Individual variability in neural event segmentation reflects stimulus content and interpretation

Clara Sava-Segal[1], Chandler Richards[2], Megan Leung[1], & Emily S. Finn[1]

[1]Department of Psychological and Brain Sciences

Dartmouth College

Hanover, NH 03748, USA

[2]Section on Functional Imaging Methods, Laboratory of Brain and Cognition

National Institute of Mental Health,

Bethesda, MD, 20892-9663, USA

Corresponding author:

Emily S. Finn

6207 Moore Hall

Department of Psychological and Brain Sciences

Dartmouth College

Hanover, NH 03755, USA

Emily.S.Finn@dartmouth.edu

Abstract

Event segmentation is a spontaneous part of perception, important for processing continuous information and organizing it into memory. While neural and behavioral event segmentation show a degree of inter-subject consistency, meaningful individual variability exists atop these shared patterns. Here we characterized individual differences in the location of neural and behavioral event boundaries across four short movies that varied in both high- and low-level features and evoked variable interpretations. We studied the same individuals across movies to investigate the extent to which individual segmentation styles are stable irrespective of the stimulus (i.e., "trait-like") versus content-dependent. Results showed that across-subject event boundary alignment follows a posterior-to-anterior gradient that is tightly correlated with the rate of segmentation; slower segmenting regions that integrate information over longer time periods show more individual variability. We found that certain aspects of movie content—namely, continuity editing and social content—drive more shared boundaries and are better suited to pull out individual differences. We identified a subset of regions, specific to each movie, in which neural boundary locations are both aligned with behavioral boundaries during encoding and predictive of stimulus interpretation, suggesting that event segmentation may be a mechanism by which narratives generate variable memories and appraisals of stimuli.

Keywords: content effects, event segmentation, fMRI, individual differences, naturalistic stimuli

Introduction

As part of everyday perception, continuous experience is spontaneously discretized into meaningful distinct units or *events* in a process known as event segmentation (Zacks et al., 2007; Zacks & Swallow, 2007). This cognitive mechanism is an automatic and adaptive part of perceptual processing, optimizing attention, and facilitating the subsequent organization of experiences into memory (DuBrow & Davachi, 2016; Kurby & Zacks, 2008; Zacks et al., 2007; Zacks & Swallow, 2007).

Event boundaries are to some extent "normative", or consistent across people. Boundary locations not only have high intersubject reliability, but are also meaningfully structured, corresponding to moments of decreased contextual stability where the prediction of the immediate future fails, such as changes in action, goals, or locations (Newtson et al., 1976; Raccah et al., 2022; Speer et al., 2009; Zacks et al., 2009). Typically, these judgments are made behaviorally, by pressing a button to indicate transitions between events (Newtson, 1973). Neuroimaging studies have identified evoked neural responses or latent brain-state changes at or around event boundaries even during passive viewing (i.e., in the absence of task demands). In particular, neural responses to normative event boundaries (i.e., those defined behaviorally in another viewing or imposed experimentally) have been reported in several regions including the hippocampus, lateral frontal cortex, medial posterior cortex (including the precuneus and the cingulate), and lateral posterior cortex (including the intraparietal sulcus, the extrastriate motion complex, angular gyrus, and the superior and inferior temporal sulci) (Baldassano et al., 2017; Ben-Yakov & Henson, 2018; DuBrow & Davachi, 2013, 2016; Ezzyat & Davachi, 2014; Speer et al., 2003, 2007; Zacks et al., 2001)

Taken together, these findings suggest that segmentation is a naturally ongoing mechanism with shared behavioral and neural correlates. However, meaningful individual differences are preserved above these shared patterns. Individual differences in behavioral segmentation patterns are stable over time (Speer et al., 2009) and correlate with age and other cognitive abilities such as working memory capacity, long-term memory retrieval, and performance on other tasks (Bailey et al., 2017; Jafarpour et al., 2022;

Sargent et al., 2013; Zacks et al., 2006). Further, disrupted segmentation is seen in certain clinical conditions such as schizophrenia (Zalla et al., 2004).

Despite these demonstrated differences in behavioral event segmentation, however, investigations into individual differences in neural event segmentation have been very limited. This is due in part to the challenge of capturing individual boundaries (as opposed to relying on normative boundaries or using stimuli with pre-defined boundaries) while also preserving natural viewing (i.e., not having subjects segment during encoding, or segment upon a second, biased viewing). Taking an individual differences approach to neural data can extend our knowledge of the functional role of segmentation as demonstrated by past work showing that individual neural activity at normative event boundaries predicts behavior (Ezzyat & Davachi, 2011). It remains unknown how the individual brain passively segments ongoing, continuous information and if and how idiosyncratic neural segmentation patterns relate to the subsequent recall and appraisal of that information. Further, although individual behavioral segmentation patterns are stable within a single stimulus (Speer et al., 2009), the extent to which either behavioral or neural segmentation patterns are stable across stimuli, versus vary with stimulus content, remains unknown.

Here we characterized how neural and behavioral event segmentation patterns vary with individual-level factors, stimulus content, and interactions between the two. Recently developed algorithms for identifying latent brain state changes (Baldassano et al., 2017; Geerligs et al., 2021) allow us to identify neural event boundaries from fMRI data acquired during passively viewed continuous "naturalistic" stimuli (ex: movies) without prior knowledge of boundary locations. We harnessed one of these algorithms (Baldassano et al., 2017) to investigate neural event boundaries within the same individuals scanned across multiple different movies. This design allowed us to discern the extent to which individual neural event segmentation patterns are stable across movies or are content- (i.e., stimulus-) dependent.

To preview our findings, we first demonstrated that our chosen algorithm for automatically inferring neural event boundaries can operate reliably at the individual level. Using individually defined

boundaries, we found that across-subject variability followed a posterior-to-anterior cortical gradient: sensory processing regions were most consistent across individuals, while higher-order association regions were most variable. We identified a strong relationship between stimulus content and segmentation, such that certain stimulus features—namely, continuity editing and social content—drive more shared segmentation in regions that have been previously functionally implicated to respond to certain content. We found a relationship between individual event boundaries and the ultimate recall and appraisal of narratives in a subset of regions that also showed normative alignment with behavioral boundaries in our movies. The specific regions showing these relationships were different across movies, suggesting that depending on stimulus content, distinct regions support the conscious chunking of a stimulus that impacts how it is later converted into memory.

Materials and Methods

Experiment 1 (main fMRI experiment).

We recruited and scanned a total of 48 subjects (all native English speakers; 27F, median age-24.5, range= (19,64)) at the National Institutes of Health (NIH). All subjects provided informed written consent prior to the start of the study in accordance with experimental procedures approved by the Institutional Review Board of the NIH. We discarded incomplete datasets without all four movies (described below), leaving 43 subjects whose data we analyzed here. Subjects were compensated $50 per hour for their time.

Subjects watched four movies (ranging from 7:27-12:27 min each) in a pseudo-random order while we collected fMRI data using a 3T Siemens Prisma scanner (TR=1 sec) with a 64-channel head coil. Functional images were acquired using a T2*-weighted multiband, multi-echo echo-planar imaging (EPI) pulse sequence with the following parameters: TR = 1000 ms, echo times (TE) = [13.6, 31.86, 50.12 ms], flip angle = 60 deg, field of view = 216 x 216 mm, in-plane resolution = 3.0mm^2, slice thickness 3.0mm, number of slices = 52 (whole-brain coverage), multiband acceleration factor = 4). Anatomical images were acquired using a T1-weighted MPRAGE pulse sequence with the following parameters: TR = 2530 ms, TE = 3.30 ms, flip angle = 7 deg, field of view = 256 x 256 mm, in-plane resolution = 1.0 mm^2, slice thickness = 1.0 mm).

The movies were projected onto a rear-projection screen located in the magnet bore and viewed with an angled mirror. The experiment was presented using PsychoPy (Peirce et al., 2019). Following each film (i.e., while still in the scanner) subjects completed a task battery designed to probe their interpretations and reactions to the film, including the following: 1) a three-minute free recall/appraisal task in which subjects spoke freely about their memories and impressions of the film, during which their speech was captured with a noise-canceling microphone; 2) multiple-choice comprehension questions designed to ensure they had been paying attention; 3) multiple-choice and Likert-style items assessing

reactions to various characters and to the film overall (**Fig. 1A**). See section "Measuring similarity in the recall" for the experimental instructions for the free recall/appraisal task.

During a separate behavioral visit, subjects completed a battery of tasks and questionnaires including selected instruments from the NIH Cognition and Emotion toolboxes as well as other psychological and psychiatric self-report scales. These data are not analyzed or reported here.

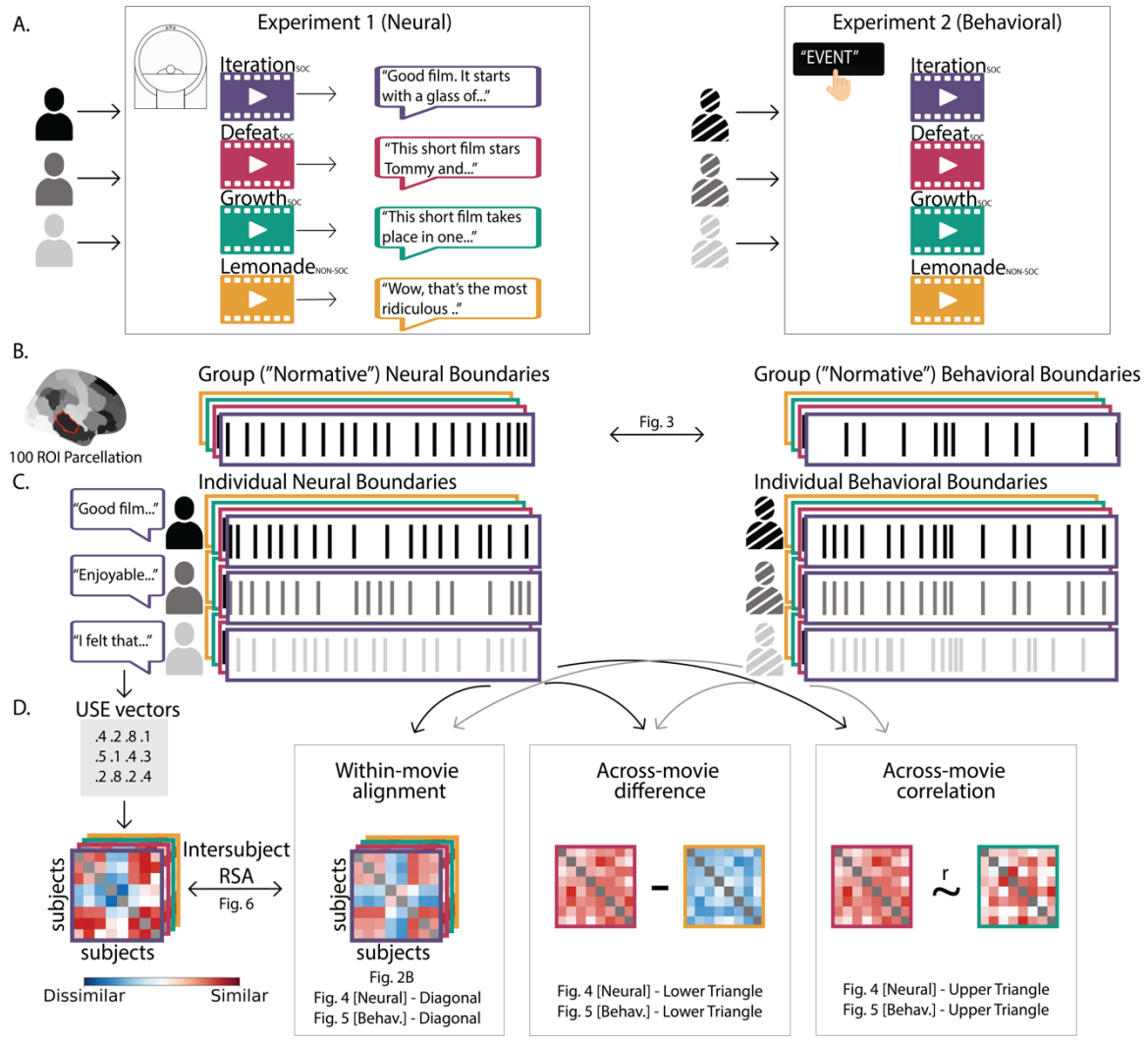

Figure 1 - Experimental protocol and methodological schematic. A. Schematic of the structure of experiments 1 and 2. Participants (n=43, n=40) came to the laboratory and either participated in an fMRI or behavioral experiment protocol. Subjects watched four movie stimuli. In experiment 1(fMRI), this was a passive task. In experiment 2, subjects were asked to report event boundaries while watching the

movies. After each movie, subjects engaged in a series of questions including a free recall and appraisal. **B.** Shared boundaries were determined at the group level ("normative") both neurally (using a Hidden-Markov Model approach - See *Methods*) and behaviorally. The locations of neural and behavioral boundaries were compared (See Fig. 3 for Results). **C.** Individual boundaries in the neural data or in the behavioral data were procured, and then a series of analyses were conducted (see D). **D.** The extent to which subjects were similar in their (neural and behavioral) event boundaries for each movie was determined ("Within-movie alignment", Fig. 2B, diagonals Fig. 4 and 5). Further, comparisons were made to determine whether subjects were aligned more in one movie as compared to the other ("Across-movie difference", lower triangles Fig. 4 and 5) or whether subjects were aligned across movies (indicating "trait-like" segmentation; "Across-movie correlation", upper triangles Fig. 4 and 5). We also looked at the relationship between recall and neural event boundaries (Fig. 6). The similarity in recall data was computed using Google's Universal Sentence Encoder (USE; Cer et al., 2018).

MRI data preprocessing.

Following the conversion of the original DICOM images to NIFTI format, we used AFNI (Cox, 1996) to preprocess MRI data. Preprocessing included the following steps: despiking, head motion correction, affine alignment with subject-specific anatomical (T1-weighted) image, nonlinear alignment to a group MNI template (MNI152_2009), combination of data across the three echoes using AFNI's "optimally combine" method, and smoothing with an isotropic full-width half-maximum of 4 mm. Each subject's six motion time series, their derivatives, and linear polynomial baselines for each of the functional runs were included as regressors. All analyses were conducted in volume space and projected to the surface for visualization purposes.

We used mean framewise displacement (MFD), a per-subject summary metric, to assess the amount of head motion in the sample. MFD was overall relatively low ($Iteration_{SOC}$ - mean=0.08, s.d. = 0.03, range = (0.03, 0.18), $Defeat_{SOC}$- mean=0.08, s.d. = 0.03, range = (0.04, 0.19), $Growth_{SOC}$ -

mean=0.07, s.d. = 0.03, range = (0.04, 0.17), $Lemonade_{NON\text{-}SOC}$ - mean=0.08, s.d. = 0.03, range = (0.04, 0.17) and did not differ across movies (repeated-measures ANOVA, F(3,126) = 1.44, p=.23).

We implemented a shared response model (P.-H. (Cameron) Chen et al., 2015) in *BrainIAK* (Kumar et al., 2021) and applied it to account for different functional topographies across individuals. This step was performed for each of the four movies separately. First, we fit a model to capture the reliable whole-brain responses to the movie across subjects in a lower dimensional feature space (features = 50). We then applied this model to reconstruct the individual voxelwise time courses for each participant. This procedure serves as an additional denoising step and makes spatiotemporal patterns more consistent across subjects.

Experiment 2 (auxiliary behavioral experiment).

We collected an auxiliary behavioral dataset at Dartmouth College to further refine and support results from *Experiment 1*. 44 subjects (21F, median age 20, range=18-32) were presented with the same paradigm as in *Experiment 1,* except that while watching each film, individuals were instructed to press a button each time they perceived that a new scene is starting (i.e., points in the movie when there is a "major change in topic, location, time, etc."; **Fig. 1A**). We used the same instructions used in Baldassano *et al.,* (2017). We discarded subjects who did not complete all four movies, leaving n=40 for our analysis. The paradigm was hosted and presented using JsPsych (de Leeuw, 2015). Subjects were compensated $15 an hour for their time or given participation credit, and the study was approved by the Institutional Review Board of Dartmouth College.

Movie overview.

The movies were four short films made by independent filmmakers that were chosen because they were rich and engaging, yet they depicted ambiguous scenarios that provoked divergent reactions and interpretations in different individuals. Three of the movies were social in nature and followed

different narratives of humans taking actions and interacting, while the fourth depicted purely mechanical information (a long, complex Rube Goldberg machine that traversed a house). We conducted a large-scale online experiment via Amazon Mechanical Turk (MTurk) to determine the optimal stimuli to use to draw out the most individual variability in both neural and behavioral responses and found that these movies did, in fact, cause individuals to diverge in their interpretations and affective reactions. We chose independent films so that subjects would be less likely to have experienced the material before. In a debriefing questionnaire, three subjects reported having seen one of the movies (*Growth*$_{SOC}$) prior to the experiment.

Here we provide brief descriptions of each film along with YouTube links. (Note that the versions presented to subjects were edited to remove credits and title pages; these edited versions are available upon request.) *Iteration*$_{SOC}$ (https://youtu.be/c53fGdK84rc; 12:27 min:sec) is a sci-fi film that follows a female character as she goes through multiple iterations of waking up and trying to escape a facility. A male character appears towards the end to help her. *Defeat*$_{SOC}$ (https://youtu.be/6yN9VH_4GSQ; 7:57 min:sec) follows a family of three (mother, two children) as the brother bullies his sister and she builds a time machine to go back and get revenge. *Growth*$_{SOC}$ (https://youtu.be/JyvFXBA3O8o; 8:27 min:sec) follows a family of four (mother, father, two brothers) as the children grow up and eventually move out amid some family conflict. *Lemonade*$_{NON\text{-}SOC}$ (https://youtu.be/Av07QiqmsoA; 7:27 min:sec) is a Rube-Goldberg machine consisting of a series of objects that move throughout a house and ends in the pouring of a cup of lemonade. This film was lightly edited to remove fleeting shots of human characters. Screen cuts were used and indicated transitions between events in *Iteration*$_{SOC}$ and *Defeat*$_{SOC}$, while *Growth*$_{SOC}$ and *Lemonade*$_{NON\text{-}SOC}$ were shot in a continuous fashion - the camera just panned from one scene to the next.

Automatic event boundary detection.

To automatically identify event boundaries from fMRI data, we fit a Hidden Markov Model (HMM) (split-merge option; Baldassano et al., 2017) as implemented in the *BrainIAK* toolbox (Kumar et

al., 2020), adapting code made available on the *Naturalistic Data* tutorial website (L. Chang et al., 2020). The HMM does not rely on annotations or hand-demarcated events, but rather infers event boundaries from shifts in stable patterns of brain activity. It relies on voxelwise patterns *within* regions. We restricted our analyses to the neocortex and used the 100-parcel, 7-network Schaefer parcellation (Schaefer et al., 2018), in keeping with using the HMM approach to identify event boundaries (Cohen & Baldassano, 2021).

The HMM approach assumes that each event has a distinct signature of activity that shifts at event boundaries. Specifically, the model assumes that (1) each subject starts in an event and then each forthcoming timepoint is either in the same event (state) or in the next state, and (2) that the voxelwise pattern of activity in a region is correlated across timepoints within the same event. The model identifies both the optimal number of events and then the transitions between these events. We were ultimately interested in individual variability in the location of event boundaries.

To focus on variability in boundary locations, we first determined the optimal number of events (k) for each region at the group level (See "Determining the number of events and deriving individual event boundaries"). For each region, we then fit an HMM to each subject's individual timeseries using this fixed value of k to identify the location of implicit event boundaries in each subject in that region. While we acknowledge that there is likely interesting individual variability in the *number* of events (i.e., segmentation rate) in each region, this is a somewhat different question than variability in the *location of event boundaries*: consider that even if two individuals have the same number of events in a region, the specific locations of their event boundaries could still be completely different. The location of boundaries is more directly related to movie content than the overall number of events, and therefore of greater scientific interest to us here. In addition to this theoretical justification, there are two related methodological justifications to fixing the number of events within a region while allowing locations to vary across subjects. First, we are using an inherently noisy method to infer neural event boundaries which was created to work well at the group level. Therefore, by fixing the number of events using group-average data, we constrain the individual HMM solutions to a reasonable number of events. Second, it is

not clear how to reliably estimate the optimal number of events at the individual level, since training and test data need to be independent.

Determining the number of events and deriving individual event boundaries.

The number of events (*k*) per region per movie was defined at the group level using a train-test split procedure (with different subjects in the train and test groups). We used a fairly liberal range of possible values for *k:* the minimum allowed *k* was 6, and the maximum (90-150) differed slightly across movies owing to their different lengths, but always reflected a minimum average event length of five seconds. For each film separately, we split the subjects into train/test groups (each with 21/22 subjects). We averaged across subjects for both the train and test data, and fit the model to the training subject data with each *k* in the allocated range. We then tested the model on the held-out subjects and took the log-likelihood of *k* events. Having repeated this procedure for each *k,* we took the *k* value with the maximum log-likelihood as the optimal number of events within this ROI for this train/test split. We repeated this procedure 100 times with different train/test splits within each ROI. For each ROI, we then calculated the median optimal *k* value across these 100 iterations and used this as the *k* value for our subsequent analyses.

We fit separate HMMs to individual-subject fMRI data using this fixed value of *k* for each region, yielding individual subject boundary locations. Therefore, we had one set of boundary locations for each region for each subject for each movie.

Computing alignment across subjects.

We used permutation testing to quantitatively assess the consistency of boundary locations across individuals. For each movie, region, and pair of subjects, we permuted boundary locations n = 1000 times to derive a z-score for the match between true boundaries relative to the null distribution. In both the true and permuted data, boundaries that were within 3 TRs of one another were counted as a match, consistent with past work (Baldassano et al., 2017). We then created subject-by-subject matrices of these z-scores in every region. Importantly, depending on which member of a pair of subjects was permuted (versus

treated as the "ground truth") there were slight differences in the resulting z-score; we took the mean of the upper and lower triangle of these matrices for relevant subsequent analyses.

Human annotation of event type.

To identify stimulus properties that seemed to trigger event boundaries, three independent raters were given the normative behavioral boundaries from *Experiment 2* and asked to describe what was happening in the film at those moments. Based on these descriptions, five event-type categories were generated with an eye toward what past work has shown drives event boundaries—i.e., prediction errors as a result of situational changes (e.g., changes in location, or character actions) or editing techniques (e.g., screen cuts). Situational changes were split into changes in "location" and changes in "human-driven action" or changes in "object-driven action." Often, event boundaries occurred at moments when a screen cut ($Iteration_{SOC}$ and $Defeat_{SOC}$) or a location change ($Growth_{SOC}$ and $Lemonade_{NON\text{-}SOC}$) was bookended by two object-driven or human-driven actions, such that the screen cut or location change served as an indication to the viewer that a new action (event) was starting. These events were marked by both their bookended action and by their visual cue (location or screen cut). We also introduced "iteration" as a type of event for $Iteration_{SOC}$, since these were meaningful movie-specific events that encompassed both screen cuts and audio-visual cues. At these moments, the character would often hear a count-down, fall, and wake up again in a new "iteration" of the sci-fi world she was in. All three raters had to agree on the event type for it to be given a label.

Comparison of neural event boundaries and behavioral event boundaries.

We used data from *Experiment 2* to identify a set of group-level or "normative" behavioral boundaries for each movie. Individuals' button presses were first rounded to the nearest second. To make the behavioral boundaries commensurate with the HMM-derived neural boundaries, we needed to account for the delay in the hemodynamic response, which may be partially offset by the delay in

behavioral motor responses (i.e., button presses) after a boundary is detected. Therefore, we added 3 TRs/seconds to each button press (without allowing for events past the end of the movie).

We took 2 approaches (described below) to compute the similarity between normative HMM-derived boundaries and behavioral-study normative boundaries (**Fig. 1B**).

Density Alignment Method.

We generated a density distribution of "button presses" to indicate, at each timepoint (second), what percentage of subjects in the behavioral study marked a boundary location at that timepoint (or within +/- 3 s of that timepoint). We then computed a Pearson correlation between this behavioral density distribution and the density distribution of individual-subject HMM-derived boundaries for each node, and compared this "true" correlation to a null distribution of correlations that were generated by "rolling" (or circle-shifting) the HMM-derived boundaries at each TR (thus the number of permutations was limited to the number of TRs for each movie). This generation of a proportion-based density distribution is similar to the "segmentation agreement" previously used to compare individuals to the group average (Bailey et al., 2013; Zacks et al., 2006).

Peak Alignment Method.

We used our permutation-based alignment metric (See "Computing alignment across subjects"), treating the brain as the "ground-truth" (i.e. permuting behavioral boundaries) to compare the behaviorally-derived boundaries to the group-average HMM-derived boundaries for each ROI. Normative (group-average) HMM-derived boundaries were computed by averaging voxelwise activity across subjects and then fitting an HMM to data within each ROI using the pre-determined number of events for that ROI. To determine the normative boundaries from the behavioral study for this approach, we needed "peak" shared behavioral boundaries. For this, we defined the peaks as timepoints when >50%, or at least 21/40, subjects marked a boundary and enforced local sparsity by limiting peaks to timepoints that were more than 3 seconds apart. (If two peaks were within 3 seconds of one another, we took the higher "peak" [more agreement]; if they were of equal height, we took the median timepoint.)

Importantly, we used 3 seconds as our tolerance for small deviations in button-press times across subjects to be in line with the fMRI data, where we had used 3 TRs (= 3 s).

Combining Methods.

To declare significant alignment between normative behavioral and normative neural boundaries in a region, that region had to show above-chance alignment ($p < .05$) in both the density and peak methods.

Controlling for possible confounds.

Several factors outside the ongoing cognitive processes of interest could contribute to higher or lower alignment in detected boundary locations between a given pair of subjects. The factors detailed below were used as regressors of no interest to control for these unwanted influences in the ensuing analyses, as described in subsequent sections.

Inter-subject correlation (ISC) of head motion.

It is possible that shared head motion at similar moments in the movie could lead the HMM to (perhaps falsely) detect similar neural event boundaries in a given pair of subjects. To control for this possibility, we computed the inter-subject correlation (ISC) of the framewise displacement across time for each subject pair. We then used this subject-by-subject motion-ISC matrix as a nuisance regressor in subsequent analyses.

Overall head motion.

Similarly, subjects with high overall levels of head motion likely have lower-quality fMRI data, which could bias the detection of event boundaries altogether (though importantly, as detailed in "MRI data preprocessing", absolute levels of head motion were relatively low in our sample). To control for this possible confound, we used each subject's median framewise displacement (FD) across all timepoints and generated a subject-by-subject similarity matrix using the Anna Karenina principle ("all low (or high) motion scorers are alike; each high (or low) motion scorers is different in their own way") by taking the mean score between each subject pair (Finn et al., 2020).

Memory performance.

Some subjects may simply have been paying better attention throughout the task and/or during certain movies, which could generate a stronger neural response and in turn drive up similarity in event boundaries among these subjects. We controlled for this possibility using subjects' performance on the four multiple-choice memory recall questions presented at the end of each movie, which were designed to be quite difficult. We computed each subject's memory performance score as the fraction of correct responses on these questions, and generated a subject-by-subject similarity matrix of these scores also according to the aforementioned Anna Karenina principle ("all low (or high) memory scorers are alike; each high (or low) scorer is different in their own way") by taking the mean score between each subject pair (Finn et al., 2020).

Across-subject alignment within movies.

The goal of this analysis was to assess the degree to which event boundary locations were aligned across subjects within each movie (**Fig. 1D**- "Within-movie alignment"**)**. Subject-by-subject matrices of alignment values (z-scores) in every region were used.

Experiment 1.

We fit a linear regression to regress out (1) head-motion ISC, (2) overall head motion, and (3) memory performance (See "Controlling for possible confounds" for more detail) from the subject-by-subject boundary alignment matrix in each ROI. Using the residuals of this regression, we took the median alignment value (z-score) across subject pairs as the summary statistic for each ROI. To determine whether alignment was significantly above chance (i.e., greater than 0), we performed this same calculation in a subject-wise bootstrapping framework (n=10,000 bootstraps) to create a non-parametric null distribution and compared our observed median residual z-score to this distribution to calculate a p-value. P-values were corrected for multiple comparisons using the false discovery rate (FDR) based on the number of regions in our parcellation (100) using an alpha of .05.

Experiment 2.

We fit a linear regression to regress out memory performance (See "Controlling for possible confounds" for more detail) from the subject-by-subject matrix of alignment in behaviorally reported event boundaries. Using the residuals of this regression, we took the median alignment value (z-score) across subject pairs as the summary statistic. To determine whether alignment was significantly above chance (i.e., greater than 0), we performed this same calculation in a subject-wise bootstrapping framework (n=10,000 bootstraps) to create a non-parametric null distribution and compared our observed median residual z-score to this distribution to calculate a p-value.

Identifying content-dependent properties by taking the difference in alignment across movies.

The goal of this analysis was to assess the degree to which stimulus content influences across-subject alignment in event boundary locations by comparing subject-by-subject alignment matrices between pairs of movies (**Fig. 1D**- "Across-movie difference").

Experiment 1.

For a given region and pair of movies, we fit a linear regression to model the difference between the two subject-by-subject alignment matrices as a function of the following regressors of no interest: (1) difference in head-motion ISC, (2) the difference in overall head motion, and (3) the difference in memory performance between the movies (See "Controlling for possible confounds" for more detail). We then took the median value from this residual difference matrix and compared it to a non-parametric null distribution (n = 10,000 bootstraps) to calculate a p-value. P-values were corrected for multiple comparisons using the false discovery rate (FDR) based on the number of regions in our parcellation (100) using an alpha of .05.

Experiment 2.

For a given pair of movies, we fit a linear regression to model the difference between the two subject-by-subject alignment matrices as a function of the following regressor of no interest: the difference in memory performance between the movies (See "Controlling for possible confounds" for more detail). We then took the median value from this residual difference matrix and compared it to a non-parametric null distribution (n = 10,000 bootstraps) to extract a p-value.

Identifying trait-dependent properties by correlating alignment across movies.

The goal of this analysis was to assess the extent to which individuals had segmentation "styles" that were consistent across movies by determining whether pairs of subjects with high alignment in one movie also showed high alignment in another movie. (**Fig. 1D-** "Across-movie correlation"). Subject-by-subject matrices of alignment values (z-scores) in every region were correlated across pairs of movies.

Experiment 1.

For each film, we fit a linear regression to regress out the effects of the following: (1) head-motion ISC, (2) overall head motion, and (3) memory performance from the subject-by-subject alignment matrix (See "Controlling for possible confounds" for more detail). Using the residuals from these regressions, we then correlated alignment matrices between each pair of movies (akin to an inter-subject representational similarity analysis) and assessed statistical significance using a Mantel test (Mantel, 1967; number of permutations (nPerms)=10,000). Resulting p-values were corrected for multiple comparisons using the false discovery rate (FDR) based on the number of regions in our parcellation (100) using an alpha of .05.

Experiment 2.

We fit a linear regression to regress out the effects of memory performance from the subject-by-subject alignment matrix (See "Controlling for possible confounds" for more detail). Using the residuals from these regressions, we then correlated alignment matrices between each pair of movies (akin to an inter-subject representational similarity analysis) and assessed statistical significance using a Mantel test

(Mantel, 1967; nPerms = 10,000). Non-existent values ('nan') - generated when subjects had no alignment in their reported behavioral boundaries - were found in this dataset. This was most commonly an issue for $Lemonade_{NON\text{-}SOC}$; for $Growth_{SOC}$ and $Defeat_{SOC}$, there was one pair of subjects (out of 903 possible pairs in the lower triangle of our matrices) where there was no alignment and for $Lemonade_{NON\text{-}SOC}$, there were 36 pairs (out of 903 possible pairs) [see *Results* - "Degree of alignment varies with individuals and movie content: behavioral event boundaries", for possible explanations as to why $Lemonade_{NON\text{-}SOC}$ may have led to more variability in responses]. When appropriate, 'nan' values were replaced with the minimum value in the corresponding matrix.

Measuring cross-subject similarity in movie recall/appraisal.

We used an inter-subject representational similarity analysis (IS-RSA) approach (P.-H. A. Chen et al., 2020; Finn et al., 2020; Glerean et al., 2016) to investigate whether pairs of subjects that were similar in their event boundaries in a region, also had shared interpretations of the movie (**Fig. 1D-** "Inter-subject RSA").

To elicit appraisals, we presented the following prompt to subjects immediately following each film: "*During this section, you will have three minutes to say what you remember about the video. You can talk about characters, events, your opinions, or anything else that comes to mind. Try to fill the whole three minutes once the timer appears and remember- there are no wrong answers!*" Subjects spoke freely while their speech was recorded using a noise-canceling microphone. These recordings were professionally transcribed and minimally cleaned to remove interjections, such as "um" or "uh", and repeated words. One subject's recall data was not able to be transcribed due to being corrupted by scanner noise and was discarded from this analysis, leaving n=42.

The total number of words spoken varied across subjects and movies: $Iteration_{SOC}$: - mean=386, s.d. = 80, range = 213-547, $Defeat_{SOC}$ - mean=362, s.d. = 78, range = 153-533, $Growth_{SOC}$ - mean=368, s.d. = 72, range = 243-520, $Lemonade_{NON\text{-}SOC}$ - mean=359, s.d. = 84, range = 180-576. The number of words differed significantly between movies (repeated-measures ANOVA, $F(3,123) = 2.87$, $p=.04$). Post

hoc tests showed that the number of words used in $Iteration_{SOC}$ was significantly greater than the number of words used in $Defeat_{SOC}$ (paired t-test, t=2.81, p=.01) and $Lemonade_{NON\text{-}SOC}$ (paired t-test, t=2.12, p=.04). Given that $Iteration_{SOC}$ was the longest movie (~12 minutes as compared to ~8 minutes for the other 3 movies), we do not find it surprising that more words were used to describe this movie. Further, given that our goal in this analysis was to compare subject-to-subject similarity in boundary location to recall within a movie as opposed to across movies, we do not think that differences in the number of words used should influence our results in a meaningful way. However, importantly, the degree of similarity in the appraisal of $Iteration_{SOC}$ is not significantly higher than the other movies. In fact, the highest similarity was in $Growth_{SOC,}$ which was greater than $Iteration_{SOC}$, $Defeat_{SOC}$, and $Lemonade_{NON\text{-}SOC}$ (paired t-tests, t≥11.84, p<.001) (median similarity − $Iteration_{SOC}$ - 0.58, $Defeat_{SOC}$- 0.59, $Growth_{SOC}$- 0.63, and $Lemonade_{NON\text{-}SOC}$ - 0.58; repeated-measures ANOVA, F(3,2580) = 86.06, p<.001). These results emphasize that there is no direct relationship between the number of words used to appraise a movie and the degree of similarity in appraisals across subjects.

The text from each individual's speech data was then encoded into Google's Universal Sentence Encoder (USE; Cer et al., (2018); implemented via https://www.tensorflow.org/hub/tutorials/semantic_similarity_with_tf_hub_universal_encoder). We chose the USE model in part because it was trained to identify similarities between pairs of sentences, and, as a sanity check, because it was best able to differentiate appraisals from different discern judgments between movies (i.e., recall between $Growth_{SOC}$-$Growth_{SOC}$ was more similar than recall between $Growth_{SOC}$-$Defeat_{SOC,}$ etc.) whereas this sensitivity was not seen with other pre-trained context-sensitive models. USE was also used on event descriptions in a recent publication (Lee & Chen, 2022). We computed the cosine similarity between the 512-dimensional vectors generated by USE to measure the semantic similarity between pairs of subjects in their interpretations, resulting in one subject-by-subject appraisal similarity matrix per movie. In an inter-subject representational similarity analysis, we then used a Spearman correlation to compare these appraisal similarity matrices to the neural boundary

alignment matrices in each region. To assess significance, we conducted a partial Mantel test (nPerms = 10,000; q<.05, FDR corrected across regions) to assess the relationship between event boundary alignment and appraisal similarity while controlling for memory performance (See "Controlling for possible confounds" for more detail; **Fig. 1D-** "Inter-subject RSA"). By controlling for memory performance in this way, we were able to more cleanly isolate relationships with more subjective impressions and interpretations rather than objective recall *per se.*

Code Accessibility.

Data analysis, including links to code and other supporting material, can be found at: https://github.com/csavasegal/individual_event_seg.

Data Accessibility.

Data from this study, including raw MRI data will be made available on OpenNeuro upon publication. Other data including the behaviorally reported boundaries and the full transcribed recall and appraisal (text) can be found at: https://github.com/csavasegal/individual_event_seg/tree/main/data.

Results

Stimulus Features.

In this work, we quantified individual variability in neural and behavioral event segmentation using four different movies that varied in both low- (ex: continuity editing) and high-level features (ex: content). By varying features across these movies, we were able to both gain valuable insight into what aspects of event segmentation are consistent irrespective of the movie and to assess how these features influence idiosyncratic segmentation.

Social content (the presence of animate characters) is thought to enhance idiosyncratic neural responses (Finn & Bandettini, 2020). Three of our movies contained social information (i.e., human characters; $Iteration_{SOC}$, $Defeat_{SOC}$, $Growth_{SOC}$), and one contained only mechanical information ($Lemonade_{NON\text{-}SOC}$). The latter was intended as a control movie in that it still had a clear trajectory (beginning, middle, and end), but involved only mechanical events as opposed to a human-driven plot line. We hypothesized that the social movies would activate both shared social schemas (ex: family dinners) and idiosyncratic responses (ex: viewer-specific memories of their own family dinners). Further, we expected that the presence of social content would lead to more shared event boundaries (shared activity: [$Defeat_{SOC}$, $Iteration_{SOC}$, $Growth_{SOC}$] > [$Lemonade_{NON\text{-}SOC}$]) and that idiosyncratic segmentation would be linked to specific interpretations of the movies.

Previous behavioral and neural segmentation work suggests that continuity editing (such as screen cuts) facilitates segmentation (Schwan et al., 2000) since cuts are highly correlated with situational changes that indicate the start of a new event. Two out of the four of our movies ($Defeat_{SOC}$ and $Iteration_{SOC}$) contained screen cuts and two lacked screen cuts ($Growth_{SOC}$ and $Lemonade_{NON\text{-}SOC}$); in these movies, the camera panned from one event to the next in a continuous manner. Thus, we expected that the presence of screen cuts would lead to more shared event boundaries ([$Defeat_{SOC}$, $Iteration_{SOC}$] > [$Growth_{SOC}$, $Lemonade_{NON\text{-}SOC}$]).

Taking into consideration both content and continuity editing, we expected shared boundaries (higher alignment) to follow the following ranked order across movies: [$Defeat_{SOC}$, $Iteration_{SOC}$] > [$Growth_{SOC}$] > [$Lemonade_{NON\text{-}SOC}$]. All four of these movies generated variable interpretations in a pilot study that we conducted. Importantly, although different types of stimuli can be more engaging, influencing the degree to which shared responses exist (Song et al., 2021), in our experimental paradigm, there were no significant differences in self-reported engagement levels after watching each movie (repeated-measures ANOVA, $F(3,126) = 1.67$, $p = .18$).

Slower-segmenting regions show more individual variability

Forty-three subjects watched all four movies during fMRI scanning. Movie order was pseudorandomized for each subject such that order was counterbalanced at the group level. Newly developed algorithms have been shown to detect event boundaries in group-average fMRI data. To define neural event boundaries, we used one of these algorithms: the Hidden Markov Model (HMM) first proposed by Baldassano et al. (2017), which does not rely on annotations or hand-demarcated events but detects event boundaries as shifts in stable patterns of brain activity. To our knowledge, this algorithm had not been previously applied to individual-subject data. Therefore, we first undertook an analysis to assess whether this model can stably and reliably detect event boundaries at the individual level. All fMRI data were parcellated into 100 cortical regions using the Schaefer atlas (Schaefer et al., 2018).

We first sought to replicate past work (Baldassano et al., 2017; Geerligs et al., 2022) showing that at the group level, the number of events (i.e., the granularity of segmentation) is higher in sensory regions and lower in higher-order association areas that are sensitive to narrative information at longer timescales (Honey et al., 2012; Lerner et al., 2011; cf. **Fig. 2A**). Using a train-test procedure to determine the optimal number of events for each region from group-average data, we demonstrate that this relationship is maintained irrespective of the stimulus: event rate (number of events per minute) follows a posterior-to-anterior gradient.

We next sought to investigate to what degree segmentation varied across individuals. Event segmentation can vary both at the level of the number of events (*k*) and the location of boundaries; we focus here on the latter, both for methodological reasons and because it is more directly related to movie content (see *Methods* section "Automatic event boundary detection" for more details). We used the fixed value of *k* for each region determined at the group level (cf. **Fig. 2A**) to fit an HMM to each individual subjects' data from that ROI (see *Methods* section "Automatic event boundary detection"). We then assessed to what degree event boundary locations were aligned across subjects using a permutation-based method that generates a z-score for observed alignment relative to an appropriate null distribution for the total number of boundaries in each region (see *Methods* sections "Computing alignment across subjects" and "Across-subject alignment within movies" for details; **Fig. 1D-** "Within-movie alignment").

Across all four movies, event boundaries were significantly and positively aligned across subjects in most regions of the brain (>89%), but this degree of alignment varied on another posterior to anterior gradient such that event boundary locations were more shared ("common") in sensory regions and more idiosyncratic in higher-order regions (cf. **Fig. 2B**). This finding is consistent with the literature on individual differences in neural activity that have been measured with inter-subject correlation (ISC). We established strong correlations (r= .78-.90) between event rate and the degree of alignment across all four movies: faster-segmenting regions showed higher alignment across subjects (less individual variability, while slower-segmenting regions showed less alignment across subjects (more individual variability; cf. **Fig. 2C**).

Notably, insignificant or negative across-subject alignment was seen in 20 unique regions across the four movies (1 for $Defeat_{SOC}$, 9 for $Iteration_{SOC}$, 11 for $Growth_{SOC}$, and 10 for $Lemonade_{NON\text{-}SOC}$). Negative across-subject alignment would indicate that permuted boundaries have a higher chance of showing alignment than the true boundaries. These regions included the prefrontal cortex (14/20), orbitofrontal cortex (2/20), temporal pole (3/20), and the cingulate (1/20). Of these, the majority were higher order "default mode" (40%) or limbic regions (25%) that have the slowest timescales of information processing (Geerligs et al., 2022; Hasson et al., 2008) and show more idiosyncratic anatomy and function

(Hill et al., 2010; Mueller et al., 2013), including during naturalistic stimulation (L. J. Chang et al., 2021; Finn et al., 2018; Gao et al., 2020; Vanderwal et al., 2017). (It should also be noted, however, that these regions are also most commonly affected by signal drop-out.) No parcel showed negative or insignificant alignment across all four movies. (All analyses in this section were controlled for the effects of head motion and overall memory performance; see *Methods* section "Controlling for possible confounds").

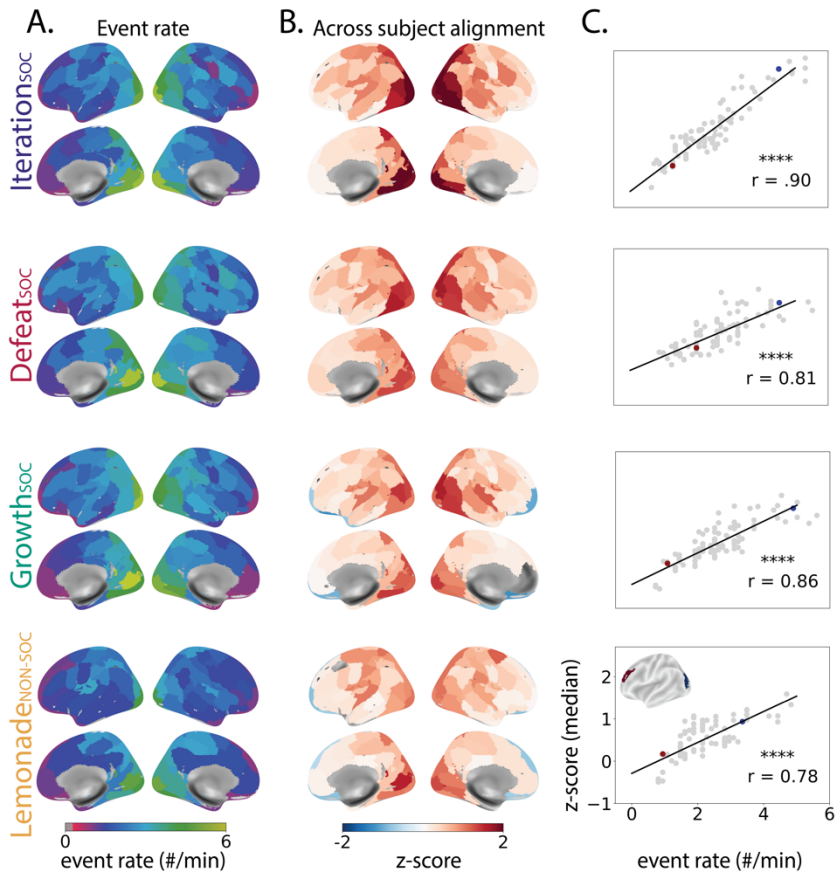

Figure 2 - Segmentation rate and degree of idiosyncrasy in boundary locations covary across the cortex. A. Event rate. The optimal number of events for each region (determined at the group level) follows the expected cortical gradient: faster segmentation (higher number of events) in posterior, sensory regions and slower segmentation (fewer events) in anterior, higher-order regions. **B. Across subject alignment within movie.** There is generally above-chance alignment in the location of event boundaries among individuals across the cortex (n=10,000 bootstraps, q<.05, FDR corrected across 100 regions). Alignment

is highest in posterior sensory regions and lowest in anterior association regions. The degree of alignment in each region varies slightly across films. While greater-than-chance alignment was found in the majority of regions (>89%) in each movie, regions that did not show significant alignment (across movies) tended to be in high-order regions such as the prefrontal cortex. **C. Correlation between event rate and within-movie alignment.** In all four movies, event rate and degree of alignment were strongly correlated across regions ($r = .78\text{-}.90$, $p < .001$) such that fast transitioning regions were more closely aligned across subjects, while slower regions were less aligned (i.e, showed more individual variability). A Spearman rank correlation was used. Each gray dot indicates a region; we highlight a visual region in blue and a prefrontal region in red. **** -p<.0001.

Validating automatically detected neural event boundaries using behavioral event boundaries.

Prior to further investigating individual variability in boundary locations, we first sought to validate neural event boundaries at the group ("normative") level. Past work has shown that automatically detected neural event boundaries tend to align with behaviorally reported event boundaries in certain regions of association cortex such as the angular gyrus or the posterior medial cortex (Baldassano et al., 2017). However, this work has been limited to single stimuli, so the extent to which behavioral-neural event alignment is content-specific has yet to be explored.

Towards this goal, we conducted an auxiliary behavioral experiment (*Experiment 2* - See *Methods*) in which a separate set of 40 subjects at Dartmouth College performed the same task as *Experiment 1* outside the scanner, except that while watching each movie, individuals were instructed to press a button each time they thought there was an event boundary; i.e., points in the movie when there is a major change in topic, location, time, etc.). For each film, we identified "normative" behavioral boundaries where subjects agreed, on average, that there was an event boundary. We then ran a content analysis to gain an

understanding of movie changes at event boundaries, and, proceeded to explore if and where these normative behavioral boundaries aligned with normative neural boundaries for each film.

In our content analysis, we identified five categories of events that tended to drive normative behavioral boundaries (See *Methods* section "Human annotation of event type"): 1) human-driven action, 2) object-driven action, 3) screen cut, 4) location change, and 5) "iteration" (an event type specific to the movie $Iteration_{SOC}$; see *Methods* section "Human annotation of event type"). Events in $Iteration_{SOC}$, $Defeat_{SOC}$, and $Growth_{SOC}$ consisted of mostly human-driven actions that were separated from one another by either, typically, a screen cut or iteration ($Iteration_{SOC}$, $Defeat_{SOC}$) or a location change ($Growth_{SOC}$). Importantly, screen cuts in our movies, as previously shown, are highly correlated with situational changes within the movies (Zacks et al., 2009). Thus, these event boundaries do not reflect just editing choices *per se,* but rather an ongoing parsing of the narrative trajectories that are cued for the viewer by screen cuts (Schwan et al., 2000). Location changes in $Lemonade_{NON\text{-}SOC}$ and $Growth_{SOC}$ had the same functional importance to the stimulus trajectory (narrative or mechanical) as a screen cut. However, location changes in $Lemonade_{NON\text{-}SOC}$ were bookended or caused by object-driven actions as opposed to human-driven actions. Events labeled with multiple colors in **Fig. 3A** depict the visual cue on screen at the time (bottom), and what was happening on either side of the boundary (top) (for instance, a screen cut between two human-driven action scenes in $Defeat_{SOC}$). These findings are in line with how situational changes, such as switches in space, goals, and character interactions and movement cues (such as changes in velocity or acceleration) that are a part of either actor-driven activity (ex: making food) or object-driven activity (ex: domino effect) are strongly predictive of segmentation across subjects (Hard et al., 2006, 2011; Speer et al., 2009; Zacks, 2004; Zacks et al., 2009).

Having validated these normative behavioral boundaries by linking them to discernible features of the stimuli, we then sought to investigate the extent to which they aligned with normative neural boundaries. We identified normative neural boundaries by fitting an HMM with k (the number of events) for each region set to the number reported in **Fig. 2A** to the fMRI data averaged across subjects. We then compared boundary locations between the behavioral and neural data to determine which brain region(s) best reflect

behavioral segmentation and whether these regions vary with stimulus content (i.e., across movies). We measured alignment between the HMM-derived and behavioral-derived event boundaries using two approaches (cf. **Fig. 3C,** see *Methods* section "Comparison of neural event boundaries and behavioral event boundaries", p<.05 across both methods). Regions where there was significant alignment between the two modalities varied across movies **(Fig. 3B)**, emphasizing the importance of stimulus content. However, within each movie, regions with neural-behavioral segmentation alignment showed anatomical consistency with previous reports of regions that are sensitive to event changes (Baldassano et al., 2017; DuBrow & Davachi, 2016; Masís-Obando et al., 2021; Speer et al., 2003, 2007; Zacks et al., 2001). These included higher-order attention or default mode regions, including the temporoparietal junction (TPJ), prefrontal cortex (PFC), insula, parietal cortex, and cingulate. The correspondence between human-annotated boundaries and automatically detected neural boundaries in certain brain regions increases confidence in the validity of the neural boundaries. Further, these two analyses gave us confidence in our ability to interpret the individual nuances in boundaries (reported below) that existed in addition to these shared responses.

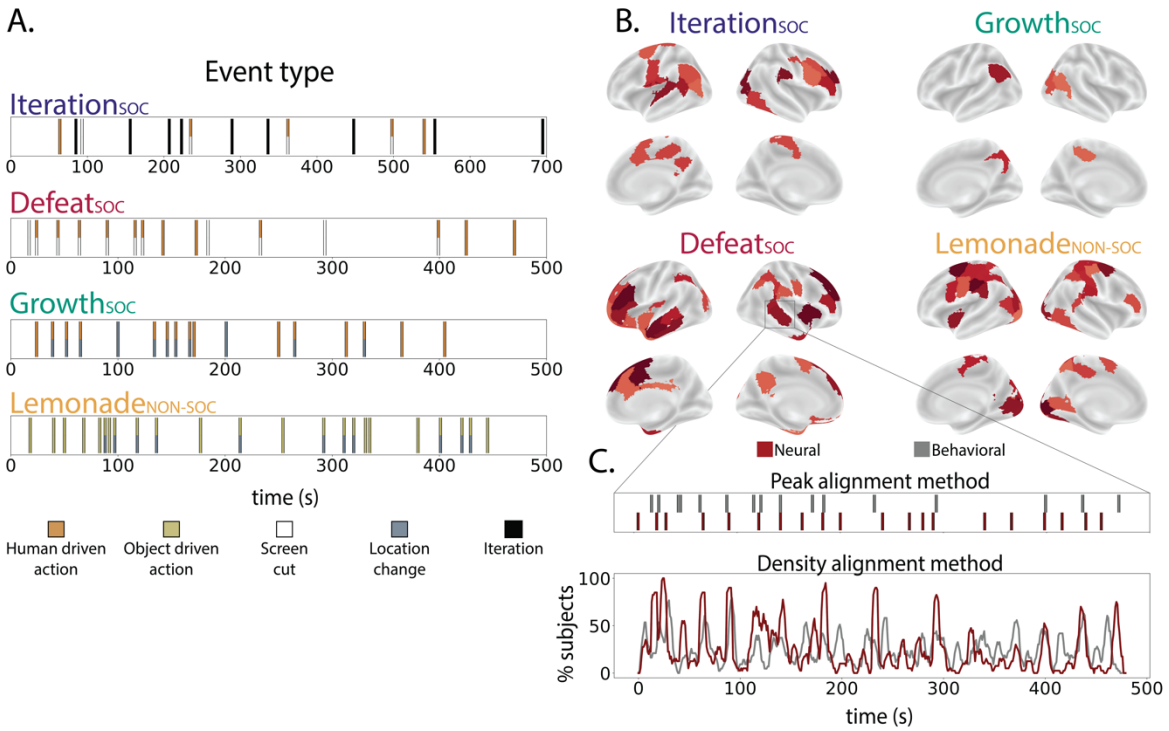

Figure 3 - Relationships between normative behavioral boundaries, normative neural boundaries, and stimulus content. A. Normative behavioral boundaries reflect definable aspects of the movies. Independent raters classified each normative behavioral boundary into one of five types, defined as follows: 1) "human-driven action" were moments where humans are actively engaged in an action that drives the plot (ex: building a time-machine, leaving their homes); 2) "object-driven action" were moments where the movement of objects drives the change to the next event (these events are unique to $Lemonade_{NON\text{-}SOC}$); 3) screen cuts were instantaneous changes in camera shot; 4) location changes typically served a similar function to screen cuts in $Growth_{SOC}$ and $Lemonade_{NON\text{-}SOC}$, which were both characterized by continuous camera panning; 5) "Iteration" was a type of event specific to $Iteration_{SOC}$ commonly consisting of both screen cuts and audio-visual cues. Screen cuts, iterations, and location changes typically were bookended by either human-driven or action-driven events. At boundaries characterized by two types of events, the top bar indicates the type of event that occurred on either side of the boundary (a human- or object-driven action) and the bottom bar represents the visual cue that illustrated the change (a screen cut or a location change). **B.** Regions showing significant alignment between normative (group-level) behavioral and neural boundaries ($q<.05$, FDR corrected across 100 regions within movie). Importantly, though there are numerous regions where this relationship holds, specific regions vary across films, suggesting that this relationship is somewhat content-specific. **C.** Alignment between normative behavioral and neural boundaries computed using two approaches (see *Methods*) depicted for an example region in the superior temporal sulcus.

Degree of alignment varies with individuals and movie content: neural event boundaries.

Having established a link between content and neural-behavioral alignment at the group level, we next sought to investigate how content influences individual variability in event segmentation, both neurally and behaviorally. We first determined the extent to which inter-subject alignment varied with stimulus content ("content-dependent") by comparing alignment between each pair of movies (**Fig. 1D- "**Across-movie difference"). (Analyses in this section were controlled for effects of head motion and overall memory

performance; see *Methods* section "Controlling for possible confounds".) We first noted that while there was significant across-subject alignment in all movies (**cf. Fig. 2B**), the overall median alignment across the cortex showed the following rank order of movies, from highest to lowest alignment: [$Iteration_{SOC}$ and $Defeat_{SOC}$] > [$Growth_{SOC}$ and $Lemonade_{NON\text{-}SOC}$] (cf. **Fig. 4A-Diagonal**). To determine this, using a linear model, we conducted an ANOVA to evaluate the fixed effect of movie on median alignment while treating region as a random effect ($F(3,300)=27.52$, $p<.0001$; Bates et al., 2015). Within the model, pairwise comparisons revealed that $Iteration_{SOC}$ showed significantly higher values than $Growth_{SOC}$ ($t=6.82$, $p<.0001$) and $Lemonade_{NON\text{-}SOC}$ ($t=7.45$, $p<.0001$), and $Defeat_{SOC}$ showed significantly higher values than $Growth_{SOC}$ ($t=5.06$, $p<.0001$) and $Lemonade_{NON\text{-}SOC}$ ($t=5.69$, $p<.0001$)**.** Thus, we found that alignment was higher in movies in which there was a human-driven, social narrative trajectory, especially when there were screen cuts cueing new scenes ($Iteration_{SOC}$, $Defeat_{SOC}$). In movies where there were no screen cuts ($Growth_{SOC}$ and $Lemonade_{NON\text{-}SOC}$), across-subject alignment was lower.

Taking a region-specific approach, we compared alignment in each ROI between pairs of movies (**Fig. 4A; Fig. 4B-lower triangle**) and found that certain movie content (social) drives more alignment in neural event boundaries in some regions than other content (cf. **Fig. 4B- lower triangle**). Specifically, while some regions showed consistently low (PFC) or high (V1) alignment across movies, canonical social regions such as the superior temporal lobe (STL) showed higher alignment in social movies compared to non-social movies ([$Iteration_{SOC}$, $Defeat_{SOC}$, $Growth_{SOC}$] > [$Lemonade_{NON\text{-}SOC}$]) (cf. **Fig. 4A; Fig. 4B bottom row**). Further, we found that among the social movies, in regions where there were significant differences in alignment, most were in the direction [$Iteration_{SOC}$ and $Defeat_{SOC}$] > [$Growth_{SOC}$]. Our non-social movie, $Lemonade_{NON\text{-}SOC}$, showed primarily higher alignment in somatomotor processing regions, likely reflecting neural state changes in response to the continuous switches in the location and object-driven motion that are specific to this movie (cf. **Fig. 3B**).

Regions where normative neural boundaries reflected normative behavioral boundaries in a given movie ("cross-modal regions"; cf. **Fig. 2B**) were more likely to show higher across-subject alignment in that movie compared to other movies. For instance, for $Iteration_{SOC}$, 60% (12) of the cross-modal regions

showed higher alignment in $Iteration_{SOC}$ in the majority (at least two out of the three) of comparisons with other movies, including 2 in the prefrontal cortex and 2 in the default-mode network; for $Defeat_{SOC}$, 64% (21) of cross-modal regions showed higher alignment in this movie compared to other movies, including 8 in the prefrontal cortex and 5 in the temporal lobe, (6 in the default-mode network). This suggests that neural regions reflecting "conscious" event segmentation also tend to have more consistent boundary locations across individuals, further highlighting that certain types of content are better suited to drive shared activity in certain regions.

Having established that there are content-dependent effects that influence the degree of individual variability, we next determined the extent to which there are "trait-like" aspects of segmentation in certain regions: i.e., if a pair of subjects segments similarly in one movie, do they also segment similarly in another movie (**Fig. 1D-** "Across-movie correlation")? If so, this would suggest that individuals have certain segmentation "styles" that hold across different movies (i.e., different content). Notably, significant relationships were mostly limited to sensory processing regions, with some exceptions in higher-order regions (e.g., TPJ, dorsolateral PFC, cingulate, insula) for certain movie pairs (cf. **Fig. 4B- upper triangle**). No single ROI emerged in all 6 pairwise movie comparisons, suggesting that an interaction between stimulus content and individual segmentation styles drive neural event boundaries. Specifically, we suggest that state changes in neural activity are driven by the ongoing external stimulus as opposed to a more "trait-like" idiosyncratic and consistent segmentation style that is specific to a region. We suggest that if future work were to carefully match stimuli across relevant, unambiguous boundary-driving features (ex: screen cuts, motion, character interaction) then, under these specific demands, neural event segmentation "styles" might emerge. However, the diverse stimuli used here allowed us to dissociate content-driven from trait-driven properties of event segmentation.

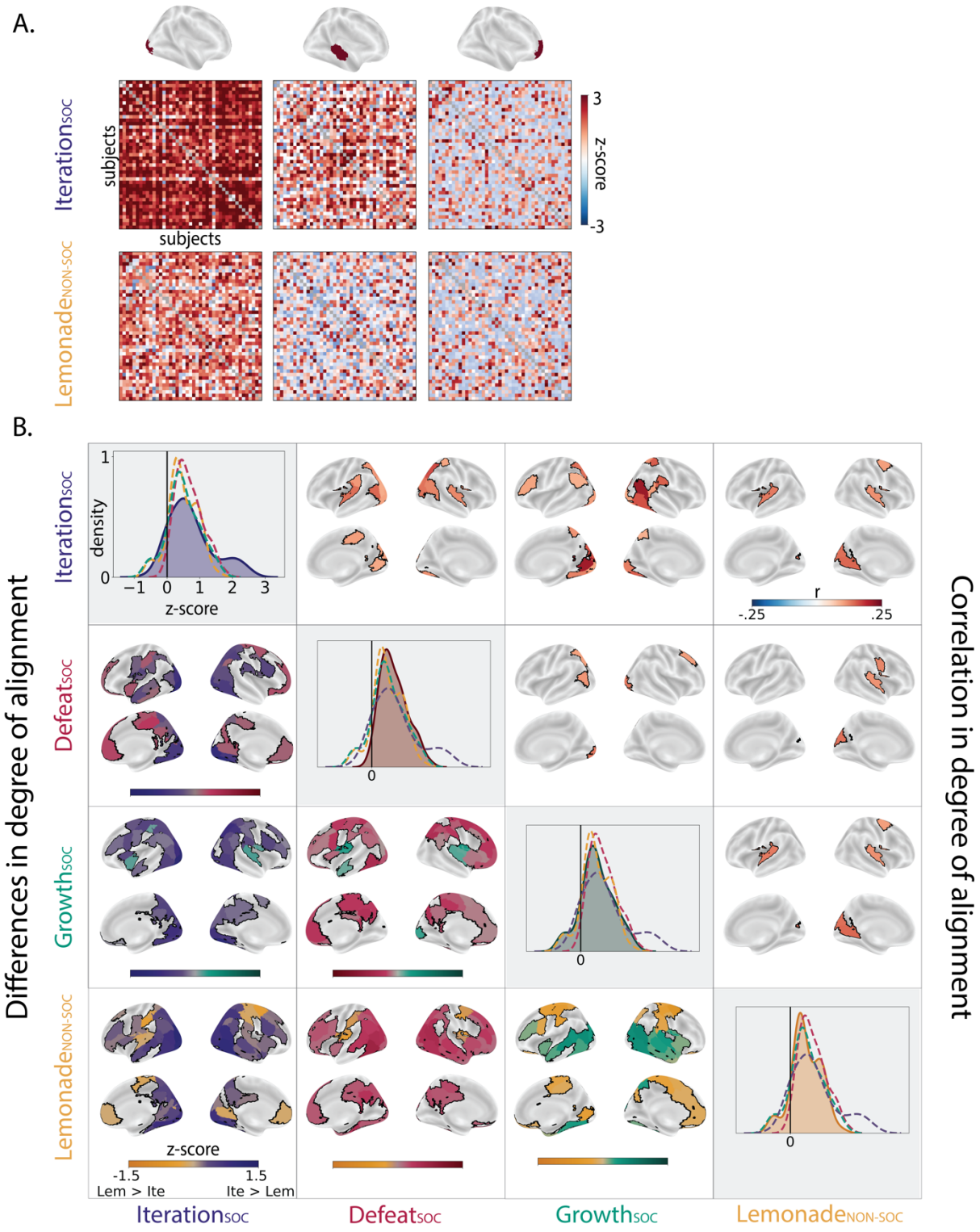
A.
Iteration SOC
Lemonade NON-SOC
subjects
subjects
3
z-score
-3
B.
Differences in degree of alignment
Correlation in degree of alignment
Iteration SOC
Defeat SOC
Growth SOC
Lemonade NON-SOC
1
density
0
−1 0 1 2 3
z-score
r
-.25
.25
0
z-score
-1.5
1.5
Lem > Ite
Ite > Lem

Figure 4 - Content-dependent and "trait-like" effects on neural event segmentation. A. Alignment differs with movie content in certain regions. Direct comparisons between matrices from $Iteration_{SOC}$ and $Lemonade_{NON\text{-}SOC}$ demonstrate that while across-subject alignment in both movies is high in a primary visual region and low in a prefrontal pole region, in the superior temporal sulcus (STS), a canonical social processing region, alignment is higher with social content than with non-social content. Notably, this same region shows significantly higher alignment in all social movies as compared to the non-social control movie (bottom row in B). These matrices are an illustration of the data that went into B. **B. Lower triangle: Degree of neural alignment differs with movie content ("content-dependent effects").** Pairwise comparisons between movies indicate regions where alignment is significantly higher in one movie as compared to another. Coloring indicates the movie where there is higher alignment and shade reflects the magnitude of the difference (maps are thresholded at $q<.05$; FDR corrected to 100 regions within movie pair). Significance was determined using the residuals after regressing the effects of head motion and memory performance. **Diagonal: Degree of overall neural alignment within movie.** Histogram depicts the distribution of median z-score values from the across-subject alignment matrices across all regions within movie (n = 100 per histogram). Whole-brain alignment shows the following rank order across films: [$Iteration_{SOC}$ and $Defeat_{SOC}$] > [$Growth_{SOC}$ and $Lemonade_{NON\text{-}SOC}$]. We suggest that this could be due to the presence of social content, and stimulus cuts. **Upper triangle: Consistent neural alignment between pairs of subjects irrespective of movie content ("trait-like effects").** Pairwise comparisons between movies indicate regions where pairs of subjects that are more aligned in one movie are also more aligned in another movie (maps are thresholded at $q<.05$; FDR corrected to 100 regions). Notably, a left hemisphere primary motor and right hemisphere frontal eye field ROI showed a significant negative correlation between $Lemonade_{NON\text{-}SOC}$ and $Iteration_{SOC}$, which are not plotted here.

Degree of alignment varies with individuals and movie content: behavioral event boundaries.

We next sought to investigate whether similar patterns of context- and trait-dependent variability in neural event segmentation were also present with behavioral event segmentation. Controlling for memory performance, we compared alignment across pairs of subjects in their behaviorally-reported event boundaries (**Fig. 1D-** "Across-movie difference") and found significant alignment in all movies (p=.0001; nbootstraps=10,000). Using rank-ordering across movies, we observed that movies that evoked more variability in neural event segmentation (across all brain regions; **Fig. 4.- Diagonal**) also evoke more variability in behavioral event segmentation ([$Iteration_{SOC}$ *and* $Defeat_{SOC}$]> [$Growth_{SOC}$] > [$Lemonade_{NON-SOC}$]; **Fig. 5.-Lower Triangle** and **Diagonal**). To specify, again, alignment was higher (more similarity) with social content and with screen cuts.

We also found strong, significant "trait-like" correlations among social-social movie pairs, suggesting that subjects that behaviorally segment social events in a similar way in one movie will also do so in another movie (**Fig. 5.- Upper Triangle**). The strongest correlation (r=.31) was seen between the two movies that were best matched in terms of social content and screen cuts ($Iteration_{SOC}$ and $Defeat_{SOC}$). While correlations were not found between $Iteration_{SOC}$ and $Lemonade_{NON-SOC}$ or between $Defeat_{SOC}$ and $Lemonade_{NON-SOC}$, which were not matched for high level content or continuity editing, we did identify a significant correlation between behavioral events in $Growth_{SOC}$ and $Lemonade_{NON-SOC}$, which were matched for having location shifts between either object-driven or human-driven actions (**Fig. 3C**). Therefore, although segmentation appears to be more "trait-like" behaviorally than neurally, as suggested above, movies may need to be matched to a certain extent in order for consistent individual segmentation styles to emerge. This finding also suggests that, somewhat paradoxically, the movies that drive more shared responses are also best suited to identifying consistent individual segmentation styles.

An alternative explanation is that the behavioral event segmentation task was harder for $Lemonade_{NON-SOC}$ given the somewhat abstract content and lack of screen cuts (as reported by subjects after the task), which could account for the significantly lower (and sometimes negative) alignment in $Lemonade_{NON-SOC}$ and for the lack of trait-like relationships between $Defeat_{SOC}$ and $Iteration_{SOC}$ and

Lemonade$_{NON\text{-}SOC}$. Crucially, this explanation still supports our inference that stimulus content matters by highlighting that segmentation is more straightforward for certain types of stimuli.

Given that we were able to detect both trait-like effects and content-dependent effects in regions that also show normative alignment, these results further validated our ability to auto-detect neural event boundaries at the individual level. Combined, our results show that individual variability in event segmentation is influenced by stimulus content irrespective of the modality (neural/HMM-derived or behaviorally-derived).

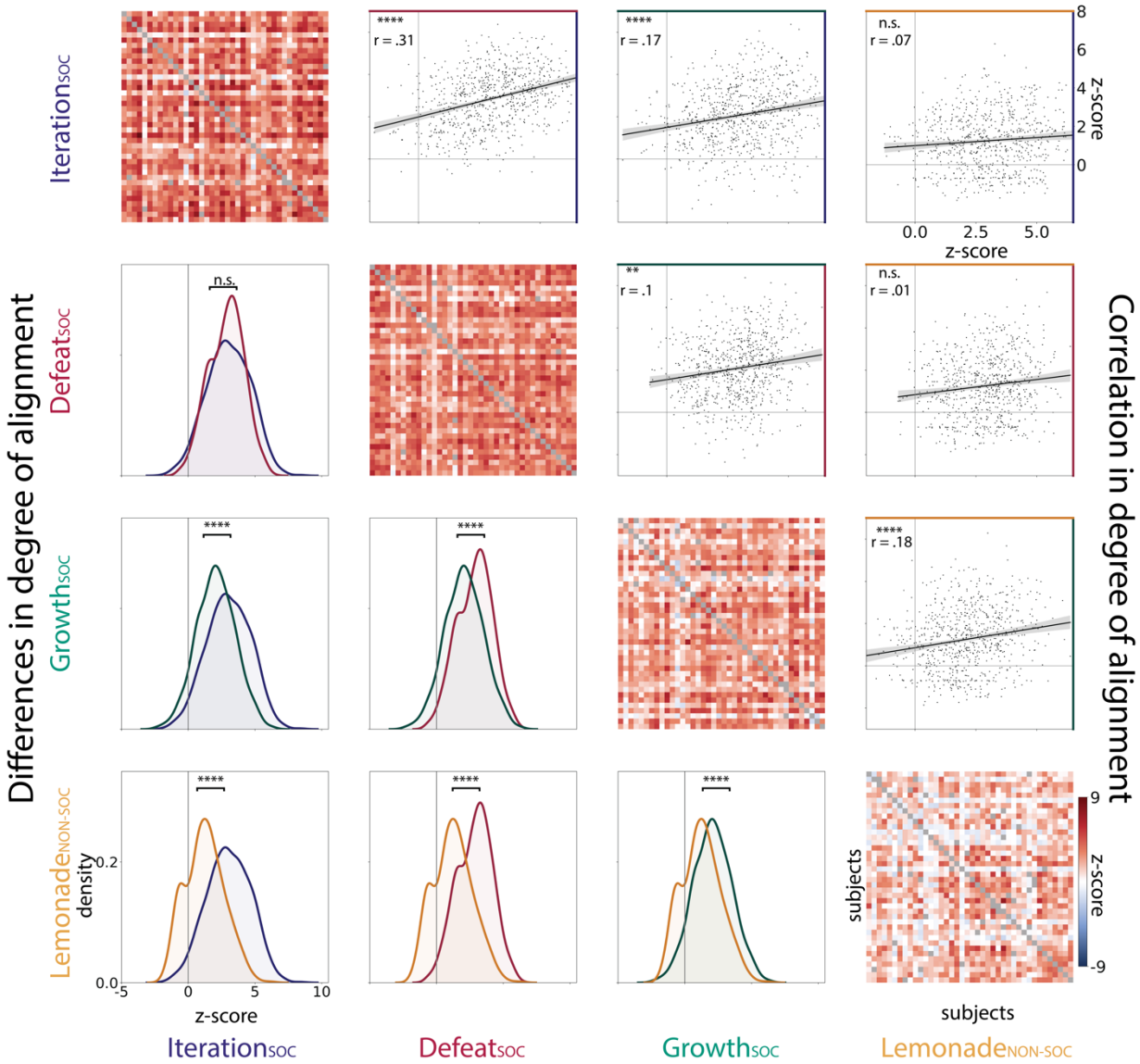

Figure 5 - Content-dependent and trait-like effects on behavioral event segmentation.

Lower triangle: Degree of behavioral alignment differs with movie content ("content-dependent effects"). Pairwise comparisons between movies indicate cases where alignment is significantly higher in one movie as compared to another. For visualization purposes, we are showing the distribution of alignment values "z-scores" between pairs of subjects (bootstraps=10,000). Significance was determined using the residuals after regressing the effects of memory performance. **Diagonal: Degree of overall behavioral alignment within movie.** Subject-by-subject alignment matrices are shown for each movie. The alignment follows a comparable rank order across movies: [$Iteration_{SOC}$ and $Defeat_{SOC}$] > [$Growth_{SOC}$] > [$Lemonade_{NON\text{-}SOC}$]. Just as with neural segmentation, we suggest that this could be due to the presence of social content, and stimulus cuts. **Upper triangle: Consistent behavioral alignment between pairs of subjects irrespective of movie content ("trait-like effects").** Pairwise comparisons between movies indicate whether subjects that are more aligned in their behavioral segmentation in one movie are also more aligned in another movie. This was seen in the majority of comparisons (except $Iteration_{SOC}$ and $Defeat_{SOC}$ with $Lemonade_{NON\text{-}SOC}$), which suggests, again, that social movies may be best for deriving "trait-like" event boundaries. For visualization purposes, we are showing the correlation of alignment values "z-scores" between movies (permutations=10,000). Significance was determined using the residuals after regressing the effects of memory performance. The stars indicate the degree of significance: *- p<.05, ** - p<.01, *** - p<.001, **** - p<.0001.

Shared neural event boundaries lead to shared interpretations.

Our final goal was to identify a link between neural event boundaries and the ultimate interpretation of a stimulus (**Fig. 1D-**"Inter-subject RSA"). We tested the hypothesis that individuals who were more similar in their neural event boundaries while watching a movie were also more similar in their interpretation of the movie. Past work has shown that, on average, group-level event boundaries have consequences for what information is remembered. We sought to extend this to *how* information is remembered, and how this differed across individuals. To test this hypothesis, we used free-speech data

acquired immediately following each of the four movies in which subjects were prompted to speak for three minutes about what they remembered and how they felt about the events and characters (see *Methods* section "Measuring cross-subject similarity in movie recall/appraisal" for exact instructions). Crucially, subjects did not simply recall the events that they watched like in a typical episodic memory task, but also shared subjective interpretations and overall reflections on the movies (appraisals), often including links to their own personal lives. Thus, we henceforth refer to this task as the appraisal task.

We recorded and transcribed each subject's speech and submitted the transcripts to Google's Universal Sentence Encoder (USE), a tool from natural language processing (NLP) that encodes text into high-dimensional vectors that reflect semantic content (Cer et al., 2018; see *Methods* section "Measuring cross-subject similarity in movie recall/appraisal"). Language embeddings provide a relatively unbiased way to quantify similarity in appraisal content. For each film, we then calculated a subject-by-subject similarity matrix from these vectors and compared it to the subject-by-subject similarity matrix of neural event boundaries in each ROI while controlling for objective memory performance (as measured by performance on memory questions) using an inter-subject representational similarity analysis (IS-RSA; **Fig. 1D -**"Inter-subject RSA").

Results (**Fig. 6A**) showed that in several brain regions, the degree of alignment in neural event boundaries during movie-watching predicted similarity in appraisal; i.e. subjects that were similar in event boundaries in these regions also tended to speak similarly about the movie afterwards. In the three social movies, these regions included high-order social cognition areas, some of which are considered part of the default-mode network (DMN). Three regions— the superior temporal lobe, the angular gyrus and the precuneus—emerged in all three social movies (**Fig. 6A;** black contours). Significant regions in the non-social control film ($Lemonade_{NON\text{-}SOC}$) had a different spatial pattern from the social movies. These regions included mostly somatomotor, dorsal-attention, and limbic regions, including a posterior temporal, ventral stream object-category selective region, the lingual gyrus, the temporal parietal junction (TPJ), cingulate, the orbitofrontal cortex, and a left somatomotor region previously linked to motor imagery (H. Chen et al., 2009; Chinier et al., 2014).

The degree of cross-modal alignment between normative neural and behavioral boundaries (density metric; cf. **Fig. 3C**) and the representational similarity between individual event boundaries and appraisal (**Fig. 6A**) were correlated across regions **(Fig. 6B)**, suggesting that regions that support conscious-level segmentation processes for a given movie at the group level are also relevant for how that movie is ultimately remembered and appraised at the individual level. Regions that showed significant cross-modal normative alignment as well as a significant relationship between individual boundary locations and appraisal are highlighted with blue contours and blue dots in Fig. 6A and 6B, respectively. We suggest that the same regions that show neural segmentation at moments corresponding to behaviorally reported boundaries were likely involved in appraising and translating these events into memory. For $Iteration_{SOC}$, this included nine regions (contoured in **Fig. 6A** in blue), such as the precuneus (DMN), a DMN medial parietal ROI, and the intraparietal sulcus; for $Defeat_{SOC,}$ this included five regions, such as the right dorsal PFC (DMN), and the STL (DMN); for $Growth_{SOC}$, this included a DMN medial parietal ROI and the TPJ; and for $Lemonade_{NON\text{-}SOC}$, this included three regions, including the previously mentioned somatomotor region implicated in motor imagery and the TPJ.

We then conducted a series of post-hoc analyses to identify what aspects of the stimulus may have driven the shared similarity structure between neural event boundaries and appraisal across subjects. We identified several instances where subjects that shared event boundaries at particular moments (**Fig. 6D**) then went on to speak about these specific events in their free recall (cf. **Fig. 6E**), supporting a link between event boundaries and recall of specific moments in the movie. We also identified that subjects who were similar in their overall event boundaries (cf. **Fig. 6C and 6D**) often also had shared general impressions of the film, i.e., not necessarily tied to particular moments. For instance, some pairs of subjects made similar comments about the broader message of the movie (**Fig. 6F**), similar references to other media (e.g., compare "*it reminds me of [sci-fi film]* "The Cube*'*"" and "*it reminds me a lot of the game* "Portal*"*), similar impressions of characters (e.g., compare *"A lot of these things show the selfish character of Tommy. Tiny [is] a child genius of some sort"* and *"So this film is about if you leave your kids unsupervised one will become a bully and because of that the other will become an evil genius"*) and similar overall appreciations

of the movie (e.g., compare "*I'd like to know who made it and why… I wondered how long it would take to set this thing up?*" and "*I was getting a little stressed out thinking about how long it is to set up … I was wondering how many people [it took] to build it*").

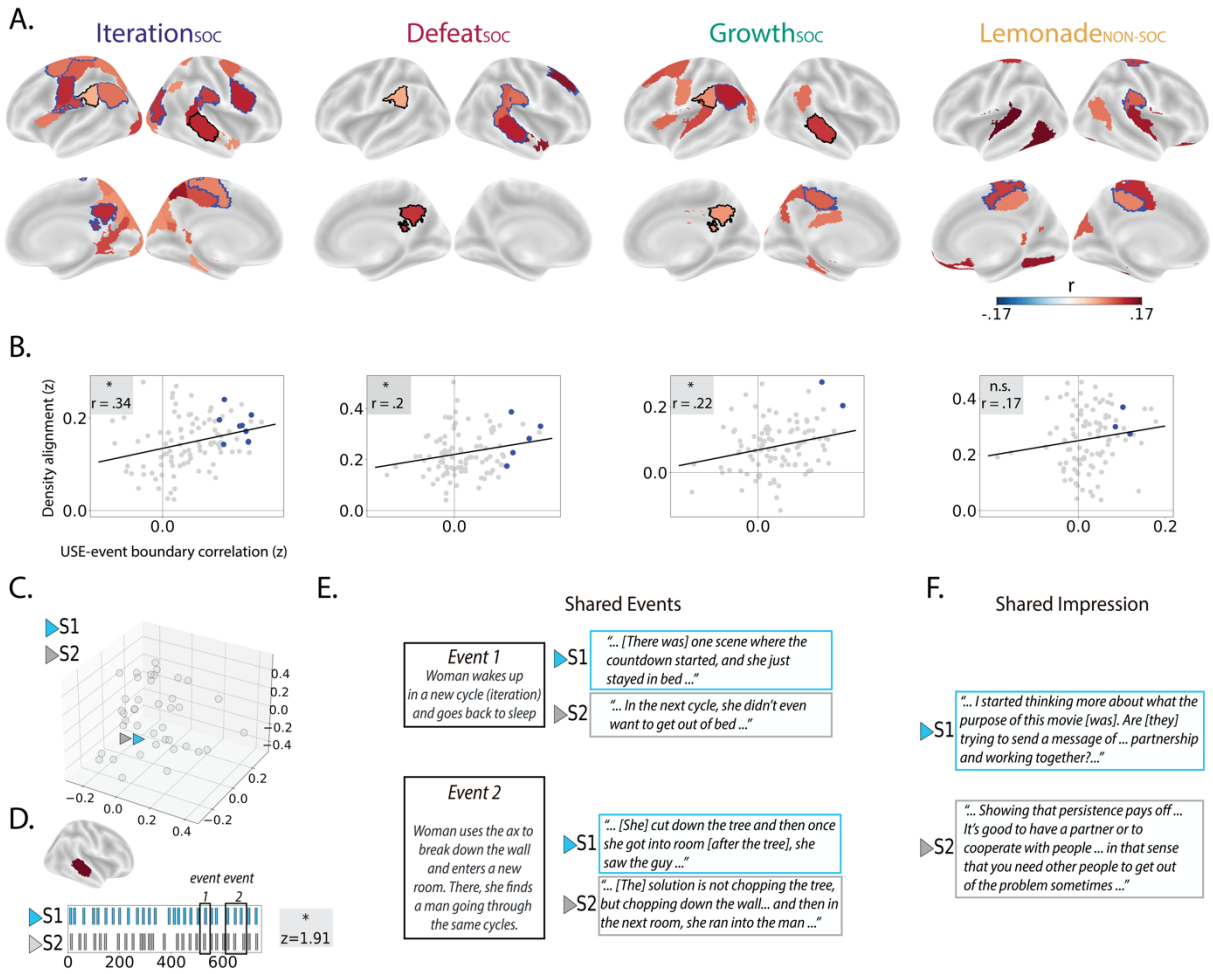

Figure 6- Degree of alignment predicts similarity of recall/interpretation.

A. Inter-subject representational similarity between event boundary locations and ultimate appraisals. Maps depict regions where pairs of subjects that were more similar in their event boundaries also had more similar interpretations of the movie (partial Mantel test controlling for objective memory performance; q<.05, FDR corrected to 100 regions within movie). Blue contours indicate regions that also

show significant alignment with normative behavioral boundaries (cf. Fig. 3B). Black contours indicate regions that showed significant representational similarity across all three social movies. Among these was a subregion of the superior temporal lobe (blue in $Defeat_{SOC}$ because it also showed a relationship with behavioral boundaries; partial Mantel test, $q<.05$, corrected within movie), the precuneus (blue in $Iteration_{SOC}$ because it also showed a relationship with behavioral boundaries), and the angular gyrus (partial Mantel test, both $p<.05$ uncorrected across all three social movies). **B. Regions are behaviorally relevant in multiple analyses.** The degree of alignment between neural and behavioral normative boundaries ("Density Alignment") is correlated with the relationship between individual event boundaries and individual recall. Gray dots represent each ROI. Regions that show a significant relationship (blue contours in A) are demarcated with blue dots. We suggest that these regions are involved in the conscious elements of event segmentation, both during encoding and recall. Values were Fisher transformed prior to running this analysis. A Spearman correlation was used. **C. Embeddings of free recall/interpretation speech data.** A PCA was used to put word-embeddings into 3D space. Grey circles indicate subjects. Triangles highlight a pair of subjects that had similar recalls and will be used throughout panels C-F. **D. Event boundary alignment.** The significant alignment between the sample subjects indicated in 5C is displayed in a superior temporal ROI for $Growth_{SOC}$. Specific events that are shared in these subject pairs are labeled and elaborated on in Fig 5E. **E. Subjects share event boundaries at important movie events that are then remembered similarly.** In their own idiosyncratic ways, subjects mentioned the same events in their recalls that they have shared boundaries for (Fig. 5D). **F. Subjects aligned in event boundaries have similar impressions of movies.** Beyond specific events (Fig. 5E), subjects that had similar alignment also had shared general interpretations.

Discussion

Individuals segment incoming information into events in different ways. Here, using four different continuous movie stimuli, we investigated individual differences in neural event boundaries and how they relate to behavior. Results showed that there is a posterior-to-anterior gradient for between-subject alignment in neural event boundaries that is tightly correlated with the rate of segmentation, such that regions that segment more slowly also show less alignment (i.e, more individual variability). Notably, we found strong content effects for regions in the middle of this gradient: while alignment was high in low-level sensory regions and low in high-order narrative processing regions across movies, regions that are more tuned to a certain type of input showed variable alignment depending on the stimulus. Although we show, both neurally and behaviorally, that the presence of social content and continuity editing increase alignment, we also found evidence for individual segmentation "styles" that are consistent across similar movies in some regions. Lastly, we describe one mechanism by which narratives may generate variable interpretations across people: individuals with more similar neural event boundaries in certain regions during movie-watching tended to have more similar appraisals of the movies.

Our first goal was to validate the use of automated event segmentation algorithms on fMRI data from individual subjects, and to characterize how much individuals vary in their neural event boundary locations. We found that although subjects show above-chance alignment in event boundaries in the majority of the cortex irrespective of the stimulus, the degree of alignment decreases from unimodal to transmodal, association regions. This finding is in line with reports using inter-subject correlation (ISC) to show that synchrony of activity decreases from posterior to anterior regions. Thus, regardless of measurement modality—ISC (continuous activity timecourses) or event segmentation (discrete state switches)— dynamic neural responses to external stimuli are more idiosyncratic in higher-order regions including the TPJ, STS, precuneus, dorsal medial and ventral medial prefrontal cortex, and the medial frontal gyrus. Notably, these regions are often considered part of the default mode network with reported

involvement in social cognition, self-referential processing, and the consolidation of autobiographical memory (for review: Yeshurun et al., 2021).

We also found that the degree of alignment is tightly correlated with segmentation rate across regions. This broadens recent findings on event rate (Baldassano et al., 2017; Geerligs et al., 2022) and on the intrinsic process memory and temporal receptive windows (the length of time before the response in which information will impact that response), which are all thought to follow a hierarchical posterior-to-anterior gradient as well (Hasson et al., 2015; Honey et al., 2012; Lerner et al., 2011). We suggest that information integration (as measured by event boundary alignment) is more stereotyped across subjects in sensory, unimodal cortex and becomes less standardized in higher-order regions with slower dynamics due, in part, to idiosyncratic processing strategies, experiences, and memories.

To support the link between idiosyncratic event boundaries and variable information integration, we demonstrate that patterns of neural event segmentation in certain regions relate to an individual's memory of the segmented experience. Using ISC, it has been shown that individuals with shared context (Yeshurun et al., 2017), shared traits (e.g., paranoia; Finn et al., 2018), or shared experimentally manipulated perspectives (Lahnakoski et al., 2014) {Citation}have similar neural responses while experiencing a narrative. We aimed to extend this past work and capture meaningful differences in how idiosyncratic neural activity during encoding relates to endogenously generated interpretations and recalls. Past work has theorized that current perceptual information interacts with long-term knowledge about event categories (schemas and scripts) when forming and updating event models (Radvansky & Zacks, 2014). These stereotyped changes likely form shared boundaries (Baldassano et al., 2018; Masís-Obando et al., 2022) and reflect central hubs in the narrative (Lee and Chen, 2022). We suggest that individual-specific boundaries—i.e., those not shared among the majority of subjects—may reflect moments with more idiosyncratic meanings (i.e., a moment activating one's own autobiographical memory) that lead to variable interpretations of a stimulus. To test this, we leveraged complex narratives that generated variable appraisals across people (as determined by our preliminary analyses), and had subjects freely discuss the movies. We then used inter-subject representational similarity analysis to show that pairs of subjects with

more similar event boundaries in numerous narrative processing regions, including the angular gyrus, superior temporal lobe, and precuneus, also had more similar appraisals of each stimulus. A large subset of these regions also reflect behaviorally reported event boundaries. Altogether, these findings uphold the functional role of these regions in the conscious encoding of an experience and its ultimate behavioral consequences, including its organization into memory.

While the idea that stimulus features systematically affect group-level segmentation is not novel—(Newtson et al., 1977) were among the first to consider how a film's features encourage segmentation of continuous streams of information into units—here, we extend this to characterize how movie content affects neural and behavioral segmentation at the individual level. Our movies were chosen to vary in both low-level features (i.e., continuity editing or screen cuts) and high-level content (i.e., social versus mechanical). The former was based on previous reports that film cuts ease segmentation (Schwan et al., 2000), generating shared responses, and the latter was based on our team's past suggestion that social content (i.e., human characters) better reveals individual differences (Finn & Bandettini, 2020). We demonstrated that although the spatial patterns of relative alignment were consistent across movies (cf. Fig. 1A), absolute degree of alignment differed with content: the presence of both continuity editing and character-driven activity drove both higher overall alignment across the cortex (cf. Fig. 3) and in behaviorally reported boundaries (cf. Fig. 4). We also found that degree of alignment in certain regions was sensitive to certain movie features; the superior temporal sulcus, temporal parietal junction (TPJ), precuneus and frontal gyrus showed higher alignment to character interactions, and the inferior parietal sulcus and the medial parietal showed higher alignment to spatial changes and interactions with objects (in line with spatial activation maps proposed by Speer *et al.,* 2009).

By scanning the same individuals watching multiple movies, we were able to tease out the extent to which individual segmentation styles are stable ("trait-like") irrespective of a movie versus movie/content-dependent. We identified stable individual effects in certain regions between certain pairs of movies. However, we found that we were best able to capture consistent patterns of segmentation between movies that were both well matched on low- and high-level features and also social in nature,

suggesting that (1) for segmentation styles to be consistent within an individual, the experience they are segmenting has to be similar since the mechanism of segmentation may differ with variable content, and (2) individual differences in segmentation are best captured when there already is some degree of a shared response. We suggest that if future work were to carefully match stimuli across relevant, unambiguous boundary-driving features (ex: screen cuts, motion, character interaction) then, under these specific demands, neural event segmentation "styles" might emerge. However, the diverse stimuli used here allowed us to dissociate content-driven from trait-driven properties of event segmentation.

For each film, a small subset of regions emerged—e.g., precuneus ($Iteration_{SOC}$), dorsolateral PFC ($Defeat_{SOC}$) and medial parietal ($Growth_{SOC}$)—with the following properties: 1) normative neural boundaries tracked normative behavioral boundaries (group-level); 2) individual boundaries were more aligned in that film compared to other movies (group-level); and 3) pairs of subjects that were more similar in event boundary locations were also more similar in their ultimate appraisal of the film (individual-level). This suggests that some degree of shared response enhances our ability to detect meaningful individual variability atop that shared response (Finn et al., 2020). Crucially, the specific regions where this held true were distinct across movies, suggesting that content plays a critical role in if and where this effect is observed.

There are some limitations to this work. First, although our choice of stimuli was somewhat principled in that it was based in existent theories about the effects of content on segmentation (Grall & Finn, 2021), our ultimate stimulus set is still arbitrary and there are plenty of other dimensions that these stimuli vary across and could be compared across. Second, our sample size (n=43) is relatively low for individual-differences work. However, a strength of our design is that we compared the same subjects across four movies. Furthermore, we did not attempt to link individual neural event segmentation patterns to trait-level behavioral measurements, an analysis for which we would likely be underpowered; rather, we focused on quantifying the overall degree of individual variability and how this differs as a function of both brain region and stimulus content, for which sample size should be a less limiting factor.

Overall, our work characterizes factors that influence individual differences in the location of event boundaries and emphasizes the importance of considering stimulus content in naturalistic neuroimaging paradigms. Although numerous studies have explored individual differences in event segmentation at the behavioral level, here, we extend existing methods to identify individual differences at the neural level during encoding and their consequences for how a stimulus is ultimately remembered and appraised. Future work should further explore this relationship by identifying whether clinical or personality traits (see Zacks & Sargent, 2010) that are stable over time are associated with particular neural styles of segmentation that impact ongoing cognition and behavior.

Funding

This work was supported by a National Science Foundation Graduate Research Fellowship to C.S.S. and by the National Institutes of Health grants K99MH120257 and R00MH120257 to E.S.F.

Acknowledgements

The authors thank Peter Bandettini, Peter Molfese, Daniel Handwerker, and Javier Gonzalez-Castillo for helpful discussions about experimental design and support with data collection. We also thank Sofia Yawand-Wossen and Payton Weiner for their assistance with the movie labeling and content analysis, and the Undergrad Research Assistantships at Dartmouth program for funding their support.

Conflicts of Interest

The authors declare no competing interests

-

References

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

Bailey, H. R., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Head, D., Kurby, C. A., & Sargent, J. Q. (2013). Medial Temporal Lobe Volume Predicts Elders' Everyday Memory. *Psychological Science*, *24*(7), 1113–1122. https://doi.org/10.1177/0956797612466676

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., Hasson, U., & Norman, K. A. (2018). Representation of Real-World Event Schemas during Narrative Perception. *Journal of Neuroscience*, *38*(45), 9689–9699. https://doi.org/10.1523/JNEUROSCI.0251-18.2018

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Ben-Yakov, A., & Henson, R. N. (Eds.). (2018). The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. *The Journal of Neuroscience*, *38*(47), 10057–10068. https://doi.org/10.1523/JNEUROSCI.0524-18.2018

Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder. *ArXiv:1803.11175 [Cs]*. http://arxiv.org/abs/1803.11175

Chang, L., Fleetwood, G., Manning, J., Parkinson, C., Finn, E., Wager, T., Haxby, J., Geerligs, L., Vega, A. de la, Lahnakoski, J., Yeshurun, Y., Shappell, H., Yarkoni, T., Baldassano, C., & Shim, W. M. (2020). *Neuroimaging Analysis Methods For Naturalistic Data* (FZJ-2020-04493). Annual meeting of the Organization for Human Brain Mapping 2020. Zenodo.

https://doi.org/10.5281/ZENODO.3937849

Chang, L. J., Jolly, E., Cheong, J. H., Rapuano, K. M., Greenstein, N., Chen, P.-H. A., & Manning, J. R. (2021). Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Science Advances*, *7*(17), eabf7129. https://doi.org/10.1126/sciadv.abf7129

Chen, H., Yang, Q., Liao, W., Gong, Q., & Shen, S. (2009). Evaluation of the effective connectivity of supplementary motor areas during motor imagery using Granger causality mapping. *NeuroImage*, *47*(4), 1844–1853. https://doi.org/10.1016/j.neuroimage.2009.06.026

Chen, P.-H. A., Jolly, E., Cheong, J. H., & Chang, L. J. (2020). Intersubject representational similarity analysis reveals individual variations in affective experience when watching erotic movies. *NeuroImage*, *216*, 116851. https://doi.org/10.1016/j.neuroimage.2020.116851

Chen, P.-H. (Cameron), Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., & Ramadge, P. J. (2015). A Reduced-Dimension fMRI Shared Response Model. *Advances in Neural Information Processing Systems*, *28*. https://papers.nips.cc/paper/2015/hash/b3967a0e938dc2a6340e258630febd5a-Abstract.html

Chinier, E., N'Guyen, S., Lignon, G., Ter Minassian, A., Richard, I., & Dinomais, M. (2014). Effect of motor imagery in children with unilateral cerebral palsy: FMRI study. *PloS One*, *9*(4), e93378. https://doi.org/10.1371/journal.pone.0093378

Cohen, S. S., & Baldassano, C. (2021). Developmental changes in story-evoked responses in the neocortex and hippocampus. *BioRxiv*, 2021.04.12.439526. https://doi.org/10.1101/2021.04.12.439526

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

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

DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of Learning and Memory*, *134*, 107–114. https://doi.org/10.1016/j.nlm.2016.07.011

Ezzyat, Y., & Davachi, L. (2011). What Constitutes an Episode in Episodic Memory? *Psychological Science*, *22*(2), 243–252. https://doi.org/10.1177/0956797610393742

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

Finn, E. S., & Bandettini, P. A. (2020). *Movie-watching outperforms rest for functional connectivity-based prediction of behavior* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.08.23.263723

Finn, E. S., Corlett, P. R., Chen, G., Bandettini, P. A., & Constable, R. T. (2018). Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nature Communications*, *9*(1), 2043. https://doi.org/10.1038/s41467-018-04387-2

Finn, E. S., Glerean, E., Khojandi, A. Y., Nielson, D., Molfese, P. J., Handwerker, D. A., & Bandettini, P. A. (2020). Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, *215*, 116828. https://doi.org/10.1016/j.neuroimage.2020.116828

Gao, J., Chen, G., Wu, J., Wang, Y., Hu, Y., Xu, T., Zuo, X.-N., & Yang, Z. (2020). Reliability map of individual differences reflected in inter-subject correlation in naturalistic imaging. *NeuroImage*, *223*, 117277. https://doi.org/10.1016/j.neuroimage.2020.117277

Geerligs, L., Gözükara, D., Oetringer, D., Campbell, K., Gerven, M. van, & Güçlü, U. (2022). *A partially nested cortical hierarchy of neural states underlies event segmentation in the human brain* (p. 2021.02.05.429165). bioRxiv. https://doi.org/10.1101/2021.02.05.429165

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

Grall, C., & Finn, E. S. (2022). Leveraging the power of media to drive cognition: A media-informed approach to naturalistic neuroscience. Social Cognitive and Affective Neuroscience, 17(6), 598–608. https://doi.org/10.1093/scan/nsac019

Glerean, E., Pan, R. K., Salmi, J., Kujala, R., Lahnakoski, J. M., Roine, U., Nummenmaa, L., Leppämäki, S., Nieminen-von Wendt, T., Tani, P., Saramäki, J., Sams, M., & Jääskeläinen, I. P. (2016). Reorganization of functionally connected brain subnetworks in high-functioning autism. *Human Brain Mapping*, *37*(3), 1066–1079. https://doi.org/10.1002/hbm.23084

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology. General*, *140*(4), 586–604. https://doi.org/10.1037/a0024310

Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition*, *34*(6), 1221–1235. https://doi.org/10.3758/BF03193267

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences*, *19*(6), 304–313. https://doi.org/10.1016/j.tics.2015.04.006

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A Hierarchy of Temporal Receptive Windows in Human Cortex. *Journal of Neuroscience*, *28*(10), 2539–2550. https://doi.org/10.1523/JNEUROSCI.5487-07.2008

Hill, J., Dierker, D., Neil, J., Inder, T., Knutsen, A., Harwell, J., Coalson, T., & Van Essen, D. (2010). A Surface-Based Analysis of Hemispheric Asymmetries and Folding of Cerebral Cortex in Term-Born Human Infants. *Journal of Neuroscience*, *30*(6), 2268–2276. https://doi.org/10.1523/JNEUROSCI.4682-09.2010

Honey, C. J., Thesen, T., Donner, T. H., Silbert, L. J., Carlson, C. E., Devinsky, O., Doyle, W. K., Rubin, N., Heeger, D. J., & Hasson, U. (2012). Slow cortical dynamics and the accumulation of information over long timescales. *Neuron*, *76*(2), 423–434. https://doi.org/10.1016/j.neuron.2012.08.011

Jafarpour, A., Buffalo, E. A., Knight, R. T., & Collins, A. G. E. (2022). Event segmentation reveals working memory forgetting rate. *IScience*, *25*(3), 103902. https://doi.org/10.1016/j.isci.2022.103902

Kumar, M., Michael Anderson, Antony, J., Baldassano, C., Brooks, P., Cai, M. B., Chen, P.-H. C., Ellis, C. T., Henselman-Petrusek, G., Huberdeau, D., Hutchinson, B., Li, P. Y., Lu, Q., Manning, J. R., Mennen, A. C., Nastase, S. A., Richard, H., Schapiro, A. C., Schuck, N. W., … Norman, K. A. (2020). *BrainIAK: The Brain Imaging Analysis Kit*. OSF Preprints. https://doi.org/10.31219/osf.io/db2ev

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Lahnakoski, J. M., Glerean, E., Jääskeläinen, I. P., Hyönä, J., Hari, R., Sams, M., & Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*, *100*, 316–324. https://doi.org/10.1016/j.neuroimage.2014.06.022

Lee, H., & Chen, J. (2021). Narratives as Networks: Predicting Memory from the Structure of Naturalistic Events. *BioRxiv*, 2021.04.24.441287. https://doi.org/10.1101/2021.04.24.441287

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. *The Journal of Neuroscience*, *31*(8), 2906–2915. https://doi.org/10.1523/JNEUROSCI.3684-10.2011

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2), 209–220.

Masís-Obando, R., Norman, K. A., & Baldassano, C. (2021). *Schema representations in distinct brain networks support narrative memory during encoding and retrieval* (p. 2021.05.17.444363). https://doi.org/10.1101/2021.05.17.444363

Masís-Obando, R., Norman, K. A., & Baldassano, C. (2022). Schema representations in distinct brain networks support narrative memory during encoding and retrieval. *ELife*, *11*, e70445. https://doi.org/10.7554/eLife.70445

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T. T., Sepulcre, J., Sabuncu, M. R., Shafee, R., Lu, J., & Liu, H. (2013). Individual Variability in Functional Connectivity Architecture of the Human Brain. *Neuron*, *77*(3), 586–595. https://doi.org/10.1016/j.neuron.2012.12.028

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

Newtson, D., Engquist, G. A., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology*, *35*(12), 847–862. https://doi.org/10.1037/0022-3514.35.12.847

Newtson, D., Harvey, J. H., Ickes, W. J., & Kidd, R. F. (1976). *New directions in attribution research*. Lawrence Erlbaum Associates Hillsdale, New Jersey:

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Raccah, O., Doelling, K. B., Davachi, L., & Poeppel, D. (2022). Acoustic features drive event segmentation in speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001150

Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition* (pp. ix, 272). Oxford University Press.

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, N.Y.: 1991)*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Schwan, S., Garsoffky, B., & Hesse, F. W. (2000). Do film cuts facilitate the perceptual and cognitive organization of activitiy sequences? *Memory & Cognition*, *28*(2), 214–223.

https://doi.org/10.3758/BF03213801

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). https://doi.org/10.1073/pnas.2021905118

Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading Stories Activates Neural Representations of Visual and Motor Experiences. *Psychological Science*, *20*(8), 989–999. https://doi.org/10.1111/j.1467-9280.2009.02397.x

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

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human Brain Activity Time-Locked to Narrative Event Boundaries. *Psychological Science*, *18*(5), 449–455. https://doi.org/10.1111/j.1467-9280.2007.01920.x

Vanderwal, T., Eilbott, J., Finn, E. S., Craddock, R. C., Turnbull, A., & Castellanos, F. X. (2017). Individual differences in functional connectivity during naturalistic viewing conditions. *NeuroImage*, *157*, 521–530. https://doi.org/10.1016/j.neuroimage.2017.06.027

Yeshurun, Y., Nguyen, M., & Hasson, U. (2021). The default mode network: Where the idiosyncratic self meets the shared social world. *Nature Reviews Neuroscience*, *22*(3), 181–192. https://doi.org/10.1038/s41583-020-00420-w

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017). Same Story, Different Story: The Neural Representation of Interpretive Frameworks. *Psychological Science*, *28*(3), 307–319. https://doi.org/10.1177/0956797616682029

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*(6), 979–1008. https://doi.org/10.1207/s15516709cog2806_5

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries.

Nature Neuroscience, *4*(6), 651–655. https://doi.org/10.1038/88486

Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition*, *112*(2), 201–216. https://doi.org/10.1016/j.cognition.2009.03.007

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., & Swallow, K. M. (2007). Event Segmentation. *Current Directions in Psychological Science*, *16*(2), 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x

Zalla, T., Verlut, I., Franck, N., Puzenat, D., & Sirigu, A. (2004). Perception of dynamic action in patients with schizophrenia. *Psychiatry Research*, *128*(1), 39–51. https://doi.org/10.1016/j.psychres.2003.12.026