PNAS

Supplementary Information for

Neural event segmentation of continuous experience in human infants

Tristan S. Yates, Lena J. Skalaban, Cameron T. Ellis, Angelika J. Bracher, Christopher Baldassano, and Nicholas B. Turk-Browne

Corresponding Author: Nicholas Turk-Browne.
E-mail: nicholas.turk-browne@yale.edu

This PDF file includes:

Supporting Information Text

Control analyses of intersubject correlation across timepoints and movies. To further evaluate the significance of neural similarity across participants, we calculated two types of chance or baseline intersubject correlation (ISC) values. First, we created a permuted rotation distribution, where the data from one participant were time-shifted for every possible shift value between one and the length of the movie (wrapping around to the beginning) and then correlated with the average of the data from all other participants at the correct timing. We averaged over held-out participants for each possible shift value and calculated significance as the proportion of time shifts where the permuted ISC value was greater than the actual ISC value (correct timing), doubled to make it two-tailed. This analysis quantifies the extent to which neural synchrony is higher when participants are temporally aligned versus misaligned. For adults (Figure S1A, left), all ROIs showed significantly higher ISC values when neural activity was aligned vs. misaligned (ps < 0.006). For infants (Figure S1A, right), this held in all but one ROI (ps < 0.011; EAC: $p = 0.117$). These results demonstrate that the main ISC findings we report (Figure 1) depend upon synchronized viewing of the movie.

Second, we re-calculated ISC between participants who watched different movies. For each participant who watched the Aeronaut movie, we correlated their neural activity with the average neural activity of different participants who watched the Mickey movie (see below for more details about this movie). Because Aeronaut was longer than Mickey (90 vs. 71 TRs), we clipped the Aeronaut data to the first 71 TRs for this analysis. Through comparison with the main ISC analysis based only on Aeronaut (Figure 1), this analysis tests whether neural synchrony is higher when participants view the same versus a different movie. For adults (Figure S1B, left), different-movie ISC values were significantly greater than zero in EVC ($p < 0.001$), LOC ($p < 0.001$), hippocampus ($p = 0.005$), and EAC ($p = 0.005$), but not in AG, PCC, precuneus, or mPFC (ps > 0.230); the significant values in some ROIs may reflect a generic correlation from the synchronized onset of visual stimulation at the start of both movies, unrelated to their specific content. Critically, the different-movie ISC values in all ROIs were significantly *lower* than when the adults being compared watched the same Aeronaut movie (ps < 0.024). For infants (Figure S1B, right), no ROI showed significant different-movie ISC values (ps >0.256) and these values in almost all ROIs were again significantly lower than the main same-movie ISC values (ps < 0.040; AG: $p = 0.068$, EAC: $p = 0.432$). These results show that the main ISC findings depend upon matching visual content.

Log-likelihood simulations. To assess whether the log-likelihood metric would be biased to higher or lower numbers of events, we tested how well we could recover event structure in simulated data. We first generated event patterns (voxels by number of events) with values drawn from a standard normal distribution. Because each voxel was treated as an independent source, we used fewer voxels (5) than our actual analyses to better simulate the correlated patterns present in real fMRI data. Event labels were assigned to each of 90 timepoints. We generated 24 "participants" by applying the simulated event patterns to each timepoint with an additional noise component (cov: the covariance matrix of a multivariate normal distribution). The resulting voxel by timepoint matrices were convolved with a double-gamma hemodynamic response function (HRF) using an fMRI simulation package (1) available in BrainIAK (2). We followed the same analysis approach described in the Methods and Materials section to estimate the optimal number of events. We calculated model error as the difference between the actual simulated number and the estimated optimal number. This was repeated across a range of different possible event numbers (from 2 to 21).

With low noise (cov = 0.1), the timepoint-by-timepoint similarity matrices showed clear block structure along the diagonal. Average error between model estimates and the correct number of events was negative (M = -0.931, $p = 0.003$), meaning that the model under-estimated the number of events (Figure S2A). When noise increased to a moderate level (cov = 2), model error did not significantly differ from zero (M = -0.008, $p = 0.980$), that is, it did not under- or over-estimate the number of events. With high noise (cov = 20), model error was positive ($M = 1.40$, $p < 0.001$), indicating that the model over-estimated the number of events.

We also explored whether this same pattern holds depending on the shape of the hemodynamic response function. In young infants, the HRF can be delayed and smaller in magnitude than adults (3). Thus, we separately varied the time-to-peak and magnitude of the HRF response before running the simulations above. Across parameter values, the model continued to under-estimate the number of events at a low noise level and over-estimate the number of events at a high noise level, with model error hovering near zero for moderate noise (Figure S2B).

Alternative behavioral boundary approach. We tested whether voxel activity patterns for timepoints within a behavioral boundary were more correlated than timepoints spanning a boundary by considering all possible pairs of timepoints within and across boundaries up to the temporal distance of the largest event. This approach has the advantage of using as many timepoint pairs as possible for calculating within vs. across boundary correlations, but may be vulnerable to increased noise due to comparing timepoints from different parts of the movie. Here we report a more conservative approach for testing how behavioral boundaries predict neural data by only considering timepoint pairs that are equated in temporal distance (as before) but additionally "anchored" to the same timepoint. Namely, for each TR we measured the correlation between the spatial activity pattern at that timepoint and timepoints forward and backward in time at a matched temporal distance. If one timepoint pair was within an event and the other was across an event, we calculated the within minus across boundary correlation. However, if both timepoint pairs were either within an event or across an event, or if one of the timepoint pairs was already included in a different calculation, the within vs. across boundary correlation was not included. We performed this for each temporal distance up to the length of the largest event and calculated the average within vs. across boundary correlation for each subject. For statistical analysis, we used the same bootstrap resampling techniques described in the Methods.

In adults, the searchlight analysis again showed that visual regions and posterior cingulate exhibited significantly greater pattern similarity within vs. across behavioral event boundaries (Figure S4A). In fact, the voxelwise map of the average within vs. across boundary correlation was highly correlated with the map using our main approach ($r = 0.818$). Nonetheless, fewer regions emerged from this more conservative behavioral boundary approach. In infants, we also found several regions showing greater pattern similarity within vs. across behavioral boundaries in infants, where again the voxelwise map was highly correlated with our main results ($r = 0.893$).

In adults, we surprisingly found greater pattern similarity for *across* vs. within timepoint pairs in LOC (M = -0.018, CI = [-0.026, -0.008], $p < 0.001$); the other regions were insensitive to the behavioral boundaries (EVC: $M = 0.001$, CI = [-0.009, 0.011], $p = 0.836$; AG: M = -0.015, CI = [-0.034, 0.004], $p = 0.114$; PCC: M = -0.022, CI = [-0.048, 0.001], $p = 0.060$; precuneus: $M = 0.008$, CI = [-0.016, 0.031], $p = 0.510$; mPFC: $M = 0.002$, CI = [-0.008, 0.011], $p = 0.686$); hippocampus: $M = 0.005$, CI = [-0.001, 0.012], $p = 0.114$; EAC: $M = 0.004$, CI = [-0.005, 0.014], $p = 0.352$). In infants, we did not find sensitivity to the behavioral boundaries (EVC: $M = 0.006$, CI = [-0.026, 0.048], $p = 0.750$; LOC: M = -0.019, CI = [-0.053, 0.016], $p = 0.342$; AG: $M = 0.014$, CI = [-0.031, 0.061], $p = 0.578$; PCC: $M = 0.012$, CI = [-0.018, 0.045], $p = 0.428$; precuneus: M = -0.012, CI = [-0.035, 0.010], $p = 0.320$; mPFC: M = -0.019, CI = [-0.053, 0.021], $p = 0.358$; hippocampus: $M = 0.027$, CI = [-0.004, 0.065], $p = 0.094$; EAC: $M = 0.025$, CI = [-0.008, 0.062], $p = 0.162$). Overall, the pattern of results was weaker, but nonetheless fairly consistent across the two approaches.

Pattern similarity drift approaching behavioral boundaries. In our main analysis relating behavioral boundaries and neural data, we considered all timepoints within an event. However, this assumes an "all-or-none" relationship between behavioral boundaries and neural data, whereas previous work has shown that neural event patterns can also change gradually (4, 5). Thus, we also explored how neural pattern similarity evolves over time, and how this relates to proximity to behavioral event boundaries. To conduct this analysis, we first calculated the absolute distance of each timepoint to the nearest event boundary. Then for each participant, we calculated the pattern similarity between adjacent time points, and correlated this with distance to the event boundary (Figure S5A). The logic of this analysis was that adjacent timepoints near a boundary may have lower similarity than adjacent timepoints farther from a boundary (in the middle of an event). We calculated significance using the same bootstrap resampling procedure described in the Methods.

In adults, a searchlight analysis revealed evidence of pattern similarity change as a function of distance to an event boundary in several regions throughout the brain, including visual and frontal regions. The unthresholded whole-brain map showed some correlation with our previous, discrete approach ($r = 0.271$). In the ROIs, we found that pattern similarity increased with distance from behavioral boundaries in LOC (relationship of pattern similarity between adjacent timepoints and absolute distance from nearest boundary: $M = 0.102$, CI = [0.060, 0.148], $p < 0.001$), AG ($M = 0.082$, CI = [0.033, 0.136], $p < 0.001$), mPFC ($M = 0.086$, CI = [0.035, 0.132], $p < 0.001$), hippocampus ($M = 0.058$, CI = [0.017, 0.096], $p = 0.006$) and EAC ($M = 0.064$, CI = [0.016, 0.117], $p = 0.018$). These results mirrored our main (discrete) behavioral boundary approach in hippocampus and EAC, whereas the precuneus was only significant in the discrete version and LOC, AG, and mPFC were only significant in the continuous version. These latter regions may exhibit less abrupt changes at behavioral boundaries, instead showing pattern similarity drift as behaviorally defined events unfold. In infants, a searchlight analysis relating neural pattern similarity and distance to behavioral boundaries revealed visual regions, supramarginal gyrus, and right lateral frontal cortex. Again, the unthresholded whole-brain map showed some correlation with our previous analysis ($r = 0.262$). For the ROIs, only one region came out as significant: EVC ($M = 0.057$, CI = [0.005, 0.109], $p = 0.040$). In infants, too, some regions may respond abruptly to behavioral event boundaries (EAC, from discrete version) whereas others may drift more gradually (EVC, from continuous version), although the particular regions involved differ from adults. This more continuous approach to relating behavioral boundaries and neural data provides a different angle on the question of how behavior is instantiated in the brain.

Of course, this analysis assumes a *linear* relationship between pattern similarity and distance to behavioral boundaries, and that the slopes are *symmetric* before and after behavioral boundaries (because distance is unsigned). However, regions may show nonlinear changes in pattern similarity around boundaries even if symmetric (e.g., U-shaped trough or inverted U-shaped peak), or asymmetric linear changes (e.g., decrease/increase before boundary but no change after, or vice versa). Thus, we also plotted the change in adjacent timepoint pattern similarity from 5 TRs before to 5 TRs after behavioral event boundaries for each ROI and age group (Figure S6).

We tested for a symmetric U-shaped (or inverted) change by fitting a quadratic model over this whole window (-5 to 5 TRs) and for asymmetric linear changes by fitting linear models over shorter windows on either side of the boundary (before: -5 to 0 TRs; after: 0 to 5 TRs). To determine significance, we resampled participant data with replacement 1000 times and defined the p-value as the proportion of quadratic or linear beta values with the opposite sign of the true effect, doubled to make the test two-tailed. For adults, a quadratic model significantly fit the data in LOC ($p = 0.002$) and EAC ($p = 0.014$), with the minimum value near zero for LOC and slightly positive for EAC. There was a significant decrease in pattern similarity when approaching the boundary in mPFC ($p = 0.004$), hippocampus ($p = 0.032$), and EAC ($p = 0.002$), and a significant increase in pattern similarity emerging from the boundary in EVC ($p = 0.048$), LOC ($p = 0.018$), and precuneus ($p = 0.018$). This analysis, combined with the results above, shows that adult behavioral event boundaries relate to adult neural pattern similarity in most of the ROIs tested. For infants, however, neither a quadratic model nor a pre-boundary linear model fit the data in any ROI. A post-boundary linear model fit the data in LOC ($p = 0.022$) and precuneus ($p = 0.032$), with a significant *decrease* in pattern similarity emerging from the boundary. This reliable but opposite pattern from adults (i.e., in which these same two regions showed a post-boundary increase), provides further evidence that adult behavioral event boundaries capture some variance in infant neural states but in a way that differs from the adult brain.

Directly relating behavioral and neural boundaries. Behavioral event boundaries may be represented in both the adult and infant brain (as shown in the main text), but yet not be picked up by the HMM model in its optimal neural event segmentation. Thus, to further understand how behavioral boundaries manifest in the brain, we examined the alignment between neural and behavioral event boundaries. For this analysis, we calculated the correlation between a continuous measure of boundary probability from the HMM and a continuous measure of boundary probability from adult behavior. Specifically, in a given region, we fit an HMM model with the optimal number of events found in our previous analysis, and then extracted a continuous measure of event change at every timepoint by taking the derivative of the expected event value timecourse (6). To generate a robust estimate of when boundaries were likely vs unlikely to occur throughout our fMRI sample, we bootstrapped this fit 1000 times (resampling participants with replacement and re-fitting the model) and averaged the event probability timecourse across all 1000 iterations. Our continuous measure of behavioral events was the proportion of behavioral participants who denoted an event boundary at each timepoint, convolved with a canonical HRF (to account for when the behavioral event would have impacted fMRI data). We then simply calculated the Pearson correlation of these two timecourses (Figure S7A). To determine significance, we created a null distribution by time-shifting the average event probability matrix for every shift between 1 and the movie length (wrapping around to the beginning). The p-value was calculated as the proportion of time-shifted permutations with values as extreme as the actual correlation, doubled to make it two-tailed.

In adults, neural and behavioral event boundary timecourses were significantly correlated in the hippocampus (r = 0.291, p = 0.044; Figure S7B); no other ROIs were significant (EVC: r = -0.041, p = 0.800; LOC: r = -0.081, p = 0.622; AG: r = -0.071, p = 0.800; PCC: r = 0.188, p = 0.311; precuneus: r = 0.024, p = 0.756; mPFC: r = 0.043, p = 0.711; EAC: r = -0.104, p = 0.822). Thus, whereas other regions (including the precuneus and EAC in the discrete analysis) represented behavioral event boundaries, only in the hippocampus were these related to the optimal neural event structure. In contrast, adult behavior was not significantly correlated with neural event structure from any ROI in infants (EVC: r = -0.157, p = 0.378; LOC: r = 0.061, p = 0.667; AG: r = -0.110, p = 0.600; PCC: r = -0.015, p = 0.889; precuneus: r = -0.034, p = 0.844; mPFC: r = -0.203, p = 0.267; hippocampus: r = -0.018, p = 0.956; EAC: r = -0.042, p = 0.778). Thus, although adult behavioral event boundaries can be represented in the infant brain in terms of pattern similarity, they are not the driving force behind their neural event structure.

Mickey dataset. We applied the intersubject correlation and event segmentation analyses to a second, previously collected convenience sample of infants watching a movie between other tasks. Although this dataset has some flaws — fewer participants, much wider age range up to three years, potential for pre-experimental familiarity with the characters, smaller stimulus display, and a mix of scanners — we felt these analyses might provide an initial assessment of generalizability to another movie. fMRI data were collected from 15 sessions with infants aged 4.00 – 32.60 months (M = 13.92, SD = 8.87; 9 female) watching a silent cartoon lasting 142 s ("Mickey"). This movie was shown on a smaller display than the Aeronaut movie (~25% of area), spanning 22.75 visual degrees in width and 12.75 visual degrees in height. In this video, a surprise party is thrown where Disney characters dance and play the piano while one character makes an exploding cake in the kitchen. Two infants participated twice after a delay (6.3 months and 2.3 months difference) and were treated as independent sessions. As before, additional infants with head motion above 3 mm (N = 5) or eyes off-screen (N = 2) for more than half of the movie were excluded. We also collected data from 15 adults aged 19 – 27 years (M = 21.47, SD = 2.90; N = 10 female). All adults and 9 infants watched the movie twice in a row within session. For these participants, data were averaged across the two viewings. Infants were collected at the Scully Center at Princeton University (N = 7) and the Magnetic Resonance Research Center (MRRC) at Yale University (N = 8). Adult participants were collected at the Brain Imaging Center (BIC) at Yale University. This study was approved by the Institutional Review Board at Princeton University and the Human Investigation Committee (MRRC) and Human Subjects Committee (BIC) at Yale University. Adults provided informed consented for themselves or their child.

Data acquisition, preprocessing, and analyses were identical to the Aeronaut dataset with two variations: First, infant data were acquired at Princeton using a Siemens Skyra (3T) MRI. Second, functional images for infants were collected under a slightly different functional EPI sequence (TE = 28ms, slices = 36). Adults were collected with the same functional sequence as Aeronaut. Gaze coding was highly reliable: coders reported the same response on an average of 91.1% of frames (SD = 5.08%; range = 78.6–98.4%). The average proportion of TRs retained after exclusion for looking off screen was high in adults (M = 99.5%, SD = 1.01%; range = 96.5–100%) and infants (M = 89.0%, SD = 13.1%; range = 58.5–100%). Eye-tracking data were not collected for one infant because of experimenter error. Timepoints with less than 3mm of translational motion were included (infants: M = 91.3%, SD =10.6%, range = 64.8%–100%; adults: 100%).

In the whole-brain analysis, ISC was strongest in visual regions for adults and infants (Figure S8). In the ROI analysis, adult ISC was significant in EVC (M = 0.472, CI = [0.423, 0.518], $p < 0.001$), LOC (M = 0.384, CI = [0.332, 0.438], $p < 0.001$), AG (M = 0.150, CI = [0.121, 0.186], $p < 0.001$), PCC (M = 0.188, CI = [0.113, 0.268], $p < 0.001$), precuneus (M = 0.228, CI = [0.178, 0.283], $p < 0.001$) and EAC (M = 0.045, CI = [0.007, 0.085], p = 0.022); but not significant in mPFC (M = 0.033, CI = [-0.003, 0.068], p = 0.070) or hippocampus (M = 0.012, CI = [-0.009, 0.030], p = 0.255). Infant ISC was significant in EVC (M = 0.117, CI = [0.084, 0.150], $p < 0.001$) and LOC (M = 0.074, CI = [0.050, 0.099], p = 0.001); and not significant in AG (M = 0.004, CI = [-0.032, 0.038], p = 0.865), PCC (M = -0.049, CI = [-0.101, 0.005], p = 0.068), precuneus (M = 0.002, CI = [-0.033, 0.037], p = 0.897), mPFC (M = 0.033, CI = [-0.002, 0.067], p = 0.062), hippocampus (M = 0.032, CI = [-0.023, 0.081], p = 0.232), or EAC (M = -0.022, CI = [-0.048, 0.002], p = 0.098).

In the searchlight analysis, we applied an HMM to one half of adult or infant participants using a range of event numbers from 2 to 18 and then tested on the second half. This maximum number of events ensured that at least some events were several TRs long, but was less than Aeronaut because Mickey was shorter. Log-likelihood was again used to assess model fit.

Similar to our main analyses, visual regions of the adult brain had more events than higher-order regions, although there were fewer events overall (Figure S9). We replicated the flattened gradient of event processing in infants, with the optimal number of events generally low across the brain.

In the nested analysis for reliability of event structure, most ROIs were significant in adults, including EVC (M = 3.99, CI = [3.68, 4.32], $p < 0.001$), LOC (M = 4.36, CI = [3.86, 4.79], $p < 0.001$), AG (M = 1.60, CI = [0.509, 2.80], $p = 0.008$), PCC (M = 3.30, CI = [2.61, 3.90], $p < 0.001$), precuneus (M = 4.65, CI = [4.19, 5.07], $p < 0.001$), and EAC (M = 2.57, CI = [1.37, 3.76], $p < 0.001$); but not mPFC (M = 0.007, CI = [-1.31, 1.17], $p = 0.996$) or hippocampus (M = 0.222, CI = [-0.902, 1.42], $p = 0.780$). In infants, reliable event structure was found in EVC (M = 2.56, CI = [1.32, 3.78], $p < 0.001$), LOC (M = 3.31, CI = [2.23, 4.32], $p < 0.001$), precuneus (M = 3.21, CI = [2.19, 4.15], $p < 0.001$), mPFC (M = 2.49, CI = [1.25, 3.51], $p < 0.001$), and hippocampus (M = 1.17, CI = [0.453, 1.86], $p = 0.002$) and not in AG (M = 0.585, CI = [-0.263, 1.41], $p = 0.196$), PCC (M = -0.267, CI = [-1.12, 0.686], $p = 0.574$) or EAC (M = 0.669, CI = [-0.581, 1.92], $p = 0.296$). By generalizing to a different movie, these results provide additional evidence for coarser event representations in infancy.

Nonlinear alignment. In our main analyses, we used a linear alignment procedure for infant anatomical images (with manual adjustments). However, dramatic developmental differences within and across ages raise the possibility that a nonlinear approach may be more appropriate. We thus used ANTS (7) to re-align infant and adult brain data from the Aeronaut dataset to adult anatomical data. We repeated the whole-brain ISC analyses and the searchlight analyses of optimal event number. The results were unchanged: ISC was highest in visual regions for both adults and infants (Figure S10A) and there was a gradient in the optimal number of events in adults but not infants (Figure S10B). Thus, our results are robust to these procedures for aligning between infant and adult brains.

Supplementary Figures

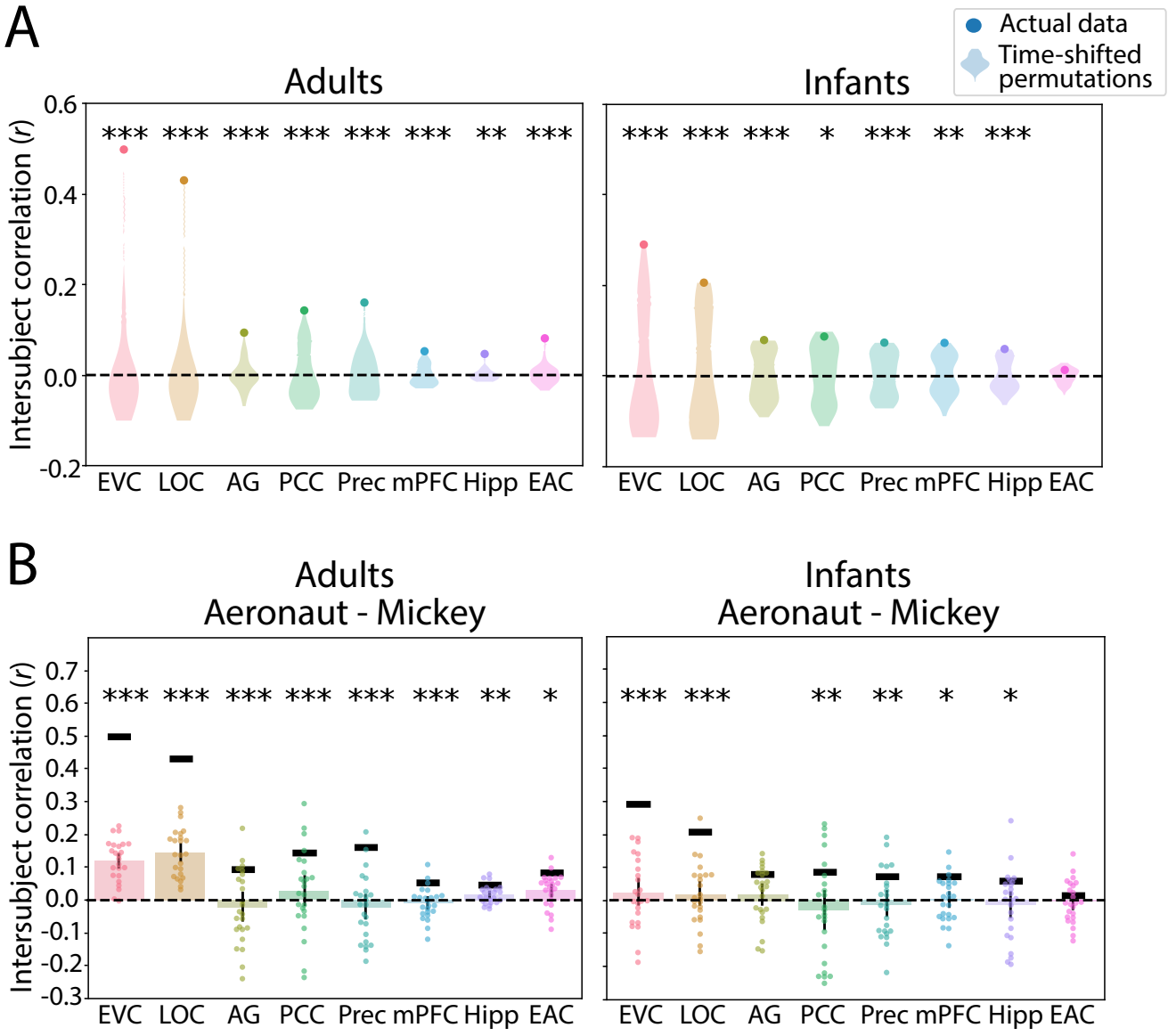

Fig. S1. Control analyses comparing intersubject correlation (ISC) values across timepoints and movies. (A) Average leave-one-out ISC in adults and infants from the actual data (dot) vs. a permuted time-shifted distribution. In all ROIs except infant EAC, ISC is significantly higher when temporally aligned. (B) Average different-movie ISC in adults and infants. Each dot represents ISC between a participant who watched the Aeronaut movie and the average of all participants from the same age group who watched the Mickey movie. Solid black lines indicate the average same-movie ISC values (from Figure 1). Although some different-movie ISC values were above zero in adults, the values in all ROIs were significantly lower than the main same-movie ISC results, except for infant AG and EAC. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

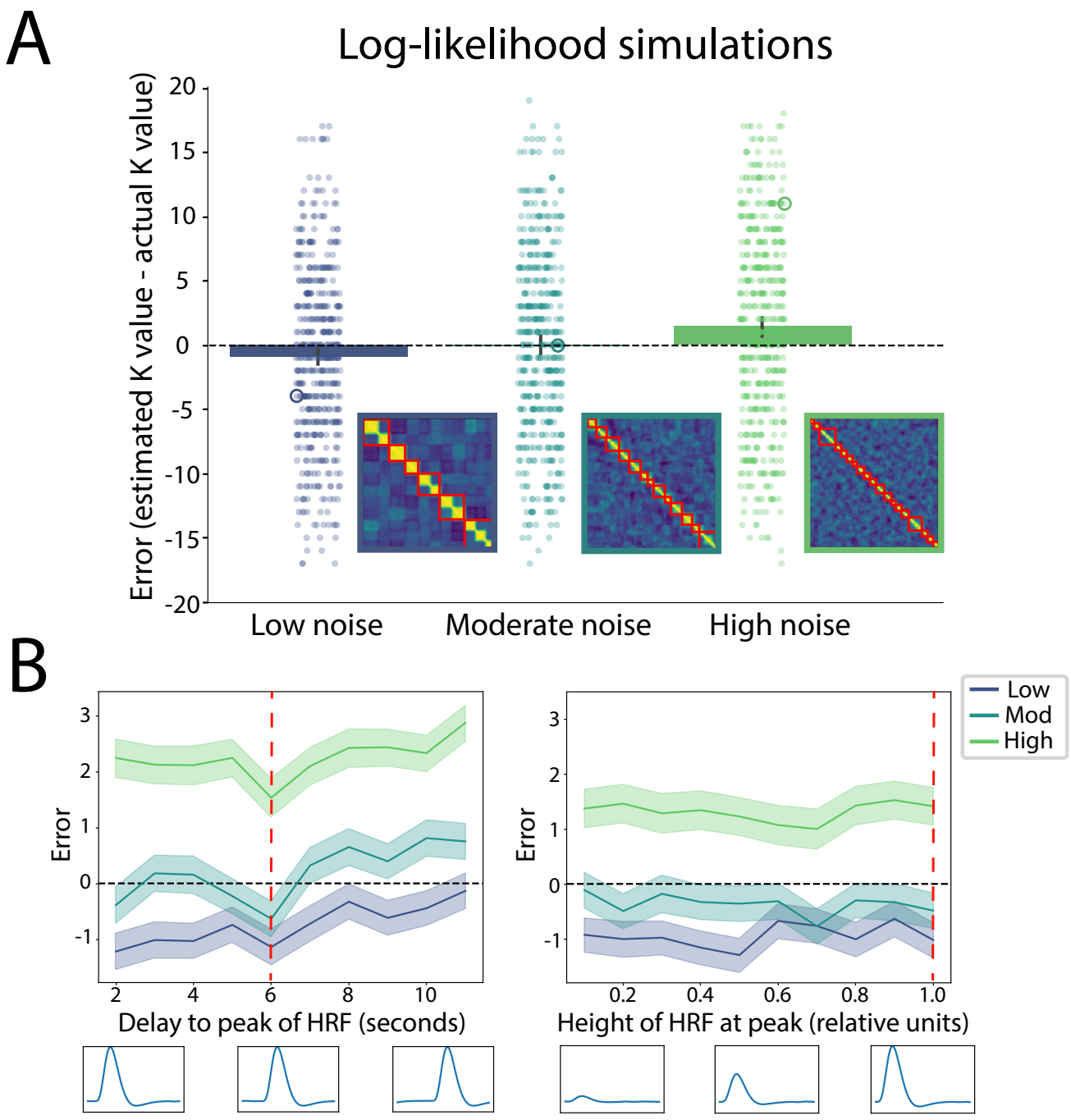

Fig. S2. Simulations of log-likelihood metric under different noise regimes. (A) Dots represent differences between the K values that maximized the model log-likelihood and the actual simulated K values for different iterations. One example error value for each noise regime is circled, and an example corresponding timepoint-by-timepoint correlation matrix is inset. Boundaries demarcating the model's best estimated K value are shown in red. In general, K values were underestimated when noise was low, guessed correctly when noise was moderate, and overestimated when noise was high. (B) This same pattern emerged when varying the delay to the peak of the hemodynamic response function (HRF) and the height of the HRF at peak. The red dashed lines indicate the parameter values used for the simulations in panel A (exact error values may differ between panels because of randomization in the simulations).

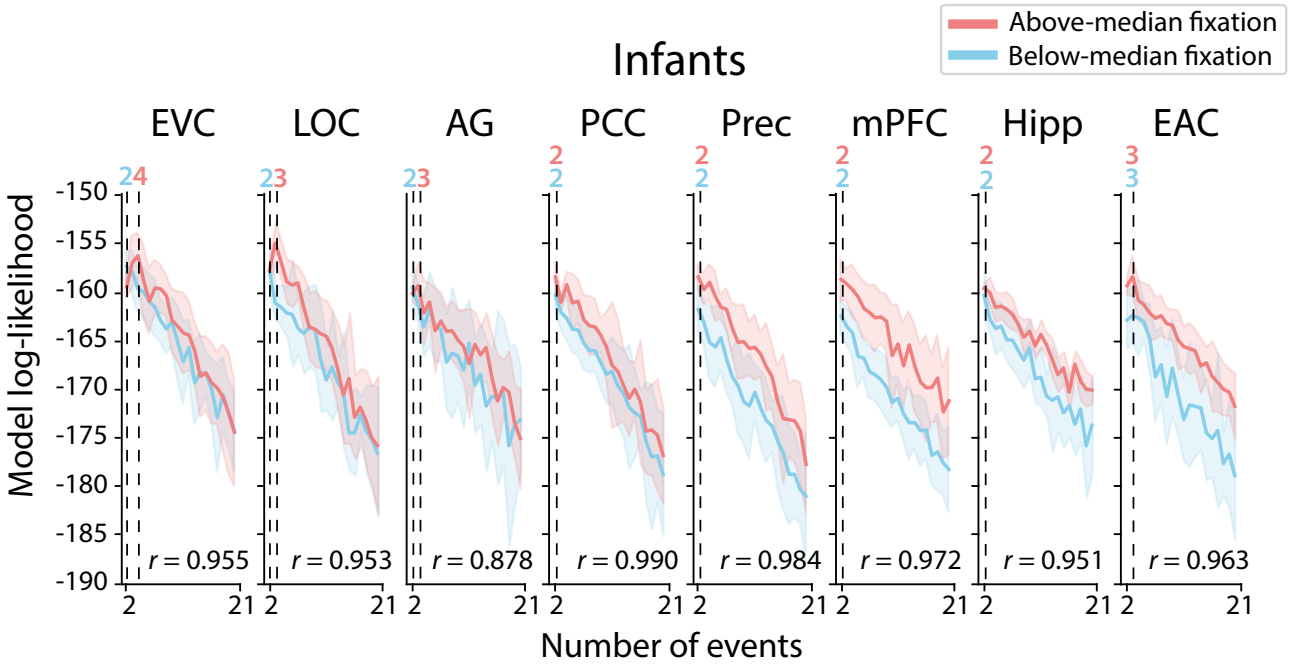

Fig. S3. Model log-likelihood for each ROI across different numbers of events in split halves of infants who were above (red) or below (blue) the median amount of looking to the screen. The correlation of log-likelihoods between above- and below-median fixation groups across event numbers is reported for each ROI. Shaded areas denote the standard deviation across iterations of model fitting. Dashed lines indicate the optimal number of events in the above- and below-median fixation groups for each ROI; this number is labeled above the dashed lines in each group's corresponding color.

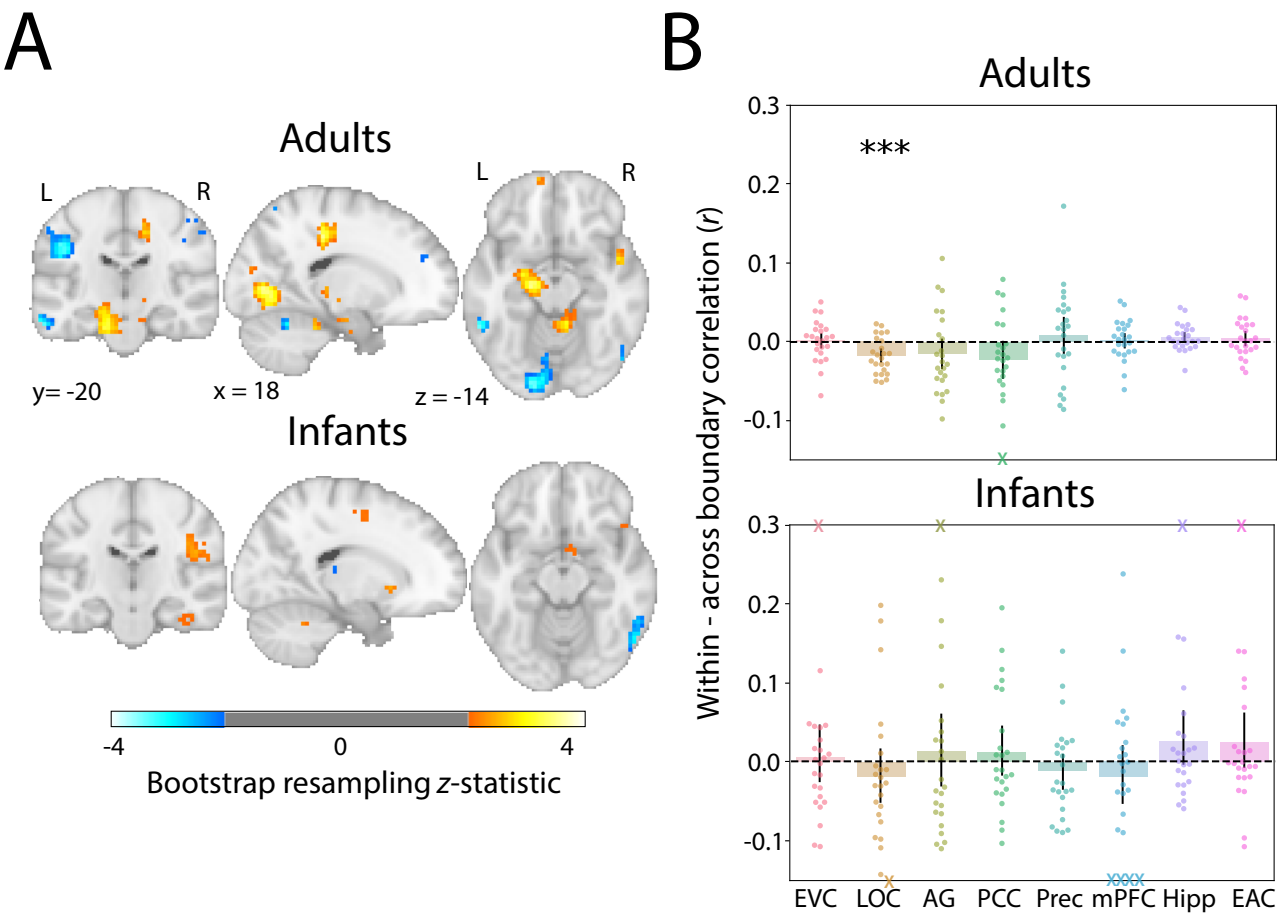

Fig. S4. Alternative approach relating behavioral boundaries to neural activity. (A) Whole-brain searchlight analysis for each age group comparing pattern similarity between timepoints drawn from within vs. across behavioral event boundaries when pairs of correlations are anchored to the same timepoint. Bootstrapped *z*-scores are thresholded at $p < 0.05$, uncorrected. (B) ROI analysis of difference in pattern similarity within minus between behavioral events. Dots represent individual participants and error bars represent 95% CIs of the mean from bootstrap resampling. One adult participant with a value beyond the y-axis range for PCC is indicated with an X at the negative edge. Infant participants with values beyond the y-axis range for EVC, AG, hippocampus, and EAC are indicated with Xs at the positive edge, and for LOC and mPFC with Xs at the negative edge. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

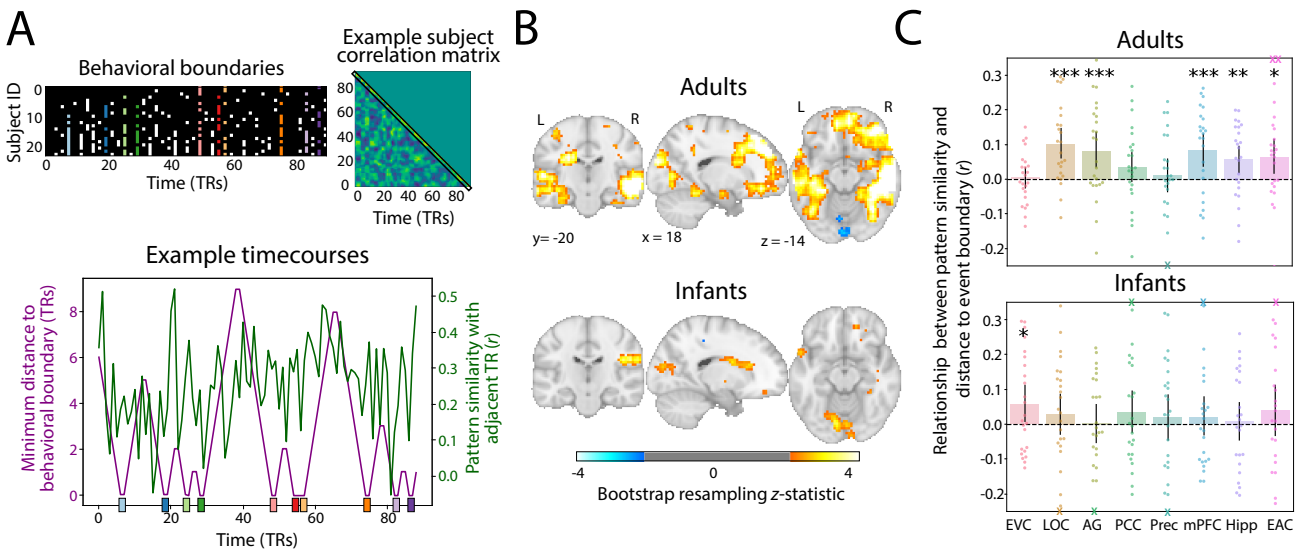

Fig. S5. Continuous approach for relating behavioral boundaries to neural pattern similarity. (A) Top left: matrix showing which behavioral participants indicated the presence of an event boundary at each TR in the movie (from Figure 5A). Top right: example timepoint-by-timepoint correlation matrix of similarity between patterns of neural activity in a given brain region for one participant, where the first off-diagonal represents the similarity between adjacent timepoints. Bottom: timecourse of pattern similarity for adjacent timepoints (green) and minimum distance to a behavioral boundary (purple). Rectangles on the x-axis indicate the 10 different behavioral event boundaries. (B) Whole-brain searchlight analysis for each age group relating pattern similarity between adjacent timepoints and their absolute distance to the nearest behavioral boundary. Bootstrapped *z*-scores are thresholded at $p < 0.05$, uncorrected. (C) ROI analysis of pattern similarity changes as a function of distance to the boundary. Positive values reflect greater pattern similarity farther from the nearest boundary. Dots represent individual participants and error bars represent 95% CIs of the mean from bootstrap resampling. Xs on the negative or positive edge indicate individual participants with a value beyond the y-axis range. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

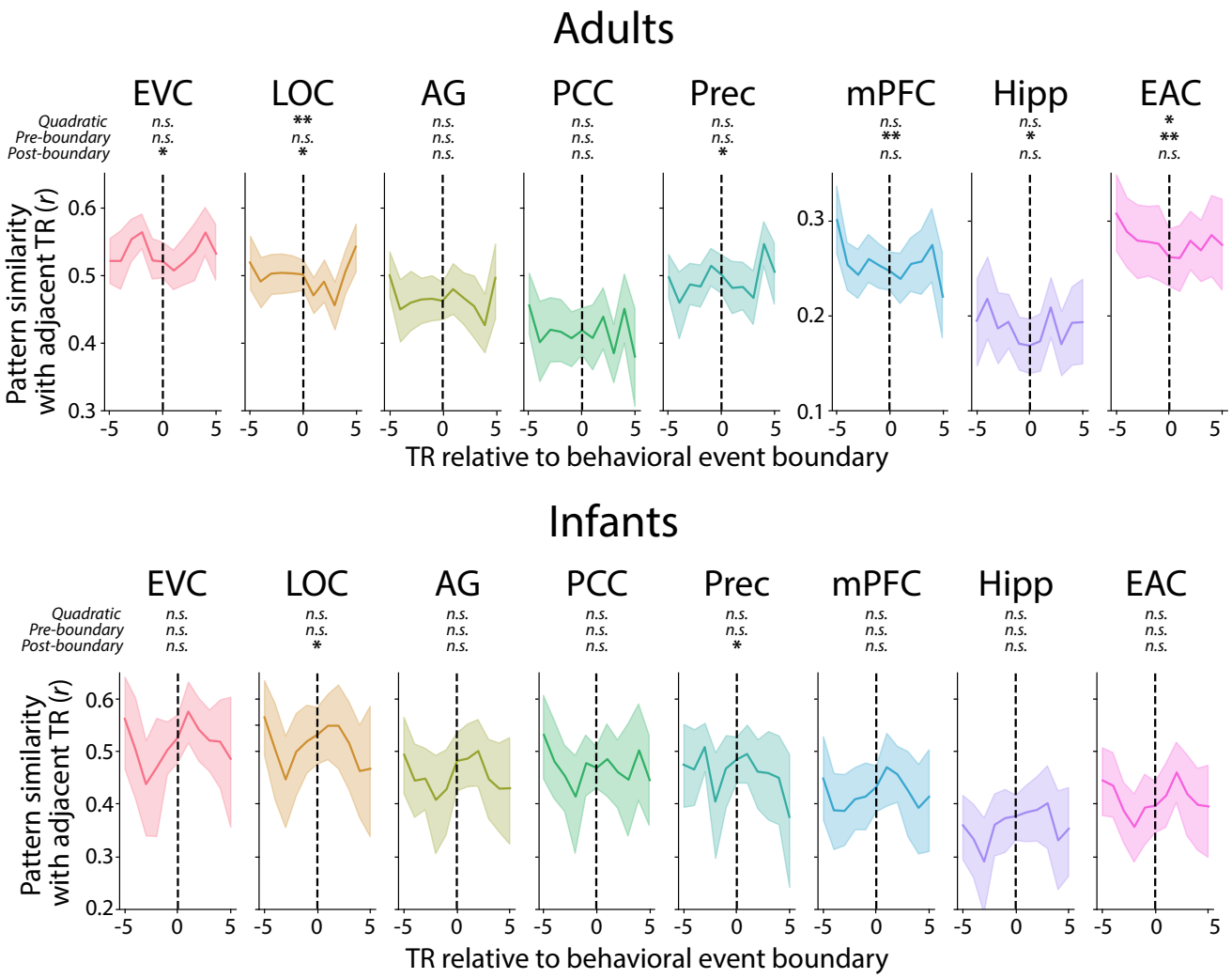

Fig. S6. Pattern similarity between adjacent timepoints before and after behavioral event boundaries (dashed line). The shaded area represents the 95% confidence interval determined through bootstrap resampling. The fits of a quadratic model, pre-boundary linear model, and post-boundary linear model are depicted at the top of the panel for each ROI and group. In adults, pattern similarity showed a reliable quadratic effect (U-shaped trough) at behavioral event boundaries in LOC and EAC, linear effect (decrease) before boundaries in mPFC, hippocampus, and EAC, and linear effect (increase) after boundaries in EVC, LOC, and precuneus. In infants, pattern similarity showed a reliable linear effect (decrease) after behavioral event boundaries in LOC and precuneus. ** $p < 0.01$, * $p < 0.05$.

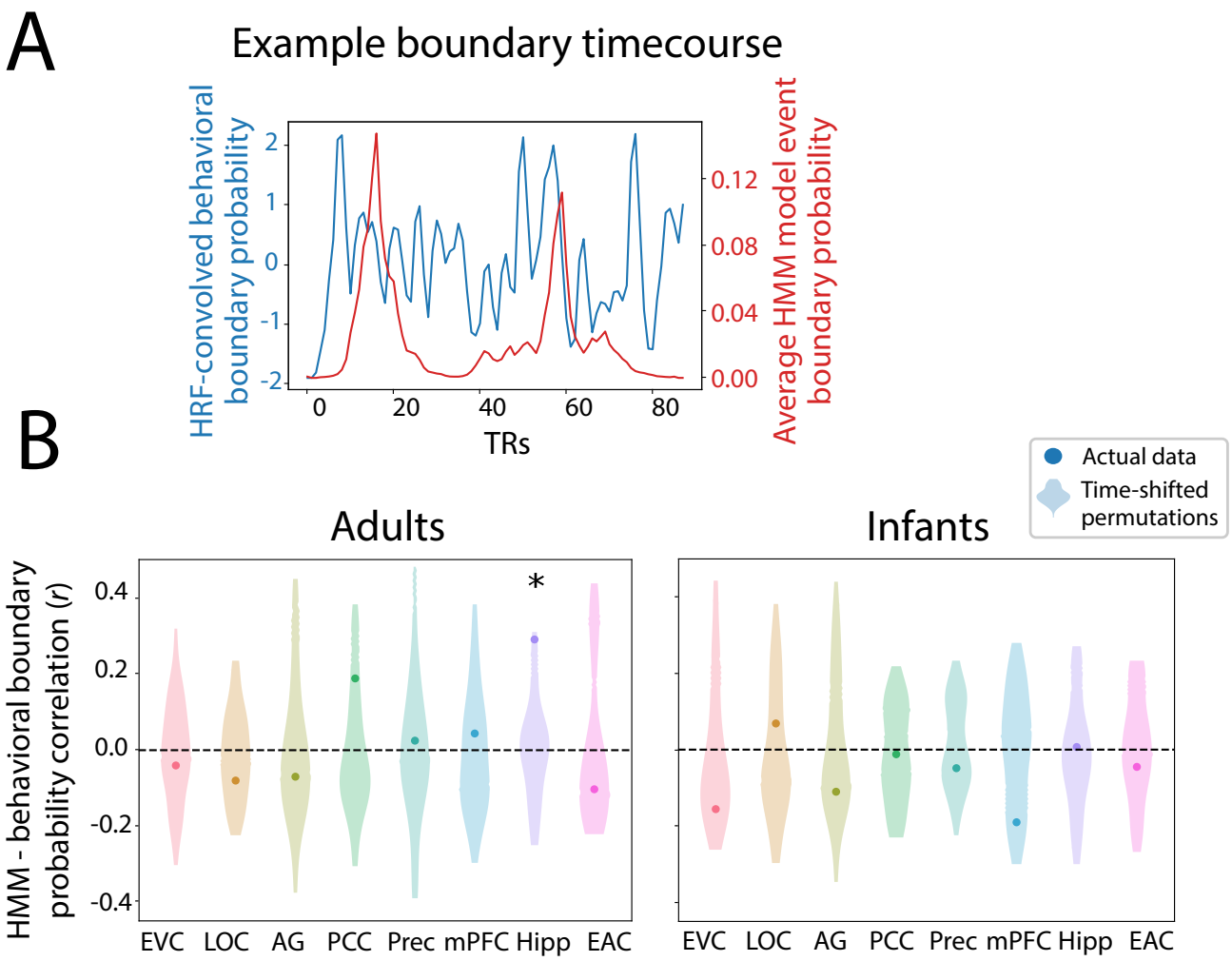

Fig. S7. Alignment of behavioral and neural boundaries. (A) Schematic of HRF-convolved adult behavioral event boundary probability in blue and adult or infant neural event boundary probability in red. (B) ROI analysis of neural and behavioral boundary alignment for actual vs. permuted timecourses. The single dot represents the actual correlation between neural and behavioral event boundary probabilities. The violin plot represents the permuted null distribution from time-shifts of the neural event boundary timecourse. * $p < 0.05$.

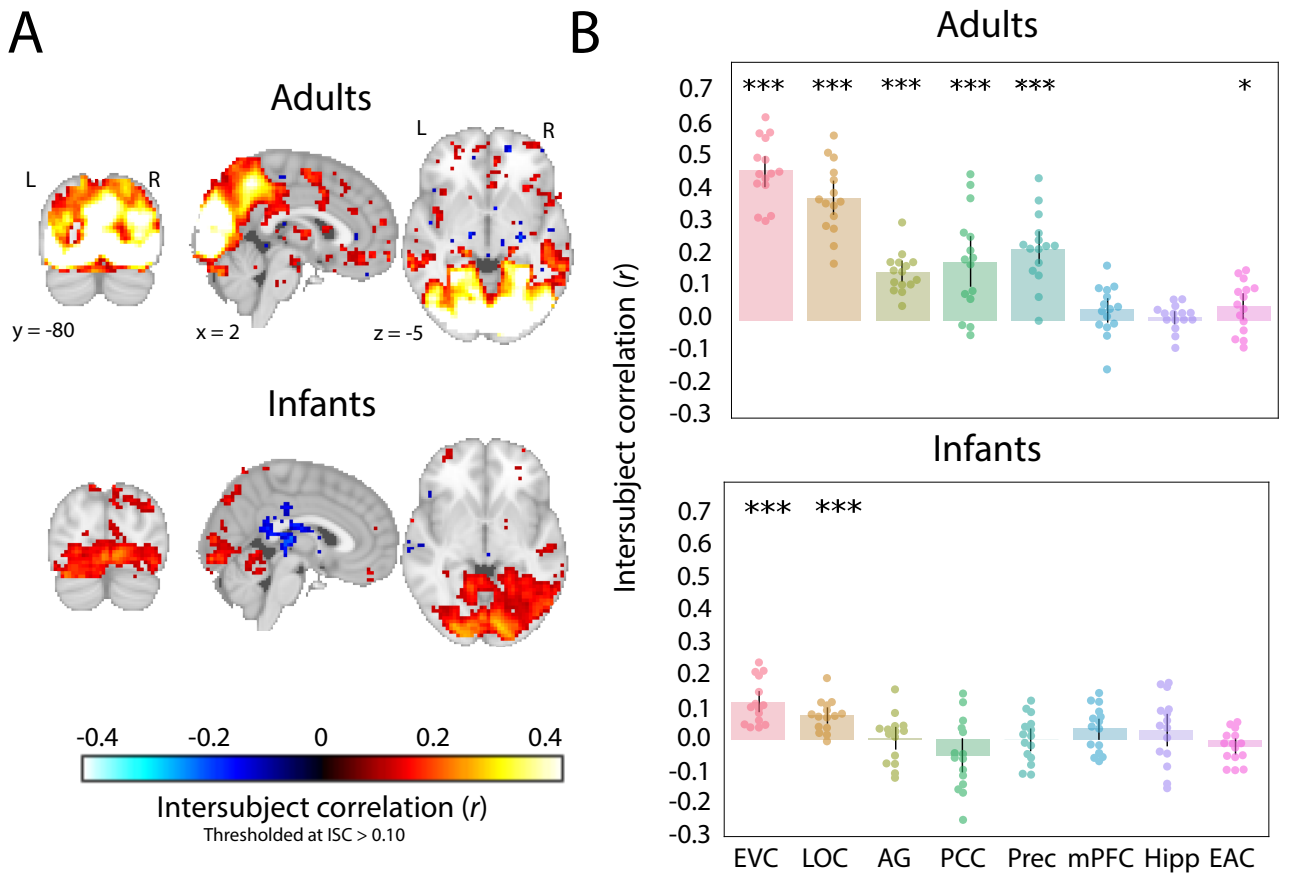

Fig. S8. Average leave-one-out intersubject correlation (ISC) in adults and infants watching the Mickey movie. (A) Whole-brain voxel-wise ISC values in the two groups, thresholded for visualization purposes at voxels with correlation values greater than 0.10. (B) ISC values in the ROIs, with the mean at the column height. Dots represent individual participants and error bars represent 95% CIs of the mean from bootstrap resampling. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

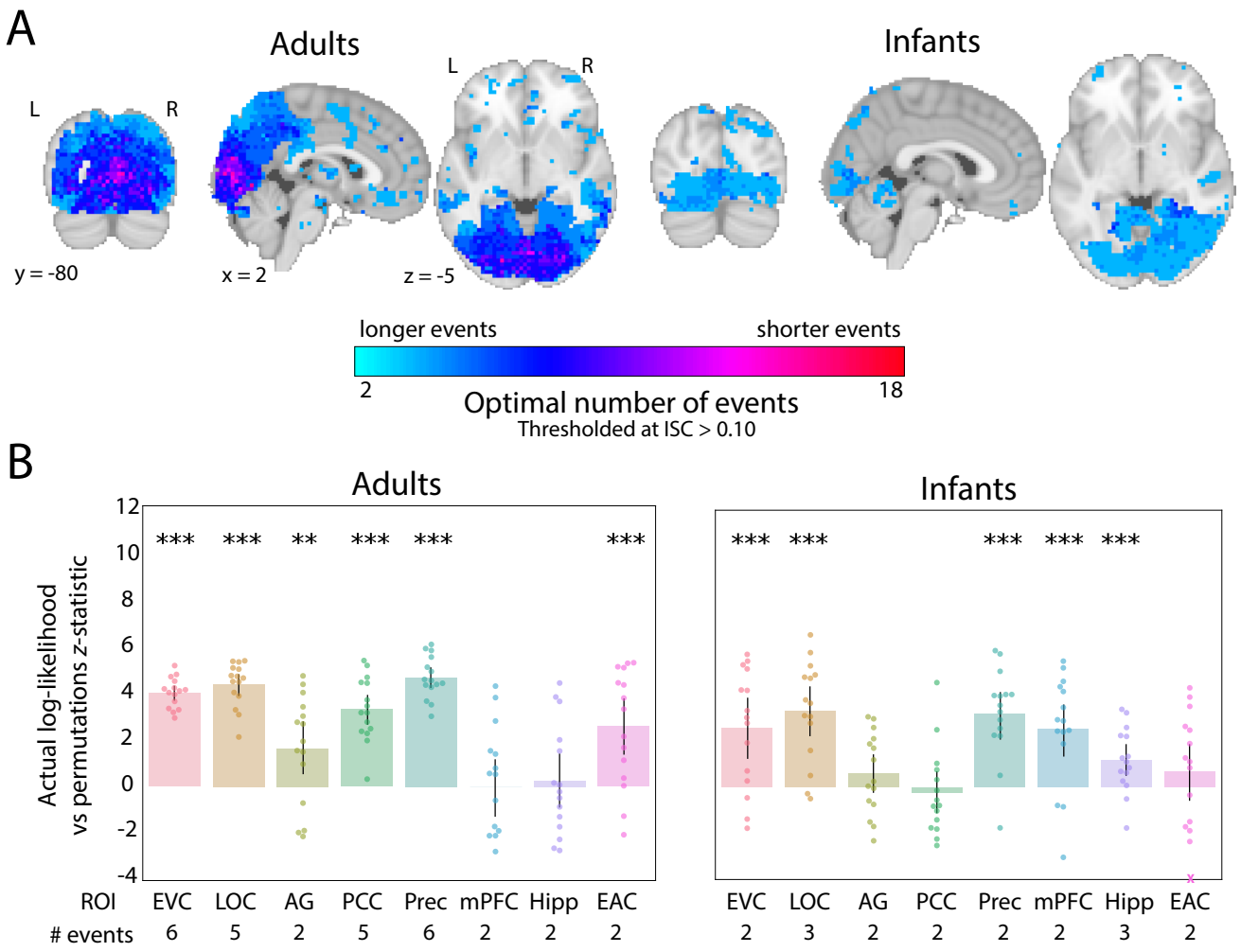

Fig. S9. Event structure and reliability for the Mickey movie. (A) The optimal number of events plotted across the brains of adults and infants. Voxels with an average ISC value greater than 0.10 are plotted for visualization purposes. (B) Results of the nested cross-validation procedure for computing the reliability of event segmentation in ROIs for adults and infants, calculated as the *z*-statistic comparing actual and permuted participant data. The number of events that optimized model log-likelihood in the full sample of participants is labeled below the x-axis. Dots represent individual participants and error bars represent 95% CIs of the mean from bootstrap resampling. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

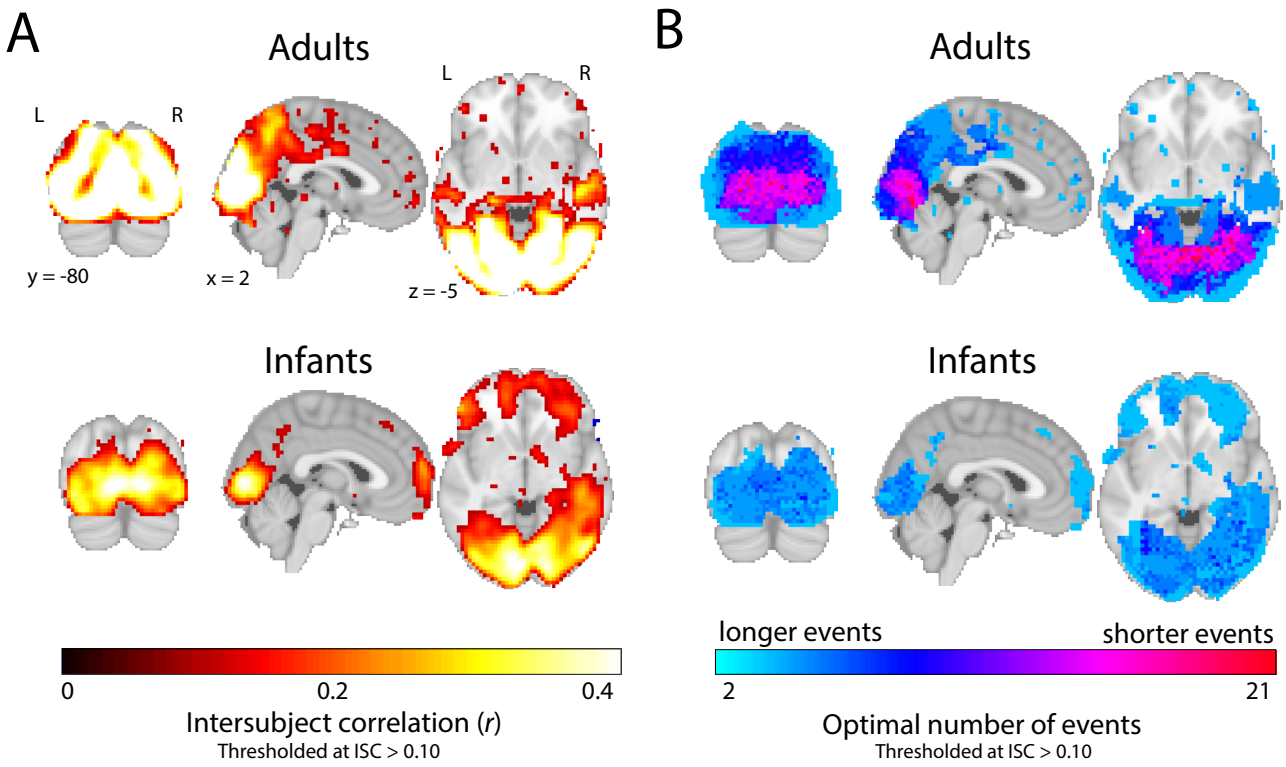

Fig. S10. Results from nonlinear anatomical alignment. (A) Whole-brain voxel-wise ISC values in the two groups. (B) The optimal number of events for a given voxel was determined via a searchlight analysis across the brain, which found the number of events that maximized the model log-likelihood in held-out data. In both panels, voxels with an average ISC value greater than 0.10 are plotted for visualization purposes.

ID	Age	Sex	Prop. incl. motion	Prop. incl. gaze	Gaze IRR	Num. coders	Min. usable
s6057_1_1	3.60	M	1.00	1.00	1.00	2	3.00
s3097_1_1	3.90	F	0.99	0.86	0.97	2	2.53
s2067_1_1	4.20	M	0.94	0.96	0.91	2	2.70
s9057_1_1	4.70	F	0.72	0.73	0.91	2	1.90
s8077_1_1	4.70	M	0.77	0.96	0.99	3	2.23
s2047_1_1	4.70	M	0.93	0.89	0.96	2	2.63
s7017_1_1	5.20	F	1.00	1.00	0.99	2	3.00
s2097_1_1	5.20	M	0.87	0.94	0.97	2	2.50
s7057_1_1	5.30	M	0.92	0.94	0.95	2	2.67
s4047_1_1	5.50	F	1.00	1.00	0.94	2	3.00
s2037_1_1	5.60	M	0.94	0.80	0.78	2	2.33
s6607_1_2	7.40	M	1.00	n/a	n/a	0	3.00
s2057_1_2	7.70	F	0.97	0.81	0.92	2	2.43
s8037_1_1	8.00	F	0.80	0.88	0.91	3	2.20
s5037_1_1	8.70	M	0.97	0.94	0.99	2	2.80
s0057_1_3	9.00	F	1.00	0.96	0.95	2	2.87
s7067_1_3	9.20	F	1.00	0.98	0.98	3	2.93
s4607_1_3	9.70	F	0.66	0.64	0.86	2	1.50
s2687_1_3	9.90	M	0.80	0.56	0.89	2	1.37
s3607_1_2	10.40	F	1.00	0.70	0.86	3	2.10
s1607_1_2	10.70	M	0.96	0.84	0.93	2	2.50
s8687_1_3	11.00	F	0.99	1.00	0.97	3	2.97
s6687_1_3	11.30	F	1.00	0.97	0.90	3	2.90
s0607_1_2	12.70	M	1.00	0.98	0.92	2	2.93
Avg.	7.43	n/a	0.93	0.88	0.93	2.17	2.54

Table S1. Demographic information for infants. 'ID' is a unique infant identifier (i.e., sXXXX_Y_Z), with the first four digits (XXXX) indicating the family, the fifth digit (Y) the child number within family, and the sixth digit (Z) the session number with that child. 'Age' is recorded in months. 'Sex' is female or male. 'Prop. incl. motion' is the proportion of TRs included after motion exclusion. 'Prop. incl. gaze' is the proportion of TRs included after gaze exclusion. 'Gaze IRR' is the proportion of frames coded the same way across gaze coders. 'Num. coders' is the number of coders per session. For one participant (s6607_1_2) gaze data was not recorded, and so there are values of n/a for the gaze reliability and number of coders. 'Min. usable' is the minutes of movie data that were usable and included in our analyses.

ID	Age	Sex	Prop. incl. motion	Prop. incl. gaze	Gaze IRR	Num. coders	Min. usable
mov_01	20.00	F	1.00	0.99	n/a	1	2.97
mov_02	20.00	F	1.00	1.00	n/a	1	3.00
mov_03	19.00	F	1.00	0.94	n/a	1	2.83
mov_04	21.00	F	1.00	1.00	n/a	1	3.00
mov_05	21.00	M	1.00	1.00	n/a	1	3.00
mov_06	18.00	F	1.00	0.99	n/a	1	2.97
mov_07	23.00	M	1.00	1.00	n/a	1	3.00
mov_08	21.00	M	1.00	1.00	n/a	1	3.00
mov_09	22.00	M	1.00	0.99	n/a	1	2.97
mov_10	23.00	F	1.00	1.00	n/a	1	3.00
mov_11	32.00	M	1.00	0.98	n/a	1	2.93
mov_12	20.00	M	1.00	1.00	n/a	1	3.00
mov_13	18.00	M	1.00	1.00	n/a	1	3.00
mov_14	20.00	F	1.00	1.00	n/a	1	3.00
mov_15	25.00	F	1.00	1.00	n/a	1	3.00
mov_16	19.00	F	1.00	1.00	n/a	1	3.00
mov_17	22.00	F	1.00	1.00	n/a	1	3.00
mov_18	24.00	F	1.00	1.00	n/a	1	3.00
mov_19	22.00	F	1.00	1.00	n/a	1	3.00
mov_20	24.00	M	1.00	0.99	n/a	1	2.97
mov_22	22.00	M	1.00	1.00	n/a	1	3.00
mov_24	26.00	F	1.00	0.99	n/a	1	2.97
mov_25	27.00	M	1.00	1.00	n/a	1	3.00
mov_26	32.00	M	1.00	0.84	n/a	1	2.57
Avg.	22.54	n/a	n/a	0.99	n/a	n/a	2.97

Table S2. Demographic information for adults. 'ID' is a unique adult identifier. 'Age' is recorded in years. 'Sex' is female or male. 'Prop. incl. motion' is the proportion of TRs included after motion exclusion. 'Prop. incl. gaze' is the proportion of TRs included after gaze exclusion. 'Gaze IRR' is the proportion of frames coded the same way across gaze coders, and is given a value of n/a when only one person coded the data. 'Num. coders' is the number of coders per session. 'Min. usable' is the minutes of movie data that were usable and included in our analyses.

References

1. Ellis CT, Baldassano C, Schapiro AC, Cai MB, Cohen JD (2020) Facilitating open-science with realistic fMRI simulation: validation and application. *PeerJ* 8.
2. Kumar M, et al. (2020) BrainIAK tutorials: User-friendly learning materials for advanced fMRI analysis. *PLOS Computational Biology* 16(1):e1007549.
3. Poppe T, Willers Moore J, Arichi T (2021) Individual focused studies of functional brain development in early human infancy. *Current Opinion in Behavioral Sciences* 40:137–143.
4. Lositsky O, et al. (2016) Neural pattern change during encoding of a narrative predicts retrospective duration estimates. *eLife* 5:e16070.
5. DuBrow S, Rouhani N, Niv Y, Norman KA (2017) Does mental context drift or shift? *Current Opinion in Behavioral Sciences* 17:141–146.
6. Lee CS, Aly M, Baldassano C (2021) Anticipation of temporally structured events in the brain. *eLife* 10:e64972.
7. Avants BB, et al. (2011) A reproducible evaluation of ants similarity metric performance in brain image registration. *NeuroImage* 54(3):2033–2044.