THE UNIVERSITY OF CHICAGO

Narrative Engagement and Polarized Neural Responses to Political Videos

By

Zhimei Niu

March 2022

A paper submitted in partial fulfillment of the requirements for the
Master of Arts degree in the
Master of Arts Program in the Social Sciences

Faculty Advisor: Yuan Chang Leong
Preceptor: Danielle Bolling

Abstract

Prior work has shown that neural activity diverges between conservatives and liberals while viewing political messages. Understanding what psychological factors contribute to these divergent responses informs our understanding of how people make political decisions. In this work, we examined the relationship between narrative engagement (i.e. the extent to which someone is both attending and emotionally involved while listening to or watching a narrative) and divergent neural responses to political videos. We measured narrative engagement by applying a predictive model that has been previously shown to predict moment-by-moment fluctuations in narrative engagement from connectivity patterns in the brain. We then examined if and how the neural measure of narrative engagement is associated with divergent neural responses between conservatives and liberals while watching political videos. The result shows a negative relationship between narrative engagement and political polarization.

Neural divergence in political narratives

People's interpretations of the same events can diverge due to different pre-existing beliefs (Yeshurun et al., 2017). Prior studies have shown that we can measure the biased assimilation of political information using functional magnetic resonance imaging (fMRI) (Leong et al., 2020; Van Baar et al.,2021). Subjects' brain activity synchronizes while watching or listening to the same story, and those with more similar interpretations of the story have higher synchrony in their brain responses (Finn et al., 2018; Nguyen et al., 2019; Regev, et al., 2019; Leong et al., 2020). The measurement of neural activity reveals important features for understanding human cognitive processes associated with subjective interpretations. Watching real-life narratives involves activation in early sensory regions like visual and auditory cortical areas; and specifically for narrative interpretation, higher-order brain areas including the dorsomedial prefrontal cortex (DMPFC), posterior medial cortex (PMC), middle temporal gyrus (MTG) are involved (Leong et al., 2020; Honey et al., 2012; Yeshurun et al., 2017; Nguyen et al., 2019; Regev, et al., 2019).

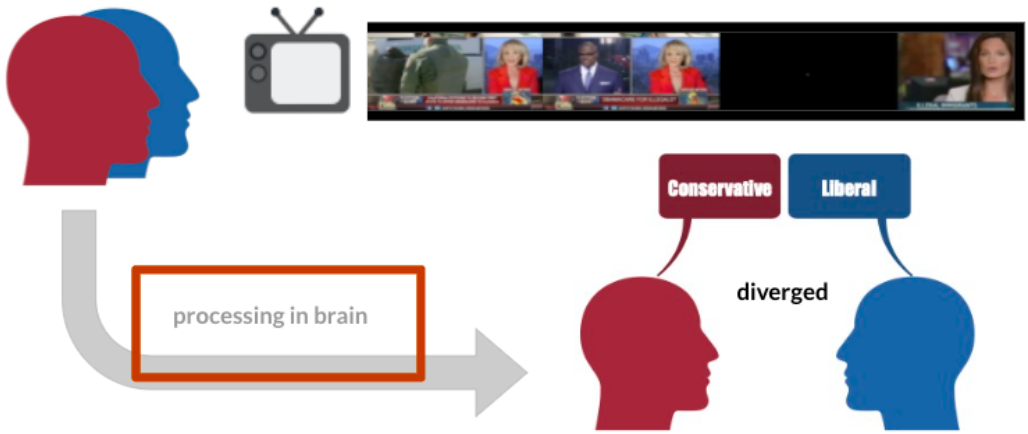
Political Polarization

In the political domain, partisans with opposing political views respond differently to the same information (Lord, Ross & Lepper, 1979, Fig. 1). For example, when opponents and proponents read studies that contain both disconfirming and supporting evidence for the deterrent efficacy of the death penalty, they are biased by their initial attitudes, and the gap between their views increases (Lord, Ross & Lepper, 1976). A recent study has shown that this is related to divergent processing of the same information in the brain, a phenomenon that has been referred to as "neural polarization" (Leong et al., 2020). In particular, Leong and colleagues had

conservatives and liberals watch real-world political videos (e.g., vice-presidential debates, news footage, campaign advertisements), and showed that activity in the DMPFC diverged between the two groups while watching the videos. This divergence predicted subsequent attitude polarization. In a related study, Van Baar and colleagues (2021) showed that neural polarization is related to individual differences in uncertainty tolerance. Specifically, individuals who are less tolerant of uncertainty exhibit greater neural polarization. Together, both studies demonstrate that divergent processing of information between conservatives and liberals can be measured using fMRI.

Figure 1

Schematic diagram of biased assimilation of political information



Note. Prior studies have shown that liberals or conservatives process the same information differently, which further increases political polarization.

Narrative Engagement

On the other hand, how engaged people are toward real-life information might also influence their interpretations and beliefs. In other words, while people are attending to the visual and auditory information from narratives, the emotional-laden attention, here defined as narrative engagement, might bias their higher processing in specific brain regions (Regev et al., 2019; Song et al., 2021). Recent work by Song et al. (2021) demonstrated that narrative engagement can be predicted from neural activity. In their study, they collected fMRI data and behavioral ratings of narrative engagement while participants watched an episode of the BBC television series, “Sherlock”. They then trained a multivariate regression model (specifically, a support vector regression model) to predict behavioral ratings of narrative engagement from fMRI data. In a first analysis, they found that engagement was associated with increased across-subject synchronization of activity in the default mode network (DMN), including the inferior parietal cortex (IPC), posterior cingulate cortex(PCC) and precuneus, medial prefrontal cortex (mPFC), and anterior cingulate cortex(ACC). These same areas have been previously associated with processing narratives (Chen, et al., 2017; Baldassano et al., 2017; Yeshurun et al., 2021). Furthermore, Song and colleagues were able to accurately predict the behavioral ratings from functional connectivity data. More impressively, their predictive model generalized to a second dataset, such that a model trained on one dataset could be used to predict behavioral engagement in a second dataset. Their work highlights the possibility of predicting narrative engagement in any fMRI dataset.

Examining the relationship between divergent processing of political information and narrative engagement

The central question of this thesis is to examine the relationship between the divergent processing of political information and narrative engagement. One hypothesis is that narrative engagement would positively predict the divergent processing of political information. For example, when partisans are engaged, they might be more motivated to engage in motivated political reasoning (cite the Granot paper we read in class). An alternative hypothesis is that narrative engagement would negatively predict the divergent processing of political information. For example, when partisans are more engaged by the narrative, they rely more on the information in the narrative and less on their prior beliefs. Thus, they are less likely to engage in motivated political reasoning. In this work, we test these two hypotheses by implementing a prediction with a computational model. Specifically, we measure narrative engagement by applying the model from Song and colleagues (2021) to a dataset of conservatives and liberals watching political videos. We then tested if narrative engagement would predict the divergent processing of these videos, with divergent processing measured by the level of neural polarization (Leong et al., 2020).

Materials and Methods

Datasets

We used two publicly available datasets. The first is the *Polarization* dataset, collected by Leong and his colleagues (2020). The dataset consists of fMRI data from 38 adults with a mean age of 31.3 (23 male, 15 female). Participants were divided into conservatives and liberals based on a median split of their responses on a seven-point scale prior to the fMRI experiment. While

undergoing fMRI, participants watched 24 videos in random order on six immigration policies (total duration: 35 min 26s). The six policies are 1) the construction of a wall along with the United States–Mexico border to reduce illegal immigration; 2) allowing illegal/undocumented immigrants to work legally in the United States without fear of deportation; 3) banning refugees from majority-Muslim countries from entering the United States; 4) allowing the use of federal funds to pay for emergency healthcare for undocumented/illegal immigrants; 5) providing a pathway to citizenship for undocumented individuals brought into the United States illegally as children; 6) cutting federal funding to sanctuary cities unless the cities agree to fully cooperate with the US immigration and customs enforcement. Each run contains one video on each policy (Leong et al., 2020). The behavioral response of political attitudes was measured again using the same questionnaire as the pre-experiment after the fMRI experiment.

The second dataset is the *Sherlock* dataset collected by Chen and colleagues (2019). In this dataset, 17 healthy subjects recruited from Princeton Community (12 male, 10 female, ages 18–26, mean age = 20.8) watched the first 50 minutes of Episode 1 of the BBC television show Sherlock. All participants are right-handed native English speakers with normal or corrected-to-normal vision and had not watched *Sherlock* prior to the experiment. The movie was split into two parts of approximately equal length (946 and 1030 TRs).

fMRI Data Acquisition and preprocessing

For both datasets, we preprocessed the fMRI data using FSL/FEAT v.5.98 (FMRIB software library, FMRIB, Oxford, UK). Motion correction, slice-timing correction, removal of low-frequency drifts using a temporal high-pass filter (100-ms cutoff), and spatial smoothing (4-mm full width at half maximum) were performed as part of preprocessing. Next, we registered

participants' functional data to their high-resolution anatomical image (rigid-body transformation with 6 degrees of freedom) and then to a template brain in Montreal Neurological Institute space (affine transformation with 12 degrees of freedom). The preprocessed data were loaded into MATLAB (Mathworks) with the NIfTI toolbox for functional analysis. In the experimental measures the video order was randomized; thus to minimize the effect of the different video combinations in each run for each participant, we z-scored the time course separately within each video. The preprocessed data were parcellated into 122 region-of-interest (ROIs) with 114 cortical ROIs from the atlas reported in Yeo et al (2015) and 8 subcortical ROIs as part of the Brainnetome atlas. The subcortical ROIs include the bilateral amygdala, hippocampus, thalamus, and striatum. For each ROI, the blood oxygen level-dependent (BOLD) time course of the voxels in the ROI was averaged to a single representative time course.

Neural polarization analysis

To calculate neural polarization from the *Politics* dataset, we extracted the BOLD time-course for each participant, and z-scored it separately for each video. We then calculated the within-group intersubject correlation (ISC) as the voxel-wise Pearson correlation between each participant and the average of all other participants in the same political group. We calculated the between-group ISC as the Pearson correlation between each participant and the average of all participants in the other political group. We then computed the difference between within-group ISC and between-group ISC. The procedure was repeated for all participants and all voxels, and then averaged across all participants.

This procedure allowed us to obtain the dorsomedial prefrontal cortex (DMPFC) ROI identified in Leong et al. (2020). For each voxel, we ran a t-test to assess if the average

difference between within-group ISC and between-group ISC was greater than zero. To generate a null distribution, we flipped the sign of the difference in *r* for a random subset of participants and recomputed the *t*-statistic. This procedure was repeated 10,000 times. The *p*-value was computed as the proportion of the null distribution that was more positive than the observed *t*-statistic. We corrected for multiple comparisons using family-wise error cluster-correction threshold of $p < 0.05$ using Gaussian Random Field theory with a cluster-forming threshold of $p < 0.001$. DMPFC was defined as the voxels that survived correction for multiple comparison in the vicinity near the reported coordinates of the DMPFC in the Leong et al. paper. We extracted the average DMPFC time course separately for conservative and liberal participants. Next, we computed the absolute difference between average conservative and average liberal time course, segmented it into 86 segments average of 24.7 seconds (average duration of 24.7 seconds, range from 12 to 38s) based on the event segmentations identified in Leong et al. (2020), and averaged the activity within each segment. This 86 segment time-course was used as our measure of neural polarization.

Neural engagement analysis

To perform the neural engagement analysis on the *Sherlock* dataset, we followed the procedure by Song et al. (2021). We re-computed time-resolved functional connectivity (FC) matrices by calculating the Fisher's r-to-z-transformed Pearson's correlations between the BOLD signal time courses of every pair of regions of interests (122 × 122 ROIs in total) using a tapered sliding window with a size of 30 repetition time (TR). The *Sherlock* dataset is a 50 minutes continuous video; however, the *Politics* dataset consists of 24 separate videos with a mean duration of 1 to 2 minutes (35 min 26 s in total). Due to the discontinuity and short video

durations of the *Politics* dataset (Leong et al., 2020), using a 30TR window size is not feasible. We thus applied a novel time-resolved FC-computation method ("Edge Time Course", ETC) developed by (Esfahlani et al., 2020) to substitute the Pearson's correlations implemented in Song et al. (2021).

The ETC method involves unwrapping the traditional Pearson correlation and directly computes a covariation measure that results in a set of time series for each time point matched with each pair of brain regions. Specifically, the networks from fMRI data were constructed by estimating the statistical dependency, how strongly ROIs are functionally connected, between every pair of time series. Each time series of each participant in each ROI were z-scored. The magnitude of moment-to-moment co-fluctuations, which is a vector by multiplying each matching paired element in the time series of two ROIs, is then calculated (i.e. multiplying the element *i* and *j* at time *t*th position). The product of the multiplication n is positive when both *i* and *j* simultaneously increase/decrease their activity relative to baseline. In contrast, the product of the multiplication is negative if *i* and *j* have opposite signs relative to the baseline. The element-wise product was normalized according to the square-root standard deviation of both time series. This procedure is repeated for all pairs of ROIs for each participant, which results in a node-by-node-by-time correlation matrix (122 x 122 x 1976 dimensions).

We validated the ETC method by training a nonlinear support vector regression (SVR) model with the *Sherlock* dataset and assessing whether we would predict behavioral ratings of engagement (Song et al., 2021). The time-resolved FC matrices are computed using edge time courses (Esfahlani et al., 2020) with a 15 sliding window. In order to limit the influence of outliers, we imposed a threshold of ±3 on the time-resolved FC matrices prior to SVR model implementation. We also tested other thresholds (±5, ±7) which yielded similar results. By

z-normalizing the time series of every functional feature across time, the temporal variance within individuals is retained while across-individual variance is excluded. The SVR model was implemented with python (sklearn.svm.SVR; rbf kernel, maximum iteration set to 1,000). We then used the model to predict t behavioral rating data (group-average behavioral measurement).. Feature selection for the model is performed in every round of cross-validation by selecting functional connections (i.e., edges) that are significantly ($p < .01$) correlated with behavioral measurement in training participants. The model was trained in all but one participant and tested on the held-out participant for each round of cross-validation. The model prediction performance is tested for each leave-one-subject-out cross-validation, and an overall prediction performance was computed as the mean r across results in all cross-validation folds was measured.

Examining the relationship between narrative engagement and neural polarization

We applied the model of engagement trained on the *Sherlock* dataset using the ETC method to the *Politics* dataset to obtain a measure of narrative engagement while participants were watching political videos. We applied the same processing steps on the *Politics* dataset, including computing and thresholding the time-resolved FC matrices (threshold of ±3). In addition, we averaged the ETC for each of 84 video segments before inputting them into the model. The proportions of pairwise regions, grouped by the predefined functional network, were selected from those significantly correlated ($p < .01$) with narrative engagement in every cross-validation fold in the within-dataset SVR prediction of the Sherlock dataset. For each cross-validation fold, we obtain a 1 x 84 vector representing the narrative engagement predicted from that held-out participant's ETC data. We then correlate this measure with neural

polarization in the DMPFC. The prediction performance is then tested against chance using a two-tailed paired t-test.

Results

Measuring narrative engagement in the *Sherlock D*ataset

As the ETC method of computing FC is novel, it is important to validate that our model is able to accurately predict neural engagement from ETC data. Specifically, we validated the edge time course using the *Sherlock* dataset to see whether it can produce a similar prediction performance in the SVR model generated in Song et al., (2021). If performance when training on ETC is comparable to those obtained in Song et al., (2021), we will then apply it to the *Politics* dataset.

In a within-dataset prediction, the SVR model was trained using FC data generated with ETC (window size = 1) from all but one participant and applied to the held-out participant's BOLD activity to predict the group-average engagement observed at every corresponding TR. The average correlation across participants between predicted engagement and behavioral engagement was statistically significant, but far weaker when the model was trained using the sliding window 30 TR approach implemented in Song et al. (2021) results ($r = 0.0615$, $p < 0.01$ vs. $r = 0.5551$, $p < 0.001$). To improve model performance, we retrained and tested the ETC model at different window sizes (3, 5, 10, 15). Model performance increased with increasing window size (window size 3TR: $r = 0.1134$, $p < 0.001$; window size 5TR: $r = 0.1475$, $p < 0.001$; window size 10TR: $r = 0.2110$, $p < 0.001$; window size 15TR: $r = 0.2964$, $p < 0.001$) (Fig. 2). We proceeded with computing ETC with a window size of 15 TR as it provided us with the highest model accuracy among the window size tested.

Edge Time Course: Thresholding FC time course

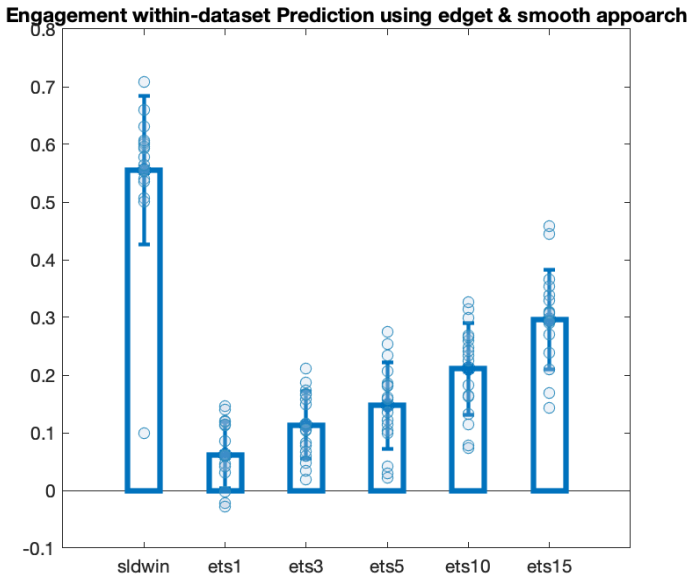
When we visualized the ETC, we noticed the presence of high-frequency spikes in the data, with the highest peak reaching ±15 in the amplitude of the FC time course (Fig. 3A) These spikes were not present when FC was computed using the dynamic sliding window approach (Fig. 3B). We note that these spikes likely reflect noise in the signal due to lower noise-to-ratio compared to the dynamic sliding window approach, and a potential imperfect alignment between videos and brain activity in every 1.5 s (1 TR).

Thus, to further improve model performance, we performed a thresholding step prior to entering the ETC FC time courses into the SVR model. Specifically, we thresholded the amplitude of ETC time courses at +3 and -3. With the increment of window size, the amplitude peak gradually reaches a smaller range, which further indicates the potential noise or the mismatching between video and brain activity with small window size (Fig. 4). Hence, to increase model performance to a level comparable to the dynamic sliding window approach, we applied a ±3 to limit the presence of high-frequency spikes (Fig. 3C).

As the threshold at +3 and -3 for the amplitude of ETC time courses best limits the presence of high-frequency spikes (Fig. 3C), the model performance increased substantially compared to non-thresholded data (Fig. 4). Model performance with a 15 TR window size with thresholding ±3 was comparable to model performance using the dynamic sliding window approach (dynamic sliding window (30TR): $r = 0.5551$, $p < 0.001$; window size 1TR, threshold ±3: $r = 0.1575$, $p < 0.01$; window size 3TR, threshold ±3: $r = 0.2213$, $p < 0.01$;window size 5TR,threshold ±3: $r = 0.2702$, $p < 0.001$; window size 10TR, threshold ±3: $r = 0.3470$, $p < 0.001$; window size 15TR, threshold ±3: $r = 0.3920$, $p < 0.001$). Hence we proceeded with computing ETC with a window size of 15 TR and a ±3 threshold.

Figure 2

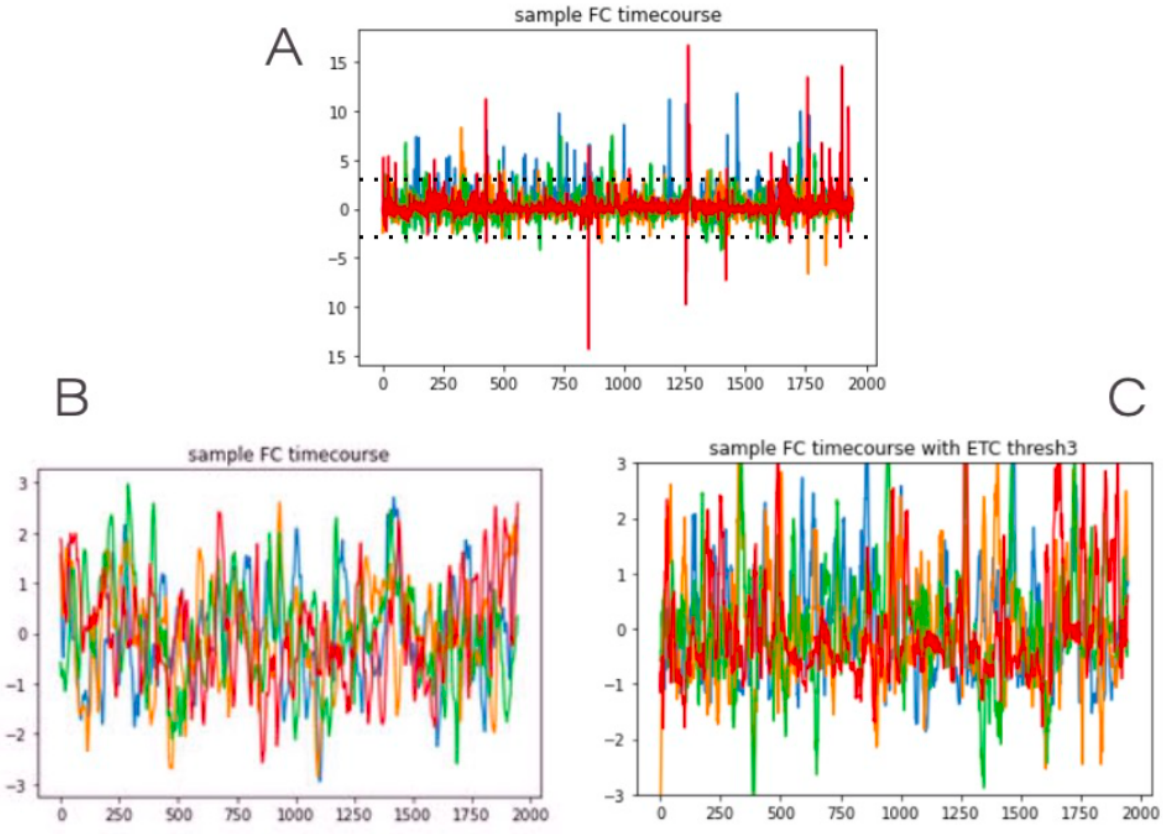
A comparison of model performance (average r) of dynamic sliding window approach and edge time course in a within-dataset prediction analysis of the Sherlock dataset



Note. Blue bars (from the left) represent the model performance of the dynamic sliding window approach with a window size of 30TR, followed by model performance of the ETC approach with window sizes 1TR, 3TR, 5TR, 10TR, and 15TR.

Figure 3

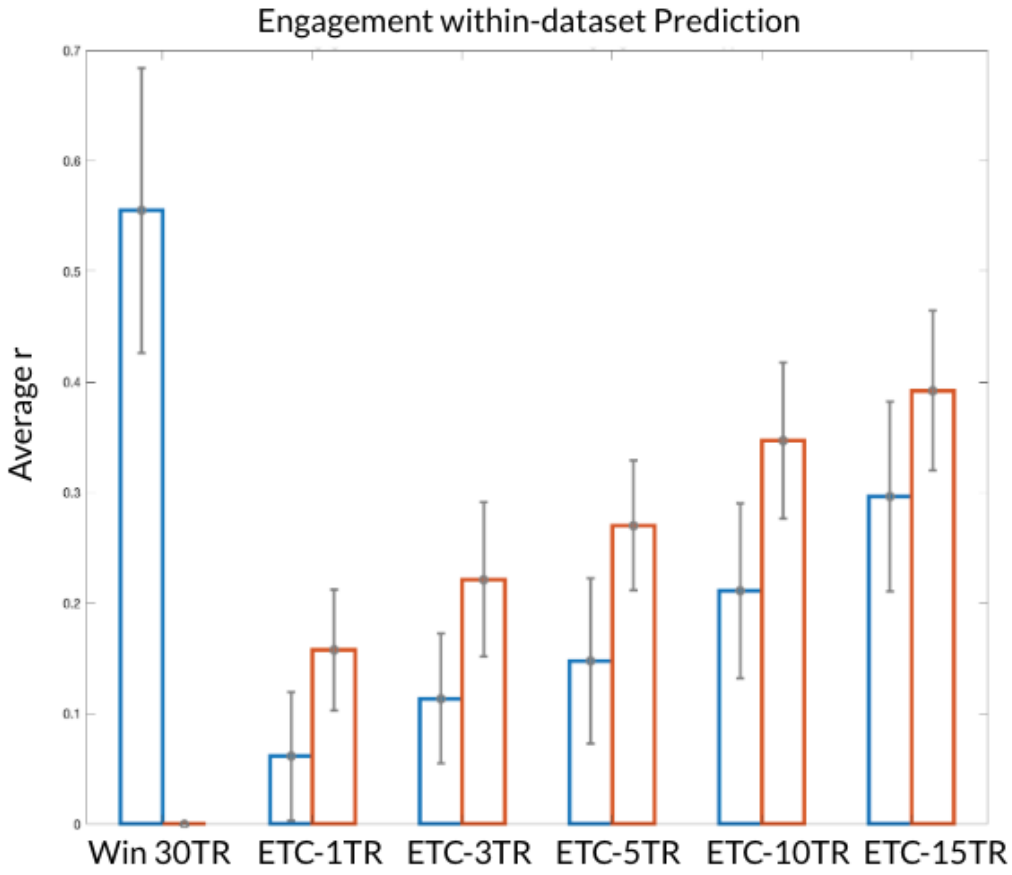
Sample functional connectivity time course computing different methods



Note. The x-axis represents the time (1TR = 1.5s), and the y-axis represents the amplitude of the functional connectivity. **(A)** Sample FC time courses computed using window-by-window ETC (Window size = 1TR), with the highest peaks and lowest troughs reaching ±15. Dotted lines indicate ±3 **(B)** Sample FC time courses computed using a dynamic sliding window (window size = 30TR) with values ranging within ±3. **(C)** Sample FC time courses computed using the ETC approach with window size 15TR, and thresholded at ±3. All timecourses presented above have been z-scored.

Figure 4

The correlation of within-dataset predicted behavioral engagement with observed behavioral engagement in the Sherlock dataset



Note. The first blue bar on the left is the result of using a dynamic sliding window with a 30TR size, other blue bars represent applying ETC in different window sizes (1TR, 3TR, 5TR, 10TR, 15TR) without thresholding. The red bars represent the correlation with applying ETC in different window sizes (1TR, 3TR, 5TR, 10TR, 15TR) with a threshold of ±3. A two-tail paired t-test is performed for all the correlations compared to a null distribution.

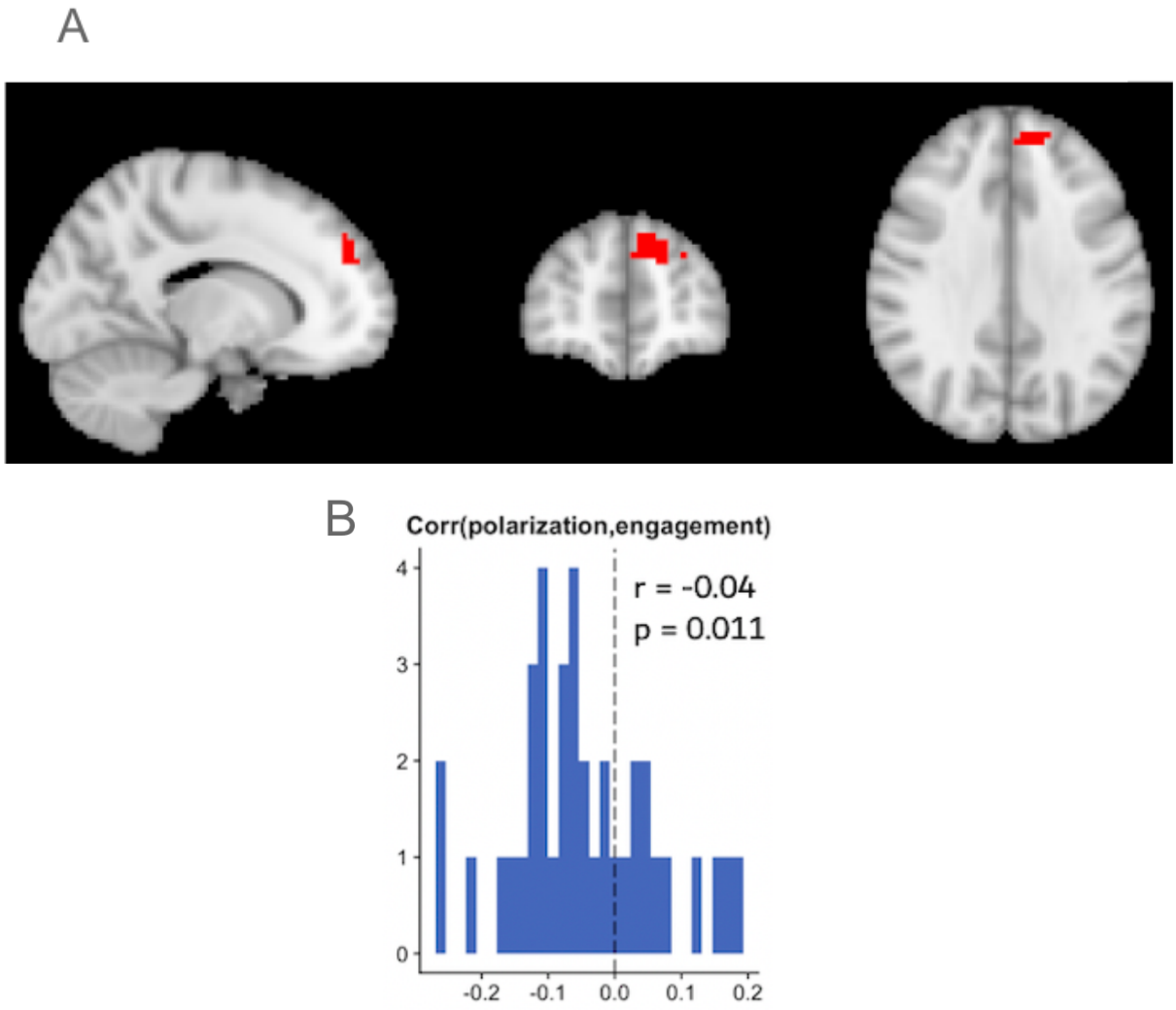
Neural polarization between conservatives and liberals watching political videos

Our initial analysis for the political dataset focused on reproducing the neural polarization results from the prior study (Leong et al., 2020). We calculated the "within-group ISC" as the correlation between each participant and the average of all other participants in the same political group (i.e., liberal vs. average liberal and conservative vs. average conservative). Also, we calculated a "between-group ISC" as the correlation between each participant and the average of all participants in the other political group (i.e., liberal vs. average conservative; conservative vs. average liberal). The result shows that "within-group ISC" was greater than "between-group ISC" only in the left DMPFC (Fig. 5A). Further, consistent with the prior result (Leong et al., 2020), the "within-group ISC" in the DMPFC was higher than the "between-group ISC" in both conservative and liberal participants. This indicates that the results in DMPFC were not driven by only one of the two groups. The difference between within-group and between-group ISC measures the similarity of neural activity between participants with similar political attitudes rather than between participants with dissimilar political attitudes. We then calculated the average difference between within-group ISC and between-group ISC, which is proven to be significant ($p<0.05$).

This procedure allowed us to obtain the dorsomedial prefrontal cortex (DMPFC) ROI. In order to calculate the absolute difference between conservatives and liberal groups, we extracted the average DMPFC time course separately for conservative and liberal participants. Next, we computed the absolute difference between average conservative and average liberal time courses and segmented it into 86 segments. This absolute difference between liberals and conservatives was used as our measure of neural polarization for subsequence analysis using the SVR.

Figure 5

(A). DMPFC time course diverges between conservatives and liberals. (B). Pearson's correlation between predicted political polarization and observed political polarization



Note. (A). Within-group ISC was higher than between-group ISC in the left DMPFC (B). Most participants' r-value were below 0.

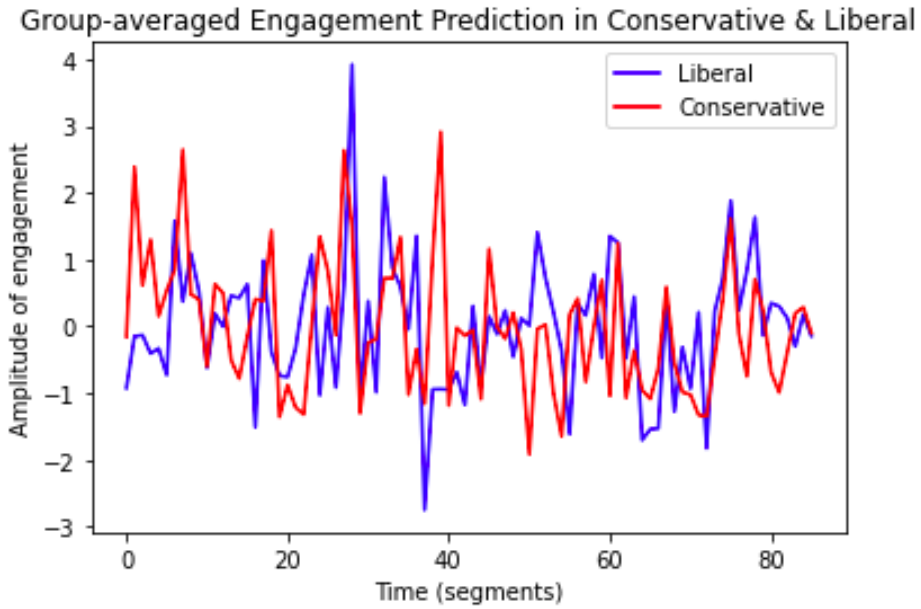
Narrative engagement predicts neural polarization

As suggested by prior evidence, the whole-brain functional connectivity (FC) predicts changes in engagement across different stories (Song et al., 2021). Here, we test whether narrative engagement is associated with neural polarization while viewing political messages. We computed neural polarization as the absolute difference between average conservative and liberal group-level time courses. We applied the SVR model trained on the *Sherlock* dataset to the *Politics* datasets to compute narrative engagement in the Politics dataset across the 86 segments. The average correlation between narrative engagement and political polarization was statistically significant ($r = -0.04$, $P = 0.011$). In particular, the distribution of correlation values is skewed to the left (Fig. 5B), indicating a negative relationship between narrative engagement and neural polarization in DMPFC. In other words, there is a stronger neural polarization when there is less narrative engagement.

We examined if our results would be different if we preprocessed the FC data in a different manner. In particular, we switched the order of the preprocessing procedure in the *Politics* FC data such that we first averaged within 86 segments first before computing the ETC and thresholding the resulting FC data at ±3. Model performance was qualitatively similar, but not statistically significant ($r = -0.038$, $P = 0.0582$).

Figure 6

The group-averaged predicted engagement time course for conservative (red) and liberal (blue) participants



Note. The line graph shows different patterns in liberal and conservative at some segments, while the correlation of predicted engagement with the political polarization is not statistically significant (liberal: r = -0.037, p = 0.1369; conservative: r = -0.051, p = 0.0708).

Examining the relationship between engagement and neural polarization separately for conservatives & liberals

Next we examined the relationship between engagement and neural polarization separately for conservative and liberal participants. Similar to the above analysis, in two separate across-dataset predictions, the SVR model was trained using FC data from both datasets, in which the *Sherlock* FC data was generated with a dynamic sliding window (30TR) and the *Politics* FC data of either conservatives/liberals group was generated with ETC, applied with a threshold of ±3 and averaged within 86 segments. The model was then used to predict the

engagement neural polarization observed in DMPFC at every corresponding TR separately for conservatives and liberals. The model performance for predicting the relationship between engagement in conservatives and neural polarization was not statistically significant ($r = -0.051$, $p = 0.0708$). Similarly, the model performance for predicting the relationship between engagement in liberals and neural polarization was also statistically non-significant ($r = -0.037$, $p = 0.1369$). Though the prediction for neural polarization is not significant, the group averaged time course of predicted engagement could possibly indicate a potential difference between conservatives and liberals in narrative engagement at every segment (Fig. 6).

Exploratory analysis

Functional connectivity pattern predicting behavioral emotion

Prior fMRI studies have shown that movies and narratives induced robust emotional responses (Finn et al., 2018; Gruskin et al., 2020; Jääskeläinen et al., 2022). Narrative engagement has been defined as the emotional-laden attention while watching or listening to a story (Song et al., 2021). Engagement largely involves emotional change, attention could drift with emotion. Hence, it is possible that changes in emotional responses are related to neural polarization in conservatives and liberals while watching political videos. However, it remains unknown whether behavioral emotional responses could be predicted from the brain pattern with a computational approach under a narrative context. Further, whether the relationship between emotion and neural polarization could be predicted. Thus, our initial approach to answering these questions is to examine whether behavioral emotion change can be predicted from FC neural activity in the *Sherlock* dataset with the SVR model. The emotional behavioral valence data in the Sherlock dataset were separated by (positive/negative) and processed with dynamic window

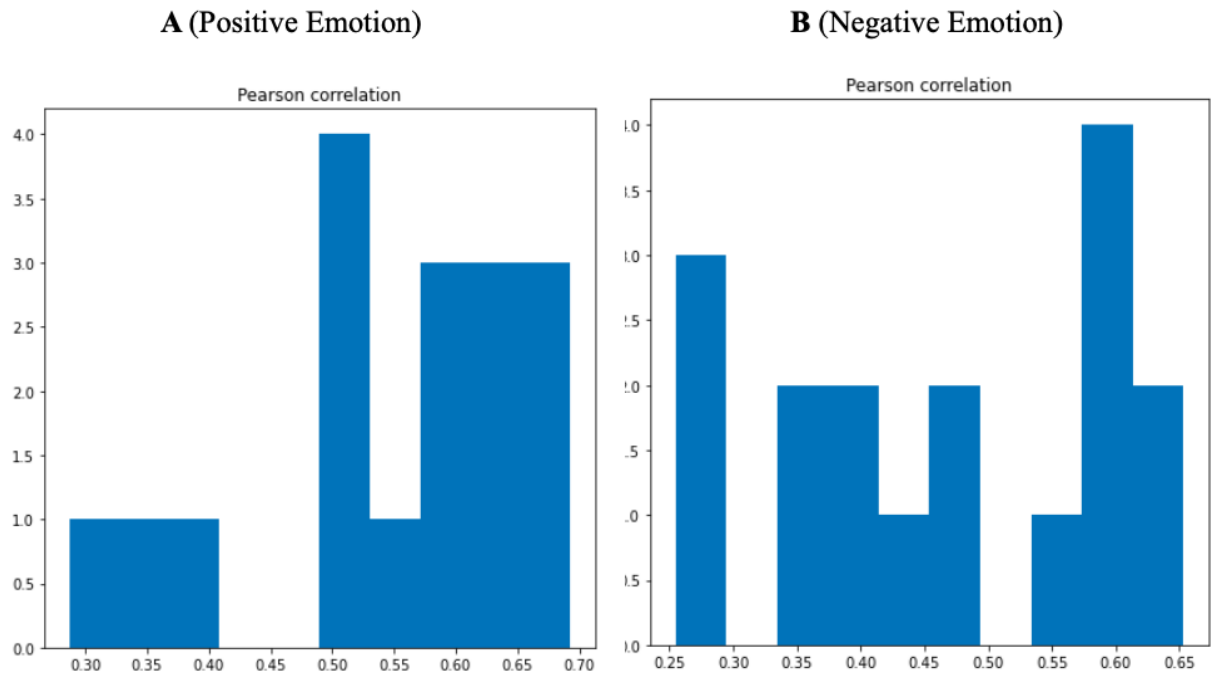
30TR. Similar to the prior analysis, in two separate within-dataset predictions, the SVR model was trained using FC data generated with 30TR dynamic sliding window circular shifting across the time course from all but one participant and applied to the held-out participant's BOLD activity to predict the group-average positive or negative emotion observed at every corresponding TR.

The average correlation across participants between predicted emotion valence and behavioral emotional response was statistically significant. Specifically, model performance successfully predict positive emotion from the FC activity ($r = 0.56$, $p < 0.01$, Figure 7A), and the negative emotion is also predicted ($r = 0.475$, $p < 0.01$, Figure 7B). This implies that the SVR model can be successfully applied as a computational tool to predict emotion valence. Both positive and negative emotion is highly predictable from brain activity. Thus, we infer that subsequent across-dataset analysis can ideally be done on neural polarization.

However, due to time limitations, we didn't manage to conduct the across-dataset analysis for predicting the relationship between emotion and neural polarization. Based on prior evidence, for example, individuals with depressive symptoms have a non-synchronization at the positive-emotion movie moment compared to healthy individuals, and a tendency of focusing on negative information in movies (Gruskin et al., 2020). It implicates that potential emotional bias driven by prior experience could potentially shift the attention or engagement towards a preferential stimulus. Thus, further analysis is necessary to be done in the future to examine the potential impact of emotional bias toward the biased interpretation of political information in the brain that drives further change in attitudes.

Figure 7

The Pearson correlation of predicted emotion with observed behavioral emotion in Sherlock Dataset



Note. (A). The predicted positive emotion significantly correlated with observed positive emotional valence ($r = 0.56$, $p < 0.01$). (B) The predicted negative emotion significantly correlated with observed negative emotional valence ($r = 0.475$, $p < 0.01$).

Discussion

Existing attitudes and the process of how individuals interpret political information could powerfully influence how they respond to political content and participate in political decisions. Recent evidence shows the occurrence of neural polarization in specific brain regions, mostly in DMPFC (Leong et al., 2020; Moore-Berg et al., 2020). The divergent patterns in conservatives and liberals may possibly be explained by the discounting of evidence that contradicts their political beliefs and reinforcing the evidence that confirms their beliefs while watching the same political content. In the current study, we first replicate the prior study (Leong et al., 2020) and generated the neural divergence that occurred in DMPFC between conservative and liberal participants watching the same political videos related to immigration policy. As the neural difference between conservatives and liberals only occurs in the DMPFC, DMPFC may play a significant role in leading to the divergence between groups. As shown in early research, increased activity in DMPFC is associated with the interpretation of narrative stimuli (Yeshurun et al., 2017; Finn et al., 2018; Nguyen et al., 2019), and is positively correlated with resistance to political belief change (Kaplan et al., 2016). These associations with DMPFC may imply that the difference in narrative interpretation and selectiveness of evidence to resist beliefs may explain the neural polarization and subsequent polarized political beliefs in conservatives and liberals.

The activity in DMPFC also has been implicated in multiple complex cognitive functions, including episodic memory retrieval, motivation in reward-seeking, inference of other's mental states, choice anxiety in decision making, emotional regulation of negative affect (Spreng et al., 2009; Overwalle, 2009; Shenhav & Buckner, 2014; Silver et al., 2015, Shigemune et al., 2017). These disparate finding further implicates a potential construction of an integrative situation model in DMPFC: the mental representation of the situations described in the narratives

combined with prior knowledge about the events, characters, actions could create a more detailed or biased representation of the information from the narratives (Yarkoni et al., 2008; Leong et al., 2020). The divergent pattern in DMPFC between conservative and liberal groups may thus reflect their different situation model towards the same content. Under the support of the situation model, here we focused on examining whether a difference exists in the attentional processes (e.g. how individuals selecting matched evidence under selective attention and emotional bias), might impact their higher-level cognition for interpreting the narrative content.

We defined "narrative engagement" as emotional-laden attention, though we acknowledge that "engagement" has been defined differently across different studies (Richardson, 2020). We examined narrative engagement as the potential factor that biases an individual's higher process in specific brain regions like the DMPFC while they are attending to information from narratives. We trained the SVR model to predict the relationship between narrative engagement and neural polarization in DMPFC. The computational model successfully predicted the relationship between narrative engagement and neural polarization, which is consistent with our expectation that machine learning combined with neuroimaging data can connect individuals' behavioral patterns with higher-level processing in the brain. Furthermore, the SVR model found a significant negative association between narrative engagement and neural polarization, which may suggest that when participants are more engaged with the political video content, they are processing it more similarly. This result is inconsistent with our second hypothesis that narrative engagement predicts more neural polarization. Several possibilities may explain this result: 1. the behavioral narrative engagement trained with the movie dataset may differ from the engagement in political news; 2. Existing individual differences in the extent of supporting political stands may affect the result; 3. while individuals

are more engaged with the video content, their brain response is more likely to be influenced by the narrative information than their prior existing knowledge. As indicated in a prior study (Leong et al., 2020), neural polarization also tracked subsequent attitude polarization. Combined with our result, it further implies that conservatives and liberals may be less likely to diverge in their interpretations and attitudes towards political information when they are more engaged with the narratives. As indicated, a potential explanation could be that while people are more engaged with the political content, they may be less likely biased by their existing political beliefs, and may interpret the political information from a more critical standpoint.

In the subsequent analysis, the SVR model failed to predict the relationship between narrative engagement and the neural polarization for conservatives and liberals separately. One potential reason is the lower statistical power due to a smaller number of subjects since we had separate conservatives and liberals participants for the analysis. Hence, increasing the subject pool by combining other political datasets may solve this problem and potentially improve our prediction results to a substantial degree.

As the engagement was defined as emotional-laden attention, the emotional bias could also potentially be a strong factor to bias individuals' interpretation of the narrative content (Finn et al., 2018; Gruskin et al., 2020; Jääskeläinen et al., 2022), which may further correlate with the neural polarization in DMPFC. In an extended analysis, to better understand how emotion plays a role in affecting the interpretation of the political content, we first examined whether the emotional valence can be successfully predicted from the brain activity in the *Sherlock* dataset. By implementing the SVR model in a within-dataset analysis, the results show we can successfully predict negative/positive emotional valence from brain activity. It demonstrates that the SVR model can be successfully applied to emotion, and a further prediction can be done to

associate the emotional response with neural polarization. Due to time limitations, we did not manage to conduct the analysis to predict the relationship between political polarization and emotional change. Future analysis can be done to examine whether emotional intensity and valence is related to the divergence in DMPFC.

Together, our findings demonstrate a computational approach for predicting the relationship between the input process (e.g. narrative engagement) and the divergence that exists in higher-level processing in the brain (e.g. neural polarization) to study the political brain under realistic conditions. With this approach, we identified a neural correlate of the biased processing of political information with narrative engagement, as well as demonstrated a potential explanation for the subsequent attitude change on political issues. Future work could be further done to obtain behavioral measurements of engagement while watching political news to understand the biased interpretation process in the brain and inform interventions to align the biased evidence selection process between conservatives and liberals.

Acknowledgment

I would like to thank my advisor Dr. Yuan Chang Leong and preceptor Dr. Danielle Bolling's support and supervision, as well as all the inspiration and support from members of the Motivation and Cognition Neuroscience lab at the University of Chicago.

References

Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. Journal of Personality and Social Psychology, 91(4), 612–625. https://doi.org/10.1037/0022-3514.91.4.612

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

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature neuroscience*, *20*(1), 115–125. https://doi.org/10.1038/nn.4450

Esfahlani, F. Z., Jo, Y., Faskowitz, J., Byrge, L., Kennedy, P. D., Sporns, O., Betzel,R.F. (2020). High-amplitude cofluctuations in cortical activity drive functional connectivity. *Proceedings of the National Academy of Sciences*, 117 (45) 28393-28401.

Finn, E.S., Corlett, P.R., Chen, G. et al.(2018). Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nat Commun* 9, 2043. https://doi.org/10.1038/s41467-018-04387-2

Gruskin, D. C., Rosenberg, M. D.,Holmes,A. J. (2020). Relationships between depressive

symptoms and brain responses during emotional movie viewing emerge in adolescence. *NeuroImage*. Vol. 216, 116217

Honey, C. J., Thompson, C. R., Lerner, Y., & Hasson, U. (2012). Not lost in translation: neural responses shared across languages. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *32*(44), 15277–15283. https://doi.org/10.1523/JNEUROSCI.1800-12.2012

Jääskeläinen, I. P., Ahveninen, J., Klucharev,V., Shestakova, A. N & Levy, J. (2022).Behavioral Experience-Sampling Methods in Neuroimaging Studies With Movie and Narrative Stimuli. *Frontier in Human Neuroscience*. https://doi.org/10.3389/fnhum.2022.813684

Kaplan, J., Gimbel, S. & Harris, S. (2016). Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Nature Scientific Reports* 6, 39589. https://doi.org/10.1038/srep39589

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology,* 37(11), 2098–2109. https://doi.org/10.1037/0022-3514.37.11.2098

Leong, Y. C., Hughes, B. L., Wang, Y., & Zaki, J. (2019). Neurocomputational mechanisms underlying motivated seeing. Nature human behaviour, 3(9), 962–973. https://doi.org/10.1038/s41562-019-0637-z

Leong Y.C., Chen, J., Willer, R., Zaki, J. (2020). Conservative and liberal attitudes drive polarized neural responses to political content. PNAS. 117(44) 27731-27739.

Moore-Berg, S. L., Parelman, J. M., Lelkes, Y., & Falk, E. B. (2020). Neural polarization and routes to depolarization. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(46), 28552–28554. https://doi.org/10.1073/pnas.2020107117

Nguyen, M., Vanderwal, T., & Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*, *184*, 161–170. https://doi.org/10.1016/j.neuroimage.2018.09.010

Overwalle, Social cognition and the brain: A meta-

Regev, M., Simony, E., Lee, K., Tan, K. M., Chen, J., & Hasson, U. (2019). Propagation of Information Along the Cortical Hierarchy as a Function of Attention While Reading and Listening to Stories. Cerebral cortex (New York, N.Y. : 1991), 29(10), 4017–4034. https://doi.org/10.1093/cercor/bhy282

Richardson, D.C., Griffin, N.K., Zaki, L. et al. (2020). Engagement in video and audio narratives: contrasting self-report and physiological measures. *Nature Science Reports* 10, 11298. https://doi.org/10.1038/s41598-020-68253-2

Silvers, J. A., Wager, T. D., Weber, J. & Ochsner, K. N. The neural bases of uninstructed negative emotion modulation. Soc Cogn Affect Neurosci 10, 10–18, doi: 10.1093/scan/nsu016 (2015).

Song, H., Finn, E.S., Rosenberg, M.D. (2021) Neural signatures of attention engagement during narratives and its consequences for event memory. PNAS. 118(33).

Van Baar, J. M., Halpern, D. J., & FeldmanHall, O. (2021). Intolerance of uncertainty modulates brain-to-brain synchrony during politically polarized perception. PNAS. 118 (20).

Yeo, B. T. T. , J. Tandi, M. W. L. Chee, Functional connectivity during rested wake-fulness predicts vulnerability to sleep deprivation. Neuroimage 111, 147–158 (2015).

Shigemune, Y., Tsukiura, T., Nouchi, R., Kambara, T., & Kawashima, R. (2017). Neural

mechanisms underlying the reward-related enhancement of motivation when remembering episodic memories with high difficulty. *Human brain mapping*, *38*(7), 3428–3443. https://doi.org/10.1002/hbm.23599

Yarkoni, T. , Speer,N. K., Zacks, J. M. (2008) Neural substrates of narrative comprehension and memory. *Neuroimage.* Vol. 41, 1408–1425.

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. Association for Psychological Science. 28(3). DOI: 10.1177/0956797616682029.

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