is more variable across the cortex than in group-modelled data **(top panel)**, but still does not significantly correlate with the granularity patterns observed in the YA data set. In the late adolescent group-modelled data **(3rd panel)**, there is no evidence for a temporal hierarchy (i.e., map is dominantly gold, indicating coarse granularity). The granularity of late-adolescent individually modelled data **(bottom panel)** is more variable across the cortex than in group-modelled data **(3rd panel)** but is still not correlated with the patterns observed in the YA dataset.

Notably, Cohen et al. (2022) observed evidence for a temporal cortical hierarchy when using a larger portion of the same public data set as we analysed here. There are several possible reasons for the difference in findings. It could be that our subset is either too small to detect this group-level pattern, and/or is different from the superset in relevant ways. In addition, our denoising pipeline may have been too aggressive for the movement profiles of developmental data. Regardless, the above analyses demonstrate that meaningful information about neural state transition timescales can be found in individually-modelled data – especially in YA populations.

Brain-Behaviour Correspondence. Having demonstrated that YA individually-modelled neural state transitions operate within a characteristic temporal hierarchy during movie viewing, we next asked if they also tended to occur at event boundaries. Replicating past research (Baldassano et al., 2017; Geerligs et al., 2023), we found significant correlations between the timing of normative behavioural boundaries and group-modelled neural state transitions across key regions of the cortex (Figure 8a, top panel). Critically, we also found a similar pattern of brain-behaviour relationships in individually-modelled fMRI data (Figure 8a, bottom panel). To more closely compare findings across modelling techniques, we separately tested the brain-behaviour relationship in three ROIs that we selected *a priori* based on the literature (Baldassano et al., 2017; Geerligs et al., 2021): the default temporal cortex (DTC), the posterior parietal cortex (PPC), and early visual cortex (Vis). In YA group-modelled data, all three regions

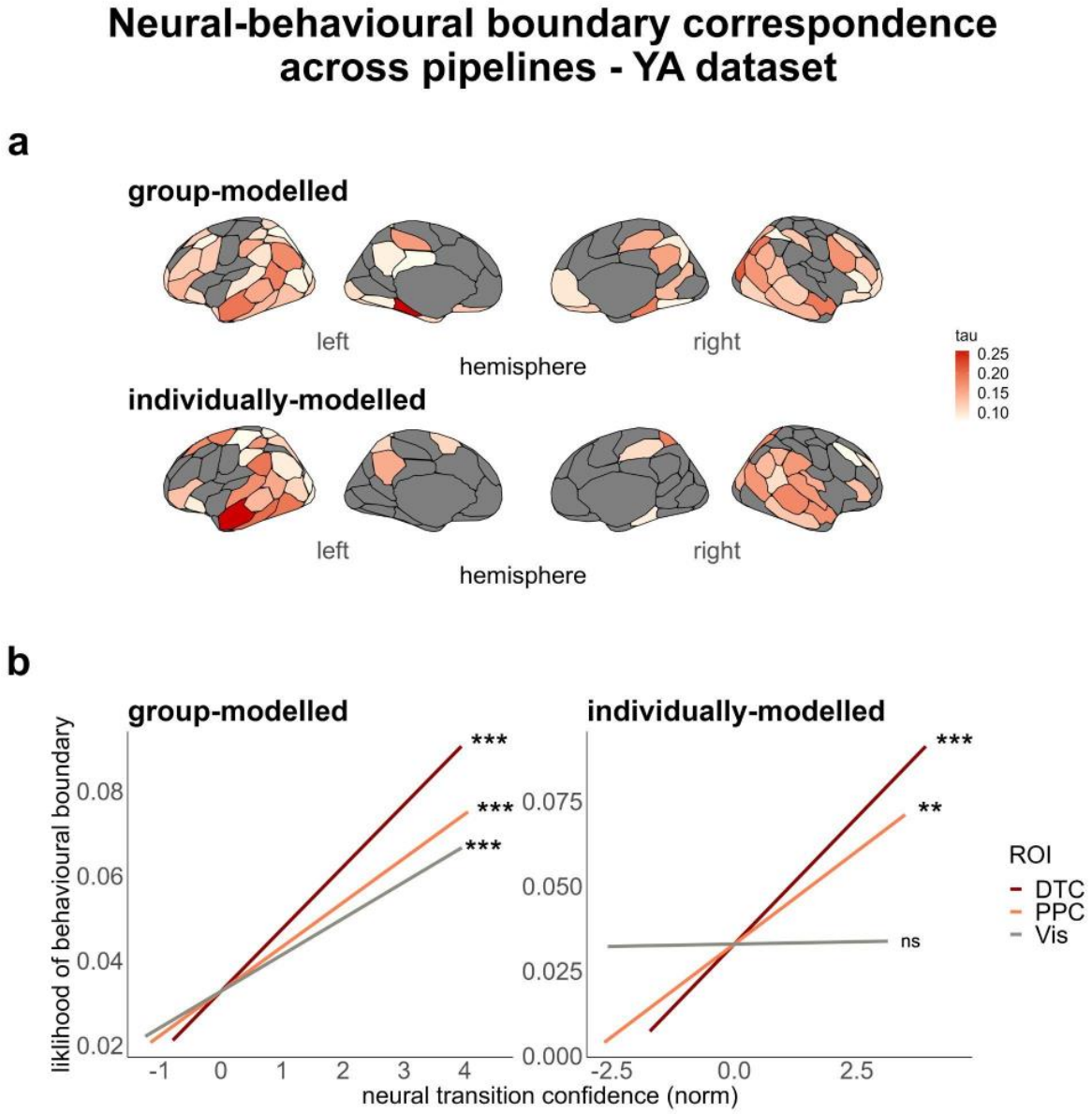

Figure 8. ***Key brain-behaviour relationships are largely preserved in individually-modelled YA data.***
a. Correlation between normative behavioural boundaries and group-modelled neural state transitions (**top panel**; similar to Baldassano et al., 2017), as well as the correlation between normative behavioural boundaries and individually-modelled neural state transitions **(bottom panel)** YA data ($p_{corrected} < .05$).
b. (left) Group-modelled transitions were significantly correlated with behavioural boundaries in three ROIs previously shown to be related to behaviour (DTC, PPC, and Vis); **(right)** individually-modelled transitions were significantly correlated with behaviour in two of the three ROIs (DTC and PPC). DTC = default temporal cortex; PPC = posterior parietal cortex; Vis = early visual cortex. *** = $p < .001$; ** = $p < .01$; * = $p < .05$; ns = not significant.

showed significant correlations between behavioural boundaries and neural transitions (Figure 8b, left; DTC, $\tau = 0.18$, $p < .001$; PPC, $\tau = 0.16$, $p < .001$; Vis, $\tau = 0.11$, $p < .001$). In the YA individually-modelled data, we also found significant brain-behaviour relationships in the DTC ($\tau = 0.16$, $p < .001$), and PPC, ($\tau = 0.09$, $p = .004$), but not in Vis, ($\tau = 0.01$, $p = .70$; Figure 8b, right). In summary, YA individually-modelled state transitions tend to correspond to normative event boundaries, though the relations may be less robust than those observed in group-modelled data.

As with the temporal hierarchy, the brain-behaviour relations in the developmental dataset were less robust. The group-modelled transitions in the developmental sample (Figure 9a, top panel) were significantly correlated with behavioural boundary timing in several regions, including in the bilateral DTC, similar to Cohen et al. (2022). Indeed, our ROI analysis found significant positive correlation in the group-modelled DTC ($\tau = 0.08$, $p = .005$; Figure 9b, left). We replicated this positive correlation in the DTC in our individually-modelled data ($\tau = 0.10$, $p = .01$; Figure 9b, right). However, individually-modelled brain-behaviour relations departed from the group-modelled pattern in notable ways. First, transitions in the PPC ROI were negatively correlated with behavioural boundaries ($\tau = -0.09$, $p = .03$; Figure 9b, right). Our denoising could significant be the culprit: recall, our pipeline created artifactual negative correlations between motion and neural transitions in children (see Figure 4 a-b), and the current sample tended to move around behavioural boundaries. Second, compared to group-modelled data, fewer individually-modelled regions correlated with behavioural boundaries across the cortex (Figure 9a, bottom panel). This weak correspondence may reflect the inclusion of younger participants with challenging motion profiles and still developing event segmentation abilities. Or our low sensitivity to brain-behaviour relationships in the developmental dataset could be due to our smaller sample (N = 62) compared to previous work by Cohen et al. (2022; N = 426).

Machine learning can predict behavioural boundaries by aggregating YA state transition evidence across the brain. Above, we demonstrated that group-modelled brain-behaviour relations are largely preserved, albeit slightly less robust, in individually-modelled YA data. We next asked if combining signals across a participant's brain could more powerfully predict behavioural boundaries. Specifically, we trained machine learning models to predict which TRs corresponded to behavioural boundaries and iteratively tested the generalizability of these models in held-out participants. Critically, this approach results in a single time series per

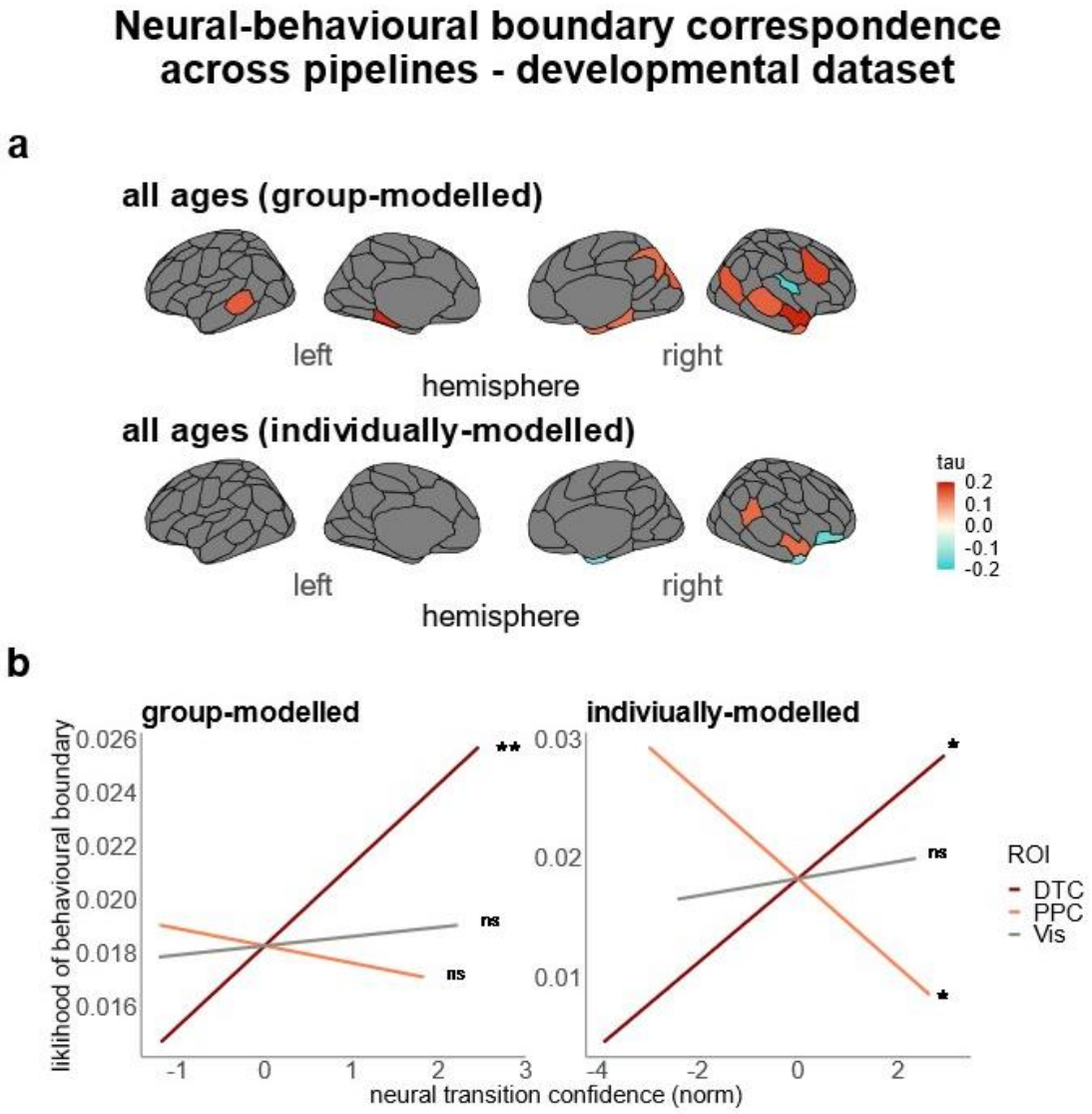

Figure 9. ***Brain-behaviour relationships partially preserved in individually-modelled developmental data.***

a. Correlation (Kendall's tau) between neural state transitions and normative behavioural boundaries for the group-modelled **(top)** and individually modelled **(bottom)** data, across all ages. Red = strong correlation, white = no correlation, blue = negative correlation, FDR corrected at $p_{corrected} < .05$, for all maps.

b. Behavioural boundaries were significantly correlated in the DTC with both the group-modelled **(left)** and individually-modelled **(right)** neural transitions. The individually-modelled PPC neural transitions were negatively correlated with behaviour (right). No other *a priori* ROIs reached significance in either modelling approach (ps > .09). ** = $p < .01$; * = $p < .05$

participant that best predicts their personal perception of a narrative shift. If predictions have high accuracy, they could be used as a personalized proxy for behavioural boundaries.

We ran a K-nearest neighbours classifier on a dimensionality-reduced (PCA) version of the neural transitions to generate personalized predictions of behavioural boundaries (see Methods for modelling details). We found that the model's most confident predictions were also the most precise. For example, if we only labelled TRs as predicted boundaries when they were among that participant's top 1% of evidence, 37% of the predicted TRs corresponded to a normative behavioural boundary, and a further 22% were also predicted to be boundaries in at least three other participants (Figure 10a). As we lowered the confidence threshold, the proportions decreased such that only 44% of predicted boundaries corresponded to behavioural boundaries or other brains. Critically, regardless of the threshold, the model predicted behavioural boundaries across the group better than all random simulations (ps = 0; Figure 10b). We could even significantly predict behavioural boundaries using neural transitions in as many as nine out of 16 of the participants (see Table 1). Lastly, to assess the value added by our machine learning procedures, we used a similar procedure to predict behavioural boundaries based on the average evidence for neural state transitions observed across the brain. This computationally simpler approach did not perform as well as machine learning (Figure 10c and 10d). For example, 34% of predicted TRs (at the 99th percentile of evidence) corresponded to behavioural boundaries and 16% to other participants (Figure 10c). Its predictions for up to seven participants were statistically above chance (Table 1), as was its performance across the group at all confidence thresholds (ps < 0.001; Figure 10d). Together, the accuracy of whole-brain predictions holds promise for their use as proxies for event boundary judgments.

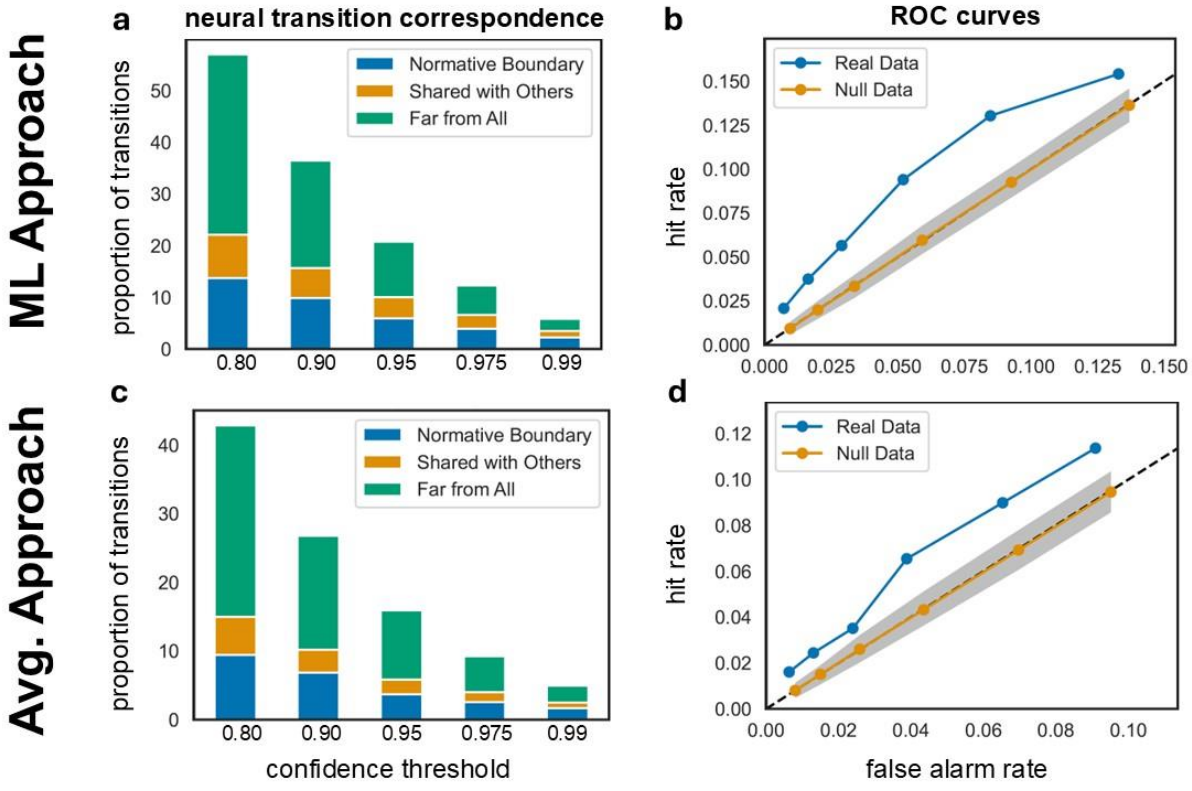

Figure 10. ***Aggregated neural transitions reliably predict behavioural boundaries***
Proportion of predicted behavioural boundaries from ML **(a)** and averaging approaches **(c)** that correspond to meaningful events. Specifically, it plots those that occurred during normative boundaries (blue), at least 3 other participants' neural transitions (orange), or neither (green) when predictions were set at different thresholds of evidence (x-axis). ROC curves for average model performance (**b:** ML; **d:** averaging) plot hit rate (y-axis) against false alarm rate (x-axis) for predictions thresholded at different levels of confidence (same percentiles used as in panels **a** and **c**). Deflections in model performance (blue curve) above the diagonal (identity) and orange (random simulations) lines reflect reliable predictions. The ribbon around the orange line marks the 95th and 5th percentiles of performance in random simulations. Note that model performance always exceeds the 95th percentile of this null distribution.

percentile	number of significant participants (precision)	
	machine learning	averaging
50	4	6
80	9	5
90	7	7
95	7	3
97.5	7	4
99	8	5

Table 1. ***More participants' neural transitions predict behaviour when using ML approach.***
Count of the number of participants whose whole-brain aggregated boundaries significantly predict behavioural boundary timing (using precision as a metric). Boundaries aggregated with our ML approach predicted behaviour numerically more often than those aggregated through averaging across the brain.

Discussion

Prior research has demonstrated that our brains undergo sharp state transitions at boundaries between meaningful events (Baldassano et al., 2017; Geerligs et al., 2021, 2022; Oetringer et al., 2024). But exploration of this exciting discovery has been limited by the available tools. This is because these tools require us to aggregate data across participants to determine neural state transition granularity and timing (Baldassano et al., 2017; Geerligs et al., 2021). Here, we present a method for overcoming this limitation, estimating fully personalized neural state transitions. First, we developed a novel pipeline to extensively denoise the individual participant fMRI data. Then, we applied the GSBS algorithm (Geerligs et al., 2021) to two fMRI datasets collected while participants viewed movies (Alexander et al., 2017; Chen et al., 2017). Personalized neural state transitions in the YA dataset were structured in a temporal cortical hierarchy characteristic of event processing (Baldassano et al., 2017; Geerligs et al., 2021, 2022) and co-occurred with behavioural event boundaries in key regions (e.g., the DTC and PPC). Further, brain-behaviour relations were so robust that the strongest personalized neural

transitions could be used to predict normative behavioural boundaries and even other participants' neural transitions. Results from the developmental dataset provide useful insights into the boundary conditions for personalized GSBS applications. Overall, we extend the GSBS–a critical tool for studying neural segmentation (Geerligs et al., 2021)–to provide fully individualized neural state transition estimates.

Our validation also extends recent work that used the HMM technique to obtain partially personalized boundary estimates (Seva-Segal et al., 2023). Critically, this application of HMM only allowed the timing, but not the granularity, of state transitions to vary across individuals. By using the GSBS algorithm–which can simultaneously fit the timing *and* granularity of neural state transitions–we estimated fully personalized neural state transitions. In so doing, we still found evidence for the temporal cortical hierarchy. This hierarchy has been observed in many group-modelled explorations of neural transition granularity (Baldassano et al., 2017; Geerligs et al., 2021, 2022; Oetringer et al., 2024) but verifying it with personalized estimates was important. As we unpacked in the introduction, the aggregation inherent to group-modelling runs the risk of exaggerating the grain of segmentation because only shared transitions are observable. Indeed, Seva-Segal et al. (2023) found that the timing of state transitions in association cortex is more variable than in sensory regions, raising concerns that group-modelling inflates evidence for cortical hierarchies. Our results quell these concerns since we nonetheless observe the cortical hierarchy at the group level when estimating granularity at the individual level.

Our work also highlights the importance of extensively cleaning motion-related noise when attempting to estimate personalized neural transitions. We explicitly tested the degree to which motion artifacts impact GSBS's detection of neural state transitions, and we characterized *how* participant motion affects neural state transition placement. Specifically, we showed that spikes in the signal created by sudden movement drive the erroneous placement of state transitions, biasing the algorithm to segment the signal at a coarser grain. We speculate that this seemingly paradoxical pattern arises because the GSBS algorithm places transitions where there is the greatest evidence for widespread shifts in the signal. Since the signal shifts caused by motion are so pronounced, they may overshadow many of the more subtle–but meaningful–neural state transitions. Importantly, we showed that our extensive denoising pipeline can effectively deal with this problem in datasets with reasonable motion profiles, uncovering these subtle shifts between brain states. However, our analysis of developmental data also showed the limits of our data-cleaning approach. In particular, while we could reduce the artifactual placement of neural transitions at moments of high motion, higher motion profiles continued to result in courser neural state granularity even after denoising. Further, since the developmental sample tended to move during event boundaries, scrubbing could have influenced the correspondence between neural transitions and behavioural boundaries in this sample. Accordingly, we recommend assessing the motion profiles of your participants before running the algorithm on your sample and setting strict thresholds for participant inclusion. We also recommend that future work combines our denoising protocol with other measures to remove noise, such as ICA denoising of multi-echo data (Steel et al., 2022), or tools to remove physiological sources of noise (Kasper et al., 2017).

On a more theoretical note, our results also show that all neural transitions do not necessarily reflect event boundaries. For example, our machine learning analysis revealed that

while many high-confidence neural state transitions occurred near event boundaries, a sizable portion of these neural transitions did not. Future work could explore whether these "unexplained" transitions might also be tied to event segmentation but reflect more subtle shifts that are the most idiosyncratic. Consistent with this possibility, highly agreed-upon event boundaries solicit stronger and more consistent activity changes across the cortical hierarchy than more idiosyncratically endorsed boundaries (Y. Lee & Chen, 2024). Alternatively, the "unexplained" transitions may have a more endogenous source. For example, spikes in norepinephrine (NE) are thought to trigger sudden shifts in network configurations and disruptions to ongoing processing (Bouret & Sara, 2005). Relatedly, Munn et al. (2021) found that increased activation of the locus coeruleus–the brain's source of norepinephrine–primes the brain for large-scale state shifts during rest. From this perspective, the idiosyncratic state transitions reported here may be instances of more general-purpose neural "motifs" relevant to a variety of cognitive processes (Breakspear, 2017; Munn et al., 2021; Shine et al., 2018; Taylor et al., 2024; Tognoli & Kelso, 2014). Our work lays the groundwork for better uncovering the broader role of neural state transitions in cognition by validating tools that can link state transitions to people's unique mental experiences.

The extraordinary promise of this technique aside, it's important to consider several limits to the application of our validation. First, individually-modelling data using GSBS is computationally time-consuming. In particular, our initial testing showed that processing time exponentially increases with the number of TRs. With that in mind, we recommend only individually modelling data when the question demands it, and to consider using datasets with shorter movies (or cropping at scene changes). Relatedly, we caution against directly comparing the sensitivity of group- and individually-modelled data in our validation as the sensitivity of our group-modelling is limited by our decision not to hyperalign (Haxby et al., 2020) brains before averaging participants' BOLD signals. This advanced registration technique is increasingly used in movie viewing studies to amplify shared signals (e.g., Geerligs et al., 2021; Oetringer et al., 2024). We chose not to use hyperalignment because it might remove idiosyncratic signals, reducing our ability to characterize individual differences in individually-modelled neural state transitions. Since group-modelled data was only used to compare to these individually-modelled results, it would be problematic to use a different registration technique in this pipeline. Additionally, because participants did not actively make segmentation judgments in the scanner, it is hard to validate that differences in neural state transitions really reflect differences in event segmentation. Here, too, the limitation reflects the management of trade-offs; active segmentation engages additional cognitive processes, such as decision-making and motor planning, which could create spurious signals at event boundaries. Indeed, this potential confound is why most related work also uses passive watching paradigms (e.g., Baldassano et al., 2017; Cohen et al., 2022; Geerligs et al., 2022; Yates et al., 2022). Future work could have the same participants segment other videos outside the scanner to test if people who tend toward coarse segmentation judgements also exhibit longer brain state timescales.

These limitations aside, the method we present here aligns nicely with broader practices in cognitive neuroscience. For example, event-related modelling of task-based data is first done at the individual participant level before drawing higher-level inferences, in part to leverage idiosyncratic performance across trials. Indeed, the inability of block designs to incorporate

participants' unique performance timecourses was a major hurdle in fMRI research before the development of event-related methods (Dale, 1999; Huettel, 2012). The practice of averaging fMRI timecourses is even more limited; not only does it prevent researchers from uncovering brain-behaviour relationships within participants, but it also restricts exploring these links across participants. Thus, personalized state transitions could significantly advance our ability to understand event cognition.

To summarise, we find that behaviourally-meaningful personalized neural state transitions can be accurately estimated using the GSBS. This advance is important given our broader understanding of individual differences in humans and their segmentation of events. We know that individual variation matters! This is why our tacit (and occasionally even explicit) practice of aggregating across people when identifying movie-related state transitions is so problematic: it assumes that what is shared is meaningful for event perception, and that which is idiosyncratic is noise. With the presented methods, we no longer have to make these assumptions and can begin to unravel pressing questions for the field. These questions include–but are not limited to–how neural state transitions relate to the perception of events and their representation in memory, the outcome of having more-or-less typical neural segmentation of experience for a person's understanding of and memory for the experience, and possible shifts in neural state segmentation with ageing and development. Beyond these pressing questions for the field, this analytical technique allows us to spotlight the individual, something that matters both biologically and theoretically (Gomez-Marin & Ghazanfar, 2019). Afterall, there is no such thing as an average person. Having a method that now allows us to tease apart idiosyncrasy from noise, enables our exploration of these pressing questions in earnest.

Data Availability

All code and data are available upon request of corresponding author.

Author Contributions

Conceptualization: K.D. and A.S.F.; Data Curation: R.E.W (with help from E.W.-S.); Formal Analysis: R.E.W (with input from E.W.-S. and supervised by K.D and A.S.F.); Funding Acquisition: K.D. and A.S.F.; Investigation: R.E.W. (with input from E.W.-S. and H.C.); Methodology: R.E.W and K.D.; Project Administration: R.E.W. and E.W.-S.; Resources: K.D. and A.S.F.; Software: R.E.W., H.C., and K.D.; Supervision: K.D. and A.S.F.; Validation: R.E.W.; Visualization: R.E.W. (with input from K.D., and A.S.F.); Writing - Original Draft Preparation: R.E.W., K.D., and A.S.F.; Writing - Review & Editing: R.E.W., K.D., A.S.F., E.W.-S., H.C.

Acknowledgements

We would like to thank Dirk Bernhardt-Walther, Alexander Barnett, Michael Mack, and Meg Schlichting for their helpful comments on the analysis approach during department meetings. We would also like to thank the members of the Duncan Lab: Thomas Biba, Catalina Yang, Ariana Youm, and Matthew Dougherty for their helpful feedback on different stages of the project during lab meetings. Finally, we would like to thank Janice Chen and colleagues for

generously publicly sharing the Sherlock dataset and its associated resources, and the Child Mind Institute for making the HBN dataset available.

Funding Information

This research was funded by: NSERC RGPIN-2023-04231, CRC-2019-00107, and CFI/ORF-2014-34479 awarded to K.D.; NSERC RGPIN-2023-05054 awarded to A.S.F; OGS-2022 and NSERC CGS-D-589163-2024 to awarded to R.E.W.

References

Alexander, L. M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega-Potler, N., Langer, N., Alexander, A., Kovacs, M., Litke, S., O'Hagan, B., Andersen, J., Bronstein, B., Bui, A., Bushey, M., Butler, H., Castagna, V., Camacho, N., … Milham, M. P. (2017). An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*, *4*(1), 170181. https://doi.org/10.1038/sdata.2017.181

Bailey, H. R., Kurby, C. A., Sargent, J. Q., & Zacks, J. M. (2017). Attentional focus affects how events are segmented and updated in narrative reading. *Memory & Cognition*, *45*(6), 940–955. https://doi.org/10.3758/s13421-017-0707-2

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, *95*(3), 709-721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Baldassano, C., Fleetwood, G., & Geerligs, L. Event Segmentation. *Naturalistic Data Analysis*. Retrieved September 2022, from https://naturalistic-data.org/content/Event_Segmentation.html

Bein, O., & Davachi, L. (2024). Event Integration and Temporal Differentiation: How Hierarchical Knowledge Emerges in Hippocampal Subfields through Learning. *Journal of Neuroscience*, *44*(10). https://doi.org/10.1523/JNEUROSCI.0627-23.2023

Benear, S. L., Popal, H. S., Zheng, Y., Tanrıverdi, B., Murty, V. P., Perlman, S. B., Olson, I. R., & Newcombe, N. S. (n.d.). Setting boundaries: Development of neural and behavioral

event cognition in early childhood. *Developmental Science*, *n/a*(n/a), e13409. https://doi.org/10.1111/desc.13409

Bouret, S., & Sara, S. J. (2005). Network reset: A simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neurosciences*, *28*(11), 574–582. https://doi.org/10.1016/j.tins.2005.09.002

Breakspear, M. (2017). Dynamic models of large-scale brain activity. *Nature Neuroscience*, *20*(3), 340–352. https://doi.org/10.1038/nn.4497

Brod, G., & Shing, Y. L. (2019). A boon and a bane: Comparing the effects of prior knowledge on memory across the lifespan. *Developmental Psychology*, *55*(6), 1326–1337. https://doi.org/10.1037/dev0000712

Carroll, J. M., & Bever, T. G. (1976). Segmentation in Cinema Perception. *Science*, *191*(4231), 1053–1055. https://doi.org/10.1126/science.1251216

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), Article 1. https://doi.org/10.1038/nn.4450

Chen, J., Leong, Y.C., Honey, C.J., Yong C.H., Norman, K.A., & Hasson, U. (2019). Sherlock. OpenNeuro. [Dataset] doi: 10.18112/openneuro.ds001132.v1.0.0

Ciric, R., Rosen, A. F. G., Erus, G., Cieslak, M., Adebimpe, A., Cook, P. A., Bassett, D. S., Davatzikos, C., Wolf, D. H., & Satterthwaite, T. D. (2018). Mitigating head motion artifact in functional connectivity MRI. *Nature Protocols*, *13*(12), Article 12. https://doi.org/10.1038/s41596-018-0065-y

Clewett, D., & Davachi, L. (2017). The ebb and flow of experience determines the temporal structure of memory. *Current Opinion in Behavioral Sciences*, *17*, 186–193. https://doi.org/10.1016/j.cobeha.2017.08.013

Cohen, S. S., Tottenham, N., & Baldassano, C. (2022). Developmental changes in story-evoked responses in the neocortex and hippocampus. *eLife*, *11*, e69430.

https://doi.org/10.7554/eLife.69430

Davachi, L., & DuBrow, S. (2015). How the hippocampus preserves order: The role of prediction and context. *Trends in Cognitive Sciences*, *19*(2), 92–99. https://doi.org/10.1016/j.tics.2014.12.004

Davidson, D. (1996). The Role of Schemata in Children's Memory. In H. W. Reese (Ed.), *Advances in Child Development and Behavior* (Vol. 26, pp. 35–58). JAI. https://doi.org/10.1016/S0065-2407(08)60505-4

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277–1286. https://doi.org/10.1037/a0034024

Ezzyat, Y., & Davachi, L. (2014). Similarity Breeds Proximity: Pattern Similarity within and across Contexts Is Related to Later Mnemonic Judgments of Temporal Proximity. *Neuron*, *81*(5), 1179–1189. https://doi.org/10.1016/j.neuron.2014.01.042

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Geerligs, L., Gözükara, D., Oetringer, D., Campbell, K. L., van Gerven, M., & Güçlü, U. (2022). A partially nested cortical hierarchy of neural states underlies event segmentation in the human brain. *eLife*, *11*, e77430. https://doi.org/10.7554/eLife.77430

Geerligs, L., van Gerven, M., & Güçlü, U. (2021). Detecting neural state transitions underlying event segmentation. *NeuroImage*, *236*, 118085. https://doi.org/10.1016/j.neuroimage.2021.118085

Glebkin, V. V., Olenina, E. O., & Safronov, N. A. (2019). Event cognition from the perspective of cognitive development. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *41*(0). https://escholarship.org/uc/item/7f2348jp

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., ... &

Perkins, L. N. (2018). Nipype. *Software*. doi: 10.5281/zenodo.596855

Gratton, C., Dworetsky, A., Coalson, R. S., Adeyemo, B., Laumann, T. O., Wig, G. S., Kong, T. S., Gratton, G., Fabiani, M., Barch, D. M., Tranel, D., Miranda-Dominguez, O., Fair, D. A., Dosenbach, N. U. F., Snyder, A. Z., Perlmutter, J. S., Petersen, S. E., & Campbell, M. C. (2020). Removal of high frequency contamination from motion estimates in single-band fMRI saves data without biasing functional connectivity. *NeuroImage*, *217*, 116866. https://doi.org/10.1016/j.neuroimage.2020.116866

Haxby, J. V., Guntupalli, J. S., Nastase, S. A., & Feilong, M. (2020). Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *eLife*, *9*, e56601. https://doi.org/10.7554/eLife.56601

Horner, A. J., Bisby, J. A., Wang, A., Bogus, K., & Burgess, N. (2016). The role of spatial boundaries in shaping long-term event representations. *Cognition*, *154*, 151–164. https://doi.org/10.1016/j.cognition.2016.05.013

Jafarpour, A., Buffalo, E. A., Knight, R. T., & Collins, A. G. E. (2022). Event segmentation reveals working memory forgetting rate. *iScience*, *25*(3). https://doi.org/10.1016/j.isci.2022.103902

Kasper, L., Bollmann, S., Diaconescu, A. O., Hutton, C., Heinzle, J., Iglesias, S., Hauser, T. U., Sebold, M., Manjaly, Z.-M., Pruessmann, K. P., & Stephan, K. E. (2017). The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. *Journal of Neuroscience Methods*, *276*, 56–72. https://doi.org/10.1016/j.jneumeth.2016.10.019

Lee, C. S., Aly, M., & Baldassano, C. (2021). Anticipation of temporally structured events in the brain. *eLife*, *10*, e64972. https://doi.org/10.7554/eLife.64972

Lee, Y., & Chen, J. (2024). The relationship between event boundary strength and pattern shifts across the cortical hierarchy during naturalistic movie-viewing. *bioRxiv*, 2024.04.10.588931. https://doi.org/10.1101/2024.04.10.588931

Liu, W., Shi, Y., Cousins, J. N., Kohn, N., & Fernández, G. (2022). Hippocampal-Medial

Prefrontal Event Segmentation and Integration Contribute to Episodic Memory Formation. *Cerebral Cortex*, *32*(5), 949–969. https://doi.org/10.1093/cercor/bhab258

Munn, B. R., Müller, E. J., Wainstein, G., & Shine, J. M. (2021). The ascending arousal system shapes neural dynamics to mediate awareness of cognitive states. *Nature Communications*, *12*(1), 6016. https://doi.org/10.1038/s41467-021-26268-x

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28–38. https://doi.org/10.1037/h0035584

Oetringer, D., Gözükara, D., Güçlü, U., & Geerligs, L. (2024). *The Neural Basis of Event Segmentation: Stable Features in the Environment are Reflected by Neural States* (p. 2024.01.26.577369). bioRxiv. https://doi.org/10.1101/2024.01.26.577369

Ren, J., Wharton-Shukster, E., Bauer, A., Duncan, K., & Finn, A. S. (2021). Events structure information accessibility less in children than adults. *Cognition*, *217*, 104878. https://doi.org/10.1016/j.cognition.2021.104878

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., Eisenberg, M. L., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241–255. https://doi.org/10.1016/j.cognition.2013.07.002

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex (New York, NY)*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Shin, Y. S., & DuBrow, S. (2021). Structuring Memory Through Inference-Based Event Segmentation. *Topics in Cognitive Science*, *13*(1), 106–127. https://doi.org/10.1111/tops.12505

Shine, J. M., Aburn, M. J., Breakspear, M., & Poldrack, R. A. (2018). The modulation of neural gain facilitates a transition between functional segregation and integration in the brain.

eLife, *7*, e31130. https://doi.org/10.7554/eLife.31130

Song, H., Finn, E. S., & Rosenberg, M. D. (2021). Neural signatures of attentional engagement during narratives and its consequences for event memory. *Proceedings of the National Academy of Sciences*, *118*(33), e2021905118. https://doi.org/10.1073/pnas.2021905118

Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(4), 335–345. https://doi.org/10.3758/CABN.3.4.335

Steel, A., Garcia, B. D., Silson, E. H., & Robertson, C. E. (2022). Evaluating the efficacy of multi-echo ICA denoising on model-based fMRI. *NeuroImage*, *264*, 119723. https://doi.org/10.1016/j.neuroimage.2022.119723

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236–257. https://doi.org/10.1037/a0015631

Taylor, N. L., Whyte, C. J., Munn, B. R., Chang, C., Lizier, J. T., Leopold, D. A., Turchi, J. N., Zaborszky, L., Műller, E. J., & Shine, J. M. (2024). Causal evidence for cholinergic stabilization of attractor landscape dynamics. *Cell Reports*, *43*(6). https://doi.org/10.1016/j.celrep.2024.114359

Tognoli, E., & Kelso, J. A. S. (2014). The Metastable Brain. *Neuron*, *81*(1), 35–48. https://doi.org/10.1016/j.neuron.2013.12.022

Yates, T. S., Skalaban, L. J., Ellis, C. T., Bracher, A. J., Baldassano, C., & Turk-Browne, N. B. (2022). Neural event segmentation of continuous experience in human infants. *Proceedings of the National Academy of Sciences*, *119*(43), e2200257119. https://doi.org/10.1073/pnas.2200257119

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and Aging*, *21*(3), 466–482. https://doi.org/10.1037/0882-7974.21.3.466

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*, 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29–58. https://doi.org/10.1037/0096-3445.130.1.29

Zheng, Y., Zacks, J. M., & Markson, L. (2020). The development of event perception and memory. *Cognitive Development*, *54*, 100848. https://doi.org/10.1016/j.cogdev.2020.100848