Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience

Luke J. Chang[1,*], Eshin Jolly[1], Jin Hyun Cheong[1], Kristina Rapuano[2], Nathan Greenstein[1],
Pin-Hao A. Chen[1], & Jeremy R. Manning[1]

[1] Department of Psychological and Brain Sciences
Dartmouth College
Hanover, NH, 03755
[2] Department of Psychology
Yale University
New Haven, CT 06511

*Corresponding author (luke.j.chang@dartmouth.edu)

Word count: 4872
Key words: vmPFC, endogenous processing, appraisal, affect

Abstract

How we process ongoing experiences is shaped by our personal history, current needs, and future goals. Consequently, brain regions involved in generating these subjective appraisals, such as the vmPFC, often appear to be heterogeneous across individuals even in response to the same external information. To elucidate the role of the vmPFC in processing our ongoing experiences, we developed a computational framework and analysis pipeline to characterize the spatiotemporal dynamics of individual vmPFC responses as participants viewed a 45-minute television drama. Through a combination of functional magnetic resonance imaging, facial expression tracking, and self-reported emotional experiences across four studies, our data suggest that the vmPFC slowly transitions through a series of discretized states that broadly map onto affective experiences. Although these transitions typically occur at idiosyncratic times across people, participants exhibited a marked increase in state alignment during high affectively valenced events in the show. Our work suggests that the vmPFC ascribes affective meaning to our ongoing experiences.

Introduction

Our brains process complex information with incredible speed and efficiency. This information can be broadly categorized into two distinct classes. First, our brains directly process *exogenous* information about the external environment by transducing physical phenomena (e.g., changes in energy, molecular concentrations, etc.) into sensory perceptions that allow us to generate and maintain a sense of what is happening around us [1,2]. Mental representations that are directly driven by the external world are likely to be highly *similar* across individuals who share the same sensory experience. Second, our brains also process *endogenous* information that reflects our current internal homeostatic states, past experiences, and future goals [3]. The integration of exogenous and endogenous information allows us to meaningfully interpret our surroundings, prioritize information that is relevant to our goals, and develop action plans [4]. Given the same input information, individuals may have unique interpretations, feelings, and plans, often leading endogenous representations to be *idiosyncratic* across individuals. How can we establish a broad functional commonality across individuals when these specific endogenous experiences may be unique to each individual?

This conceptual distinction between exogenous and endogenous processes may reflect a broad organizing principle of the brain. Recent work characterizing the functional connectomes of human and macaque brains has revealed topographic gradients from cortical regions that process exogenous information originating from a single sensorimotor modality (e.g., auditory, visual, somatosensory/motor) to transmodal association cortex such as the default-mode network (DMN) [5,6]. The DMN, which encompasses the ventromedial prefrontal cortex (vmPFC), posterior cingulate cortex (PCC), dorsomedial prefrontal cortex (dmPFC), and temporal parietal junction (TPJ) was initially identified to be metabolically active in the absence of any psychological task [7–11]. Subsequently, the DMN has been linked to a number of endogenous processes such as: engaging in spontaneous internal thought and mind wandering [12–14], thinking about the self [15,16], prospecting into the future [17,18], and recalling autobiographical memories [19–21].

Within the DMN, vmPFC responses are particularly variable across individuals. Voxels in this region show little evidence of temporal synchronization across individuals when listening to auditory stories [22] or watching movies [23,24]. Moreover, decoding accuracy using pattern classification is generally lower in the vmPFC compared to unimodal sensory areas [25], and functional connectivity patterns in resting state fMRI appear to be more variable across individuals in the vmPFC than in other areas of cortex [26,27].

Anatomically, the vmPFC projects directly to regions involved in affect and peripheral regulation, such as the hypothalamus, amygdala, ventral striatum, periaqueductal grey (PAG), and brainstem nuclei [28,29], as well as cognitive systems involved in conceptual processing such as the medial temporal lobe and dorsal and medial prefrontal cortex [30]. With connections to systems involved in both affective and conceptual processing, the vmPFC is thought to be instrumental in making self-relevant evaluative appraisals [3,4,15], which are critical in generating feelings [31–34], evaluating the value of a choice option [35–39], and comprehending abstract narratives [40–45]. The affective meaning generated by these appraisal processes are likely to be highly idiosyncratic across individuals.

Beyond an individual's unique appraisals, there are several potential methodological factors contributing to the variability of vmPFC responses across individuals. First, the vmPFC is particularly vulnerable to susceptibility artifacts due to its close proximity to the orbital sinus, which can increase noise due to signal dropout and spatial distortions. These artifacts can be slightly mitigated by optimizing scanning parameters and sequences [46–48]. Second, because the vmPFC is transmodal and integrates information from parallel processing streams, there may be greater variability in the temporal response profile of this region. While all brain regions exhibit some form of temporal integration, the length of temporal receptive windows are thought to be hierarchically organized based on levels of information processing [49]. In early sensory regions, this integration occurs over short temporal receptive windows that reflect rapidly changing inputs. By contrast, regions involved in abstract processing (e.g., narrative comprehension) exhibit much longer temporal receptive windows [50]. Together, this suggests that the heterogeneity observed in vmPFC signals arises from a combination of individual variability, spatial distortion, and temporal variation.

Establishing functional commonalities that transcend this variation in vmPFC response profiles requires identifying new ways to align signals across individuals. One such alignment procedure commonly used in the decision neuroscience literature employs an idiographic approach whereby experimental paradigms are customized to specific individuals [37,51–53]. Based on an individual's preferences, stimuli are mapped onto ordinal value signals, which are then used to identify brain regions that exhibit a linear correspondence [54]. Another promising alignment approach attempts to remap regions across participants based on common temporal response profiles to the same time-locked external stimulation. Functional alignment algorithms such as hyperalignment [55,56] and the shared response model [57] can be highly effective at improving spatial alignment across people based on common functional responses while still maintaining individual differences [58]. This approach might be particularly well suited for improving alignment in regions that process exogenous information and also potentially useful for mitigating variability arising from spatial distortions. However, because vmPFC responses reflect idiosyncratic appraisals based on endogenous information that are far removed from the exogenous input stimuli (e.g., goals, memories, etc), vmPFC activity patterns likely cannot be aligned across individuals in space or time using such approaches. Rather, aligning vmPFC responses across individuals might require developing a new approach that detects changes in latent states (e.g., mental, physiological).

In the present study, we investigate how the vmPFC processes endogenous information while viewing a rich naturalistic stimulus. Participants watched a 45-minute pilot episode of a character-driven television drama (*Friday Night Lights*) while undergoing functional magnetic resonance imaging (fMRI). This show was chosen in particular for its engaging story arc, parallel plotlines, and range of dynamics in affective experiences. We test the hypothesis that vmPFC activity reflects changes in latent states associated with the interaction between exogenous and endogenous processes. These states are hypothesized to reflect appraisals of meaningful plot elements as a narrative unfolds (e.g., “that is amazing!”, “that is terrible!”), that ultimately result in the emergence of feelings. More specifically, we predicted: (a) greater heterogeneity in spatiotemporal dynamics in vmPFC across individuals relative to primary sensory regions; (b) longer processing timescales in the vmPFC compared to primary sensory regions; and (c) that changes in vmPFC states correspond to meaningful psychological experiences.

Results

Idiosyncratic spatiotemporal dynamics of the vmPFC

We hypothesized that the vmPFC plays a critical role in processing endogenous information by integrating information from the external world with internal states, past experiences, and future goals. Unlike regions that directly process exogenous information (e.g. early visual cortex), we predict that there should be little across-participant consistency in vmPFC responses during natural viewing. We tested this hypothesis using two separate studies. Participants in Study 1 (n=13) were scanned on a Philips 3T scanner while watching the first two episodes of *Friday Night Lights*. Participants in Study 2 (n=35) were scanned on a Siemens Prisma 3T scanner while watching only the first episode of the same show. First, we examined consistency in temporal responses across the 45-minute episode using inter-subject correlations (ISC; see *Methods*) [23]. To reduce computational complexity, all analyses were performed on 50 non-overlapping regions based on a whole-brain parcellation using meta-analytic coactivations from over 10,000 published studies [59,60]. We then computed the average pairwise temporal similarity of mean activity within each parcel across all participants. This yielded a number-of-subjects by number-of-subjects correlation matrix for each of the 50 parcels. ISC is simply the average of these pairwise similarities across participants. In Study 1, early sensory regions (e.g., V1) exhibited the highest level of temporal synchronization (mean ISC=0.37, sd=0.13), whereas PFC parcels largely showed little evidence of synchronization across participants (mean vmPFC ISC=0.06, sd=0.07; Figure 1A). These results were replicated in Study 2; V1 was associated with high levels of temporal ISC (mean ISC=0.33, sd=0.15), whereas the vmPFC exhibited a low degree of synchronization (mean r=0.07, sd=0.05; Figure 1B). The overall spatial pattern of ISC across brain parcels was highly similar across the two studies, $r(48)=0.99$, $p < 0.001$.

Although vmPFC activity was idiosyncratic when considering the full duration of the episode, it is possible that activity might briefly align during particular moments of the episode. For example, an engaging pivotal moment in the story might evoke strong responses across individuals, whereas less salient moments might encourage person-specific endogenous processing. We defined a metric of instantaneous synchronization in Study 1 using the average spatial ISC computed across patterns of voxel activity within each ROI during each time point (i.e., each TR; see *Methods*). Unlike other methods of dynamic synchrony, this approach does not require specifying a sliding window length or shape [61] or restricting the analysis to a specific temporal frequency band [62]. Instead, it characterizes the average alignment of spatial activity patterns across participants. Across all timepoints, we observed a modest spatial ISC in sensory regions, such as V1 (mean r=0.1, sd=0.05) and virtually no synchronization in the vmPFC (mean r=0.03, sd=0.03). It is possible that spatial distortions from susceptibility artifacts might have increased variations in spatial configuration of voxels in the vmPFC for each participant or that anatomical alignment techniques were insufficient for aligning fine-grain spatial patterns across individuals. For this reason, we functionally aligned each region into a common 1000-dimensional space by fitting the shared response model [57] to independent data (i.e., episode 2). Consistent with previous work, we found that functional alignment dramatically improved spatial synchronization across participants in regions involved in exogenous processing such as V1 (mean r=0.23, sd=0.16), but had no appreciable impact on vmPFC synchronization (mean r=0.02, sd=0.03) [55–57].

These results demonstrate that vmPFC activity patterns do not appear to align spatially or temporally across people, which could arise from at least two possible explanations. First, it is possible that incoming stimulus-driven information simply may not directly impact vmPFC activity. For example, the vmPFC might play a role in representing or driving endogenous thoughts that are specifically *unrelated* to the exogenous experience (e.g., mind-wandering). Alternatively, the vmPFC may be involved in representing information or thoughts related to the stimulus (e.g., appraisals), but this information may be idiosyncratic to each individual.

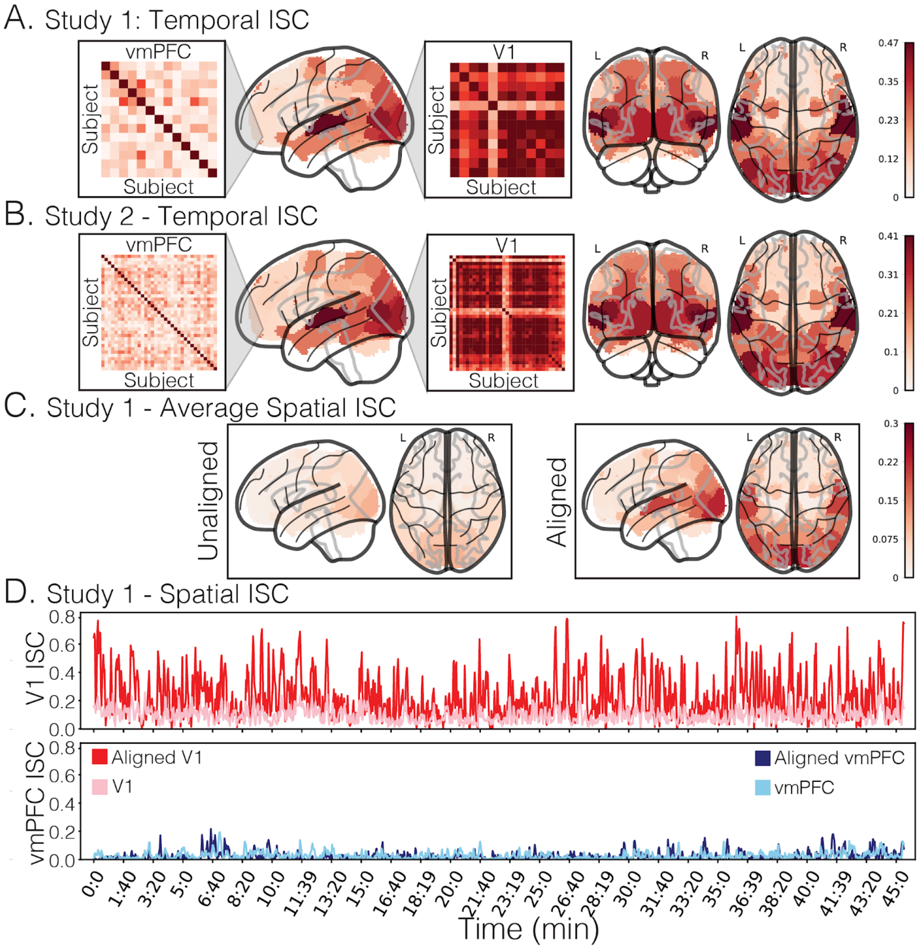


Figure 1. Intersubject correlations (ISC). *A) Temporal ISC for each ROI for Study 1 (n=13). Temporal ISC reflects the average pairwise correlation between each individual participant's activation time series while viewing the 45-minute episode. Heatmap blowouts depict all pairwise temporal ISCs between each participant for the vmPFC and V1 ROIs. B) Temporal ISC for each ROI for Study 2 (n=35); same format as Panel A. C) Spatial ISC across participants averaged across TRs for Study 1. Spatial ISC reflects the pairwise spatial similarity of activity patterns across participants within each ROI at each TR. Functional alignment in the right sub-panel was performed using the Shared Response Model* [57] *on an independent dataset. D) Spatial ISC at each TR across the 45-minute episode for the vmPFC and V1 ROIs for Study 1. Lighter colors denote anatomical alignment of voxels across participants. Darker colors denote functional alignment.*

Temporal recurrence of spatial patterns

To more precisely characterize the structure of vmPFC responses across the episode, we examined the temporal recurrence of each individual's' unique patterns of vmPFC activity. For each participant, we computed a number-of-timepoints by number-of-timepoints correlation matrix reflecting the similarity of spatial patterns of vmPFC activity over the 45-minute episode. This analysis can aid in identifying time points during the show where a participant demonstrates a similar pattern of vmPFC activity, which might reflect a recurrence of a similar psychological state.

As shown in Figure 2A, this temporal correlation matrix exhibited a striking block diagonal structure. Specifically, the vmPFC exhibited periods of stable activity that persisted for up to several minutes, punctuated by moments of rapid change (See Figs S1-S4 for all participants). This suggests that there are likely several different spatial configurations of vmPFC activity that are maintained for long periods of time. To determine if these response profiles are consistent across participants, we computed a one-sample *t*-test on Fisher *z*-transformed correlation values at each pair of time-points across all participants (Study 2, n=35). After correcting for multiple comparisons, we found very few time points that exhibited a consistent recurrence structure across participants (Figure 2B). Rather than treating each time point independently, we also examined whether the relative pattern of spatiotemporal dynamics was consistent across participants (i.e., the similarity of the lower triangles of the time x time spatial similarity matrix) (Figure 2C) and found little evidence of spatiotemporal synchronization (mean correlation=0.03, sd=0.01). In comparison, we observed a modest but reliable spatiotemporal synchronization across participants in V1 (mean correlation=0.13, sd=0.07). Together, these results indicate that individual participants appear to share a block diagonal structure revealing long periods of sustained vmPFC activity, but the specific dynamics and patterns of recurrence appear to be unique to each individual participant.

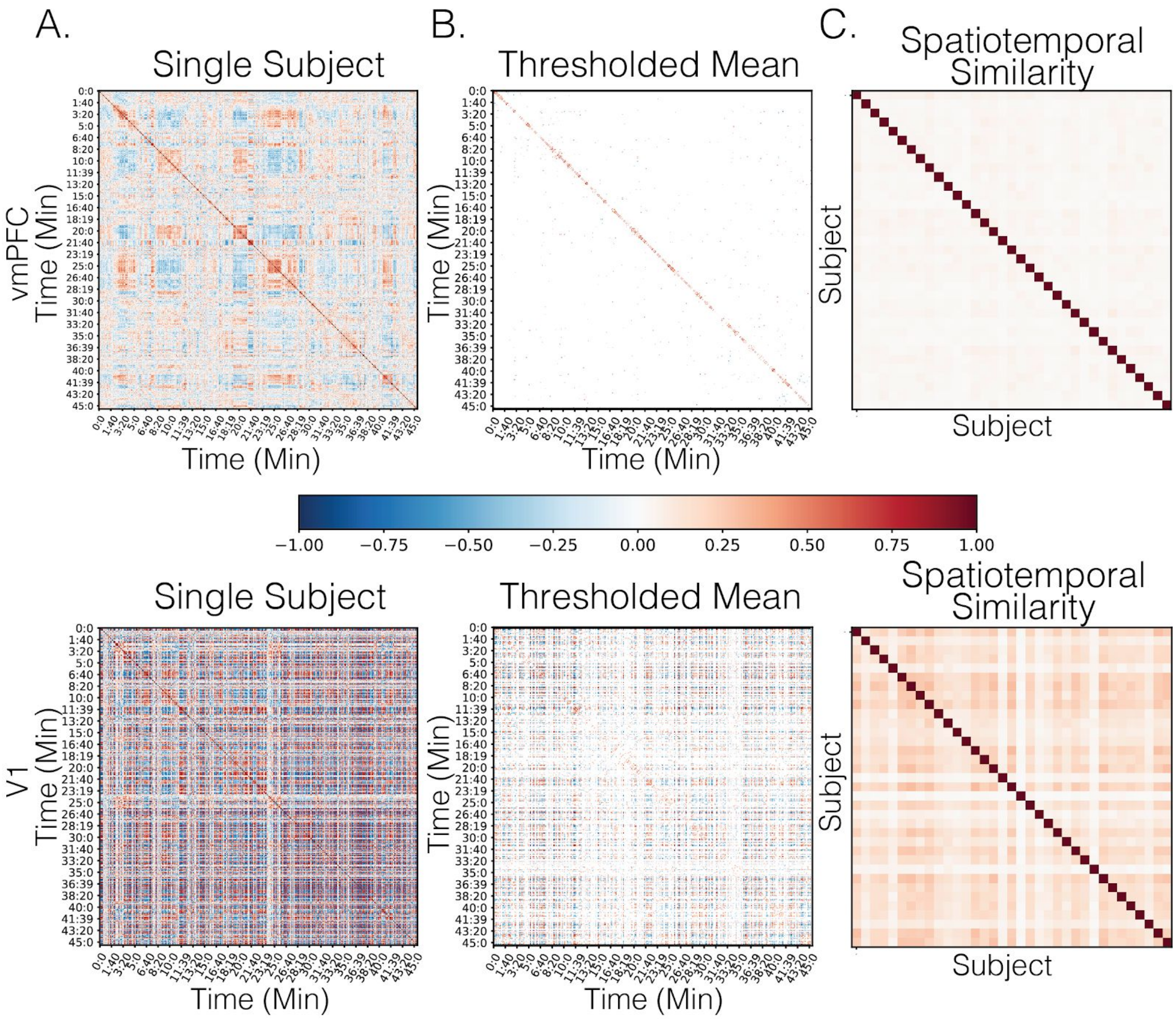


Figure 2. Temporal recurrence of spatial patterns*. A) A recurrence matrix of vmPFC spatial patterns for a representative participant. B) A map of t-tests over each pair of time-points across all participants from Study 2 (n=35) thresholded using FDR q < 0.05 (vmPFC: p=0.00042; V1: p=0.012). C) A subject-by-subject similarity matrix representing the consistency of spatiotemporal recurrence patterns across participants. Colorbar is scaled between [-1,1], note that the scaling on the t-map depicted in panel B is one order of magnitude larger [-10,10].*

The vmPFC has a long temporal autocorrelation function

To quantify the rate of temporal integration of information in the vmPFC, we estimated the temporal autocorrelations of activity throughout the brain. Specifically, we estimated the amount of autocorrelation for each individual voxel in the brain up to 50 TR lags. We then fit an exponential function to the median of these estimates within each ROI using a robust estimator and calculated the number of time-points before the function decayed to an arbitrary threshold of 0.1 (see *Methods*). Figure 3A illustrates the estimated amount of autocorrelation at the voxel level across each of the 50 parcels. We found that the degree of autocorrelation in voxels in the vmPFC (6 TRs) appeared to be similar to other regions such as V1 (6 TRs).

Next we sought to evaluate whether the multi-voxel spatial pattern within each ROI might be integrating information over a longer time window. While individual voxel responses changed at a similar rate in the vmPFC and V1, it is possible that the dynamics of the pattern of population responses across voxels within an ROI might differ across regions. To do so, we applied the same autocorrelation estimation procedure employed above using the multi-voxel pattern within each ROI. We found that the vmPFC displayed much more gradual changes in pattern responses (10 TRs; 20s lag to reach an average autocorrelation of 0.1) compared to sensory regions such as V1 (4 TRs; 8s). This suggests that information encoded in spatial patterns in the vmPFC is being integrated over longer periods of time than sensory regions. Unlike previous work that has demonstrated similar findings by causally manipulating presented information at different temporal scales (e.g., scrambling words, sentences, or paragraphs) [22,50], this method allows us to estimate the longevity of spatial patterns within an ROI in a context-general way (including non stimulus-driven contexts).

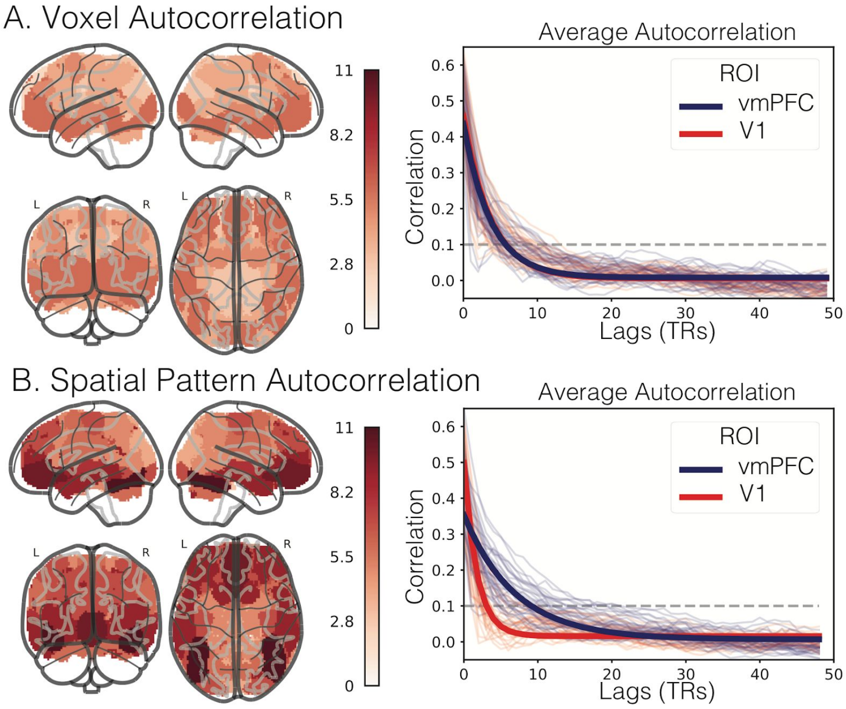


Figure 3. Temporal autocorrelation. *A) The left panel displays the lag in TRs taken to reach an autocorrelation of 0.1 (arbitrarily selected). Larger values indicate more gradually changing voxel responses. In the right panel, we display the median autocorrelation function across voxels for vmPFC (blue) and V1 (red) for each participant. The darker lines reflect an exponential function fit to the median of the distribution across voxels for each ROI, and the thin lines represent autocorrelation functions for each individual participant. B) Identical analysis to Panel A, but for spatial patterns across all voxels within each ROI rather than for individual voxels.*

Spatial alignment of vmPFC latent states

The ISC analyses described above establish that the vmPFC does not show consistent responses across participants in its average time course or spatial patterns using either anatomical or functional alignment. Our temporal recurrence analysis revealed that each participant exhibited a similar block diagonal structure characterized by long periods of sustained vmPFC activity that recurred at multiple points throughout the episode. This suggests that the vmPFC may slowly transition between different states. Though the overall patterns of recurrence did not appear to synchronize across participants, it is possible that a subset of these states might be shared across participants, but expressed at different moments in time. We used Hidden Markov Models (HMMs) to segment patterns of activations in the vmPFC into discrete latent states. In implementing this analysis, we made several simplifying assumptions. We assumed that (a) voxel activations were drawn from orthogonal Gaussian distributions, (b) the transitions in latent states had the form of a first-order Markov chain, and (c) all participants experienced the same number of states. For each participant, we fit a separate HMM to a set of orthogonal principal components that explained 90% of the variance of the participants' multivariate vmPFC time series. We used maximum *a posteriori* (MAP) estimation to determine the probability that the participant was in each given state at each moment of viewing the episode (Figure 4A) and aligned the states across participants by maximizing the spatial similarity using the Hungarian algorithm [63]. To determine the number of states for the HMM estimation procedure, we identified the number of states that most frequently resulted in the largest increase in individual model fit across participants. Overall, we found that a low number of states (k=3) appeared to consistently yield the biggest jump in model fit across participants in both V1 and vmPFC in Study 1.

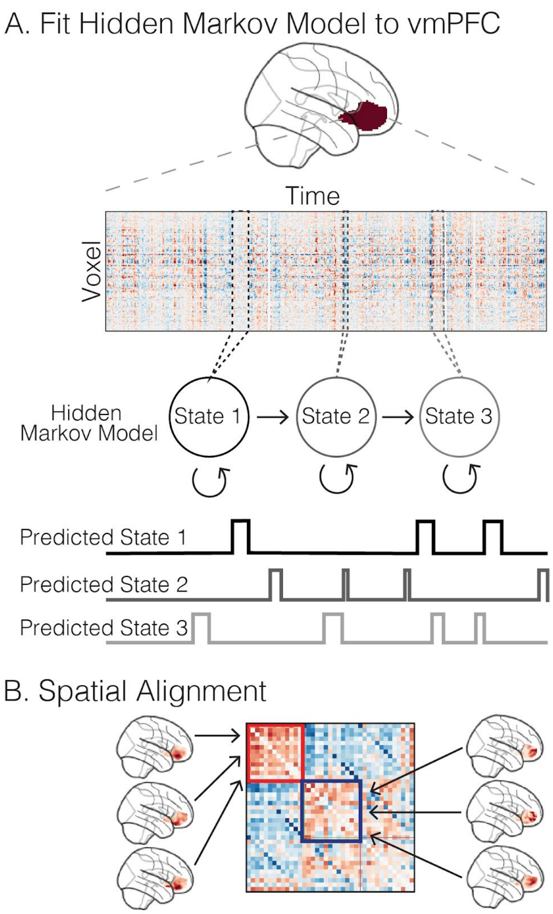


Figure 4. *HMM estimation procedure. A) A schematic of how an HMM is fit to each participant's ROI time series. We fit an HMM with a prespecified number of states to a PCA-reduced time series of voxel responses. This yielded estimates of the starting state, a transition matrix reflecting the likelihood of transitioning into any state at time t from time t-1, and a pattern of activity (assumed to be an emission from an orthogonal multivariate Gaussian distribution). We then applied the model to the same data to predict the MAP latent state at each moment in time. B) We aligned states across participants by maximizing the spatial similarity of the estimated states.*

Interestingly, in direct contrast to the spatial ISC analyses, the HMM analysis revealed that a subset of the latent vmPFC states appeared to be shared across participants (Fig. 5A). The states with the highest spatial consistency in vmPFC and V1 are depicted in Figure 5B, and the average spatial pattern is shown in Figure 5C. The magnitude of the HMM spatial similarity is in the upper range of what we observed in our ISC analyses.

Next we examined whether these latent states exhibited any temporal consistency using a concordance metric, which we define as the proportion of participants in a given state at each moment of the episode. We plot the overall concordance of the vmPFC and V1 states that exhibited the highest spatial consistency across time in Figure 5D. This analysis indicates moderate state concordance in both vmPFC (Study 1 mean=0.32, std=0.15; Study 2 mean=0.17, std=0.08) and V1 (Study 1 mean=0.27, std=0.19; Study 2 mean=0.29, std=0.18; see Table S1 for all states). The timecourses of these state concordances appeared to replicate across both studies. We observed a strong correlation in the temporal concordance between Studies 1 and 2 for V1, (r(1362)=0.70, p< 0.001) and a more modest correlation for vmPFC (r(1362)=0.22, p< 0.001). Together, these analyses suggest that the vmPFC has a consistent spatial pattern for specific states, but these states tend not to be experienced by the same participants at the same moments in time, except at a few specific timepoints (i.e., local maxima in Figure 5D).

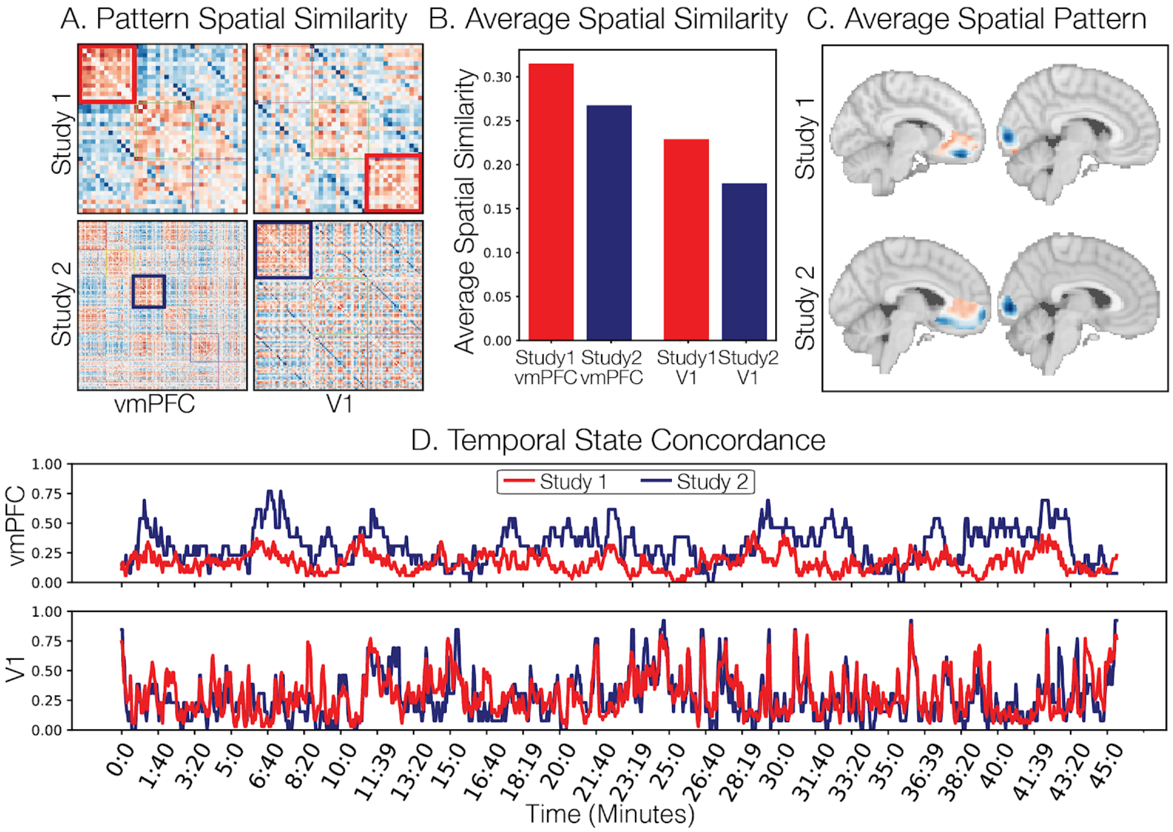


Figure 5. Alignment of spatial patterns corresponding to latent states. *A) Subject by subject similarity matrix of individual spatial patterns identified using an HMM in vmPFC and V1 in Study 1 (n=13) and Study 2 (n=35). Outlines highlight the states with the highest intersubject spatial similarity. B) Average spatial similarity across subjects within the highlighted states in panel A. C) Average spatial pattern across participants for vmPFC and V1. D) Temporal concordance of the highlighted predicted states across individuals within each study. Higher values indicate a higher proportion of participants in the same putative latent state within each study (red: Study 1; blue: Study 2).*

vmPFC state concordance is associated with affective experience

The above results show that the vmPFC HMM states were similar with respect to their spatial activity patterns and, to a lesser extent, their time courses, across two studies collected on separate scanners. We next sought to understand what types of thoughts or processes might be reflected by these vmPFC states using three distinct approaches.

First, we performed an informal reverse correlation analysis [23] to identify which scenes from the show corresponded to the periods of high across-participant state concordance (Fig. 6A). We found that high concordance intervals occurred during scenes with strong narrative importance for the episode. For example, these key moments include scenes such as the start of the important football game (~28 min), the protagonist getting severely injured (~35 min), and the backup quarterback leading the team to a surprising victory despite the setbacks (~42 min). Further, these states appear to be maintained for long periods of time, on the order of multiple minutes, and are consistent across different participants and scanners.

Second, we collected an additional independent dataset outside of the scanner to gain further insight into the identified vmPFC states based on facial expression behavior. In Study 3 (n=20), participants watched the first episode while their facial expressions were video recorded [64]. We used a computer vision algorithm to identify 20 facial action units (AUs), a standardized system to describe the intensity of facial muscle movements [65], for each frame of the video during the episode [66]. We used principal components analysis (PCA) to project these AUs into a set of orthogonal components and correlated each component with the two vmPFC states associated with the highest level of spatial similarity across participants (Figure S5; Table S2). This allowed us to identify two distinct combinations of action units that maximally aligned with the vmPFC state concordance across both Study 1 (state 1: $r=0.13$, $p < 0.001$; state 2: $r=0.26$, $p < 0.001$) and Study 2 (state 1: $r=0.11$, $p < 0.001$; state 2: $r=0.2$, $p < 0.001$) and survived multiple comparisons correction (FDR $q < 0.05$). State 1 is most associated with increased intensity of the AU12 (zygomaticus major; PCA loading=0.47), which pulls the lip corners to form a smile [67]. State 2, by contrast, is most associated with AU25 (depressor labii; PCA loading=0.38), AU4 (corrugator; pca loading=0.16), AU 5 (levator palpebrae superioris; PCA loading=0.16), which parts the mouth, lowers the brow, and raises the upper eyelid respectively [67]. This suggests that vmPFC state 1 is associated with a feeling of positive affect, whereas state 2 is associated with a feeling of surprise, concern, or worry. A visual depiction of the AU loadings is displayed in Figure 6C (also see Figure S6).

Third, we collected an additional dataset to examine self-reported subjective feelings. In Study 4, participants (n=188) recruited from Amazon Mechanical Turk watched the same television episode used in Studies 1, 2, and 3. The 45-minute episode was periodically randomly paused (n=12) and participants were instructed to report their subjective feelings on 16 different emotional dimensions. We used a collaborative filtering technique (see *Methods*) to predict how each participant was feeling for the remaining time points across all 16 dimensions. Similar to the facial expression data, we estimated a full set of principal components using these emotion dimensions and separately correlated each component with the two vmPFC states associated with the highest state concordance (Table S3; Figure S7). Consistent with the facial expression findings observed in Study 3, we identified two components that correlated with the vmPFC state concordance time series in both Study 1 (state 1: r=0.12, p < 0.001; state 2: r=0.33, p < 0.001) and Study 2 (state 1: r=0.19, p < 0.001; state 2: r=0.19, p < 0.001). State 1 was associated with a component that included high loadings for positive feelings such as joy, pride, hope, satisfaction, elation, and relief, whereas state 2 was associated with a separate component that included high loadings for negative feelings such as sadness, guilt, envy, and pride (Figure 6D).

Overall, we observe strong evidence that participants collectively experience at least two latent affective states while watching the episode, with convergent interpretations across vmPFC activity, facial expressions, and subjective ratings (Figure 6B). The components for the facial expressions and subjective ratings were consistently expressed across participants at specific scenes. This suggests that vmPFC states are linked to exogenous inputs and are more likely to align across individuals in contexts that elicit more intense affective responses [4,33,68].

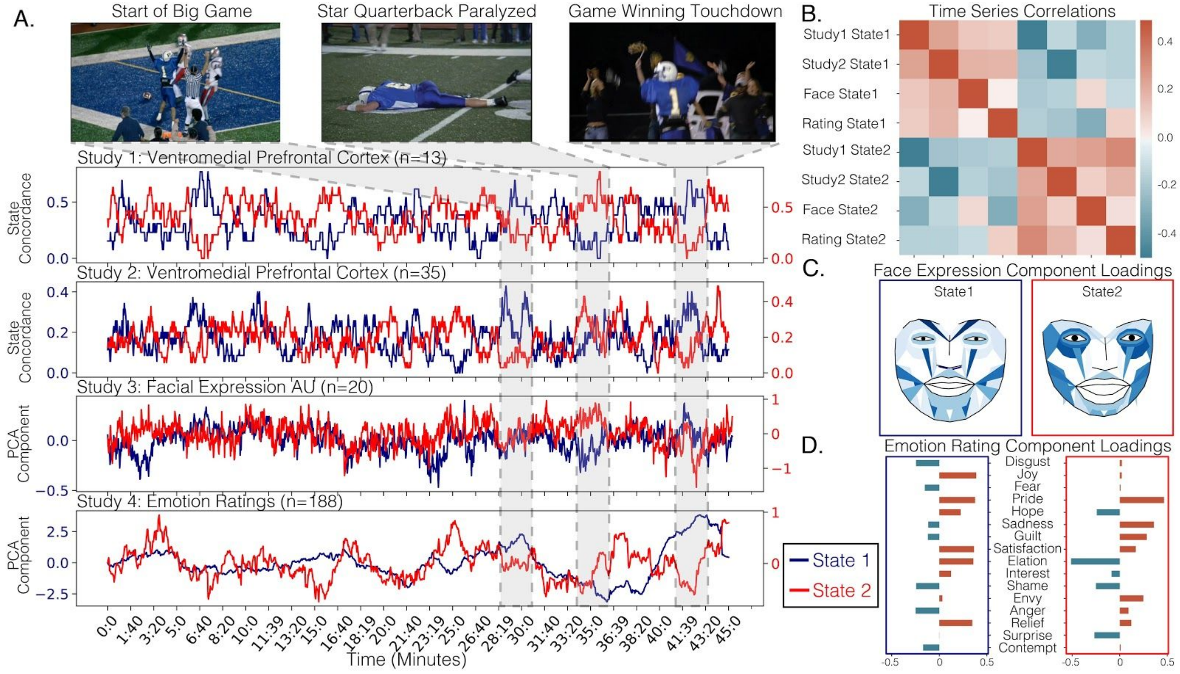


Figure 6*. Emotions associated with increased state concordance. A) State concordance at each moment in time for the two states with the highest overall spatial similarity in the vmPFC (Studies 1 and 2). Two additional behavioral studies linking these time points to affective states were estimated via facial expressions (Study 3) and subjective ratings (Study 4). For Studies 3 and 4, data were reduced using PCA. The component that most strongly correlated with vmPFC state concordances is visualized for each state (State 1, blue; State 2, red). B) Overall temporal similarity between each pair of time series data across the four studies. C) Loadings of facial action units onto each facial expression component. D) Loadings of subjective ratings onto each component. Note: to maintain compliance with biorxv preprint policy, example participant face expressions have been removed from this figure.*

Discussion

In this study, we investigated how the vmPFC processes information while viewing a rich, evocative, multimodal naturalistic stimulus. We found evidence that the spatiotemporal response profiles of the vmPFC were heterogeneous across individuals, even after applying spatial and temporal alignment algorithms to the data. We also found that individual vmPFC spatial patterns appeared to persist for long periods of time and recurred periodically over the course of the episode. The long window of temporal integration in the vmPFC was specific to these spatial patterns and not observed in individual voxels. We used HMMs to segment patterns of vmPFC activity into discrete latent states and found that a subset of these states appeared to be shared across individuals. Although these states were most often expressed at different moments in time across individuals, scenes that evoked strong affective responses appeared to synchronize these latent states across participants, which replicated in an independent sample collected on a different scanner. Moreover, we observed converging evidence that these states synchronized with specific patterns of facial expressions and subjective affective ratings in two additional independent samples viewing the same television episode outside of the scanner. Taken together, these vmPFC activity states appear to reflect latent endogenous psychological processes involved in conferring affective meaning to ongoing events. This helps to explain the heterogeneity of this region's responses across people, and also has implications for experiments designed to probe the function of this region.

The vmPFC appears to be broadly involved in making meaning of the external world [3,4]. It has been associated with representing and navigating cognitive maps and conceptual spaces [69–71] and interpreting narratives [41,42,45,72], which often requires abstracting temporal structure of abstract models of situations [44]. At its most basic level, making a simple evaluation about the valence of an object or event (i.e., is it good or bad?) requires making an interpretation based on an individual's past experiences, current homeostatic states, and future goals. This appraisal process appears to be directly linked to the vmPFC [35,37]. For example, evaluating a particular food option requires integrating multiple features such as taste, cost, and caloric content [73] with broader goals such as eating healthy [74] and internal homeostatic drives [75]. Beyond food, the vmPFC appears to code general representations of subjective valence appraisals that transcend stimulus modalities [33,76,77]. Countless studies have implicated the vmPFC in processing the valence of emotional memories [32] and chemosensory sensory olfaction [78,79], and gustation [53] signals. These types of affective experiences of feelings, smells, and tastes can be causally induced with direct electrical stimulation of this region via intracranial electrodes implanted in patients with epilepsy [34]. In this study, we find the highest level of synchronization of vmPFC states at the most emotionally evocative narrative events, which is consistent with work using pupillometry indicating that people temporally synchronize mental states at affectively arousing events [68].

Extracting meaning from our subjective experiences requires integrating multimodal information at a variety of timescales. Consistent with work that has demonstrated a cortical hierarchy of processing temporal information [22,49,50,80], we find that the vmPFC maintains stable states over timescales on the order of tens of seconds to minutes. This is notable for two reasons. First, this slow processing does not appear to be reflected in the autocorrelations of individual voxels, but rather only emerges when examining temporal autocorrelations of spatial patterns. Second, standard preprocessing protocols for fMRI data routinely apply high pass filters, which would have removed this signal [81]. Our autocorrelation approach complements experimental designs that shuffle information at multiple timescales [22,50]. While most methods studying neural processes have focused on studying fast signals that are closer to the firing of individual neurons, fMRI may be be particularly well suited for studying slower signals that involve integrating information across time such as transforming experiences into memories [24,82].

Participants exhibited unique spatiotemporal response patterns in the vmPFC while viewing a common television episode. We do not believe that this can be solely accounted for by methodological challenges (e.g., susceptibility artifacts, variations in hemodynamic response functions). This presents a significant challenge to traditional neuroimaging analysis methods that assume a common response profile across participants (e.g. two-level univariate analyses [83], resting state analyses [84], multivariate pattern analysis approaches [85,86], intersubject synchrony [23,41,87], and functional alignment [55–57,87]). Because subjective endogenous experiences are not typically shared across participants, regions like the vmPFC that appear to exhibit idiosyncratic stimulus-driven activity may be mischaracterized by these approaches. Our state-based analysis framework provides a means of characterizing this endogenous stimulus-driven activity, even when the response patterns do not align spatially or temporally across individuals, or to the external stimulus. Our approach might be also useful in translational applications where patient groups are often highly heterogeneous compared to healthy controls.

We note several limitations of our work. First, we assume that the anatomical demarcation of the vmPFC is consistent across all participants. We used a large ROI that was defined by a meta-analytic coactivation [60]. However, the vmPFC can be further subdivided into more functionally specific regions [29], each which may have a unique response profile. We note that our pattern-based approach leverages variability across these subregions to identify state changes. Second, our HMM analyses assume that all participants experience the same number of latent states while viewing the television episode. This simplifying assumption was necessary in order to constrain the model and to evaluate spatial and temporal consistency across participants. However, we do not believe that there is a ground truth “correct” number of states, nor that all participants experience the same states. Finally, we do not directly link emotional experiences to vmPFC activity within a single participant. Instead we rely on providing converging evidence across different studies which are time-locked to the stimulus presentation. Video measurements of facial expressions while scanning is difficult due to occlusions from the head coil. However, simultaneous facial electromyography with fMRI has shown promise by demonstrating a relationship between displays of negative affect via facial expression and vmPFC activity [88] and is an area we are continuing to explore.

In summary, we find evidence indicating heterogeneity in exogenous and endogenous information processing across the brain. The vmPFC appears to be involved in ascribing meaning to ongoing experiences with respect to the unique endogenous thoughts individuals bring to bear on processing each new moment. These processes may be reflected in spatial patterns [89–93], which appear to be unique to each appraisal and experience, and change at relatively slow timescales. However, we observed increased synchronization of these spatial patterns at emotionally salient scenes [68], suggesting that manipulating internal states (e.g., homeostatic states, memories, goals, or feelings) can increase synchronization of the vmPFC across people. Our work demonstrates the potential of characterizing the temporal dynamics of spatial patterns to study the experiences of a single individual elicited during naturalistic multimodal scanning paradigms.

Methods

Study 1

Subjects

Thirteen subjects (*mean [sd] age = 24.61 [4.77]; 6 female)* were undergraduate and graduate students at Dartmouth College participating for either monetary compensation ($20/hr) or for partial course credit. All participants gave informed consent in accordance with the guidelines set by the Committee for the Protection of Human Subjects at Dartmouth College.

Procedure

Participants completed two separate sessions. In the first session, participants watched episodes 1 and 2 of the television drama *Friday Night Lights* (FNL) while undergoing functional Magnetic Resonance Imaging (fMRI). This show was chosen in particular for its engaging story arc, parallel plotlines, and range of affective experience dynamics. Participants viewed each episode with a break in between followed by two runs of passively viewing images of characters. In a second session, participants watched episodes 3 and 4 outside of the scanner while we recorded their facial expressions using a head-mounted camera. All episodes were approximately 45-minutes in length. After each episode, participants made ratings of characters on a computer. The character viewing session, facial expressions of episodes 3 and 4, and the character ratings will be discussed in more detail in forthcoming manuscripts; those data were not reported in the present manuscript.

Imaging Acquisition

Data were acquired at the Dartmouth Brain Imaging Center (DBIC) on a 3T Philips Achieva Intera scanner (Philips Medical Systems, Bothell, WA) with a 32-channel phased-array SENSE (SENSitivity Encoding) head coil. Raw images were saved directly to NIfTI format. Structural images were acquired using a high-resolution T1-weighted 3D turbo field echo sequence: TR/TE: 8200/3.7ms, flip angle = 8°, resolution = 0.938 x 0.938 x 1.00 mm voxels, matrix size = 256 x 256, FOV = 240 x 240mm^2. Functional blood-oxygenation-level-dependent (BOLD) images were acquired in an interleaved fashion using single-shot gradient-echo echo-planar imaging with pre-scan normalization, fat suppression and an in-plane acceleration factor of two (i.e. SENSE 2): TR/TE: 2000/30ms, flip angle = 75°, resolution = 3mm^3 isotropic voxels, matrix size = 80 x 80, FOV = 240 x 240mm^2, 40 axial slices with full brain coverage and no gap, anterior-posterior phase encoding. Functional scans were acquired during each episode in a single continuous run per episode (episode 1: 1364 TRs; episode 2: 1317 TRs).

Study 2

Participants

Thirty-five participants (*mean [sd] age=19.0 [1.07] years; 26 female*) were recruited from undergraduate introductory psychology and neuroscience courses for either course credit or monetary compensation at Dartmouth College. All participants gave informed consent in accordance with the guidelines set by the Committee for the Protection of Human Subjects at Dartmouth College.

Procedure

The experimental procedure was identical to Study 1, with the exception that each participant only watched the first episode of FNL (45 minutes) while undergoing fMRI.

Imaging Acquisition

Data were acquired at the Dartmouth Brain Imaging Center (DBIC) on a 3T Siemens Magnetom Prisma scanner (Siemens, Erlangen, Germany) with a 32-channel phased-array head coil. Raw DICOM images were converted to NIfTI images and stored in the brain imaging data structure (BIDS) format [94] using ReproIn from the ReproNim framework [95] [95]. Structural images were acquired using a T1-weighted, single-shot, high-resolution MPRAGE sequence with an in-plane acceleration factor of two (i.e. GRAPPA 2) and pre-scan normalization: TR/TE: 2300/2.32ms, flip angle = 8°, resolution = 0.9mm^3 isotropic voxels, matrix size = 256 x 256, FOV = 240 x 240mm^2. Functional blood-oxygenation-level-dependent (BOLD) images were acquired in an interleaved fashion using gradient-echo echo-planar imaging with pre-scan normalization, fat suppression and an in-plane acceleration factor of two (i.e. GRAPPA 2), and no multiband (i.e.

simultaneous multi-slice; SMS) acceleration: TR/TE: 2000/25ms, flip angle = 75°, resolution = 3mm^3 isotropic voxels, matrix size = 80 x 80, FOV = 240 x 240mm^2, 40 axial slices with full brain coverage and no gap, anterior-posterior phase encoding. All functional scans were acquired in a single continuous run (1364 TRs).

Imaging Preprocessing

All MRI data for both studies 1 and 2 were preprocessed using a custom pipeline written using nipype [96]. The pipeline involved trimming the first five non-steady-state TRs of functional each scan using nipy [97] and subsequently realigning each functional volume to the mean functional volume in a two-pass procedure implemented via FSL's MCFLIRT [98]. In parallel, brain-extraction (i.e. skull-stripping) was performed on T1-weighted structural images using ANTS [99]. Subsequently, transformations for linearly coregistering realigned functional volumes to skull-stripped structural images, and non-linearly normalizing skull-stripped structural images to the ICBM 152 2mm MNI template were calculated using ANTS. These transforms were then concatenated and applied in a single step using basis spline interpolation in ANTS. Data were spatially smoothed using a 6mm FWHM Gaussian kernel implemented using fslmaths [100]. To perform denoising, we fit a voxel-wise general linear model (GLM) for each participant using custom software implemented in Python (https://github.com/cosanlab/nltools). We used this toolbox to remove variance associated with the mean, linear, and quadratic trends, mean activity from a cerebral spinal fluid (CSF) mask, the effects of motion estimated during the realignment step using an expanded set of 24 motion parameters (6 demeaned realignment parameters, their squares, their derivatives, and their squared derivatives), motion spikes between successive TRs identified using ART [101] and global signal-intensity spikes greater than three standard-deviations above the mean intensity between successive TRs (Note: Study 1 did not include global signal intensity spikes). All analyses occured on the residual time-series after GLM estimation. Code for this custom pipeline is available at: https://github.com/cosanlab/cosanlab_preproc

Parcellation

All imaging analyses were conducted over a set of 50 parcels (parcellation available at http://neurovault.org/images/39711). The parcellation was created by performing a whole-brain parcellation of the coactivation patterns of activations across over 10,000 published studies available in the Neurosynth database [59,60]. The use of a parcellation scheme has several advantages over the more conventional voxelwise and searchlight approaches. First, it is several orders of magnitude less computationally expensive as analyses are performed only 50 times compared to 352k voxels. Second, the parcels are non-overlapping and contain bilateral regions that reflect functional neuroanatomy, whereas a searchlight approach is limited to local spheres that do not adapt to different areas of cortex.

Imaging Analyses

Intersubject Correlation

We take the intersubject correlation (ISC) to be the lower triangle of the pairwise correlation matrix of signals between each subject.

$$\mathrm{corr}(x,y) = \frac{\sum_{n=1}^{N}(x_n-\bar{x})(y_n-\bar{y})}{\sigma_x\sigma_y} \qquad \text{(eq 1)}$$

$$\sigma_z = \sqrt{\sum_{n=1}^{N}(z_n-\bar{z})^2} \qquad \text{(eq 2)}$$

$$\bar{z} = \frac{1}{N}\sum_{n=1}^{N} z_n \qquad \text{(eq 3)}$$

$$\mathrm{ISC}_{1\ldots N} = R\left[\frac{2\sum_{i=1}^{N}\sum_{j=1}^{i} Z\left(\mathrm{corr}(S_i,S_j)\right)}{N^2-N}\right] \qquad \text{(eq 4)}$$

where

$$Z(r) = \frac{\log(1+r)-\log(1-r)}{2} \qquad \text{(eq 5)}$$

is the Fisher z-transformation and

$$R(z) = \frac{\exp(2z)-1}{\exp(2z)+1} \qquad \text{(eq 6)}$$

is its inverse.

S is a matrix of subject timeseries data from a total of *N* subjects and *i* and *j* indicate the subject indices. We calculated the ISC for each subject's mean vmPFC timeseries, their vmPFC spatial pattern at each moment in time, and their unique spatiotemporal pattern, defined as the vectorized lower triangle of the participant's spatial recurrence matrix across all voxels in their vmPFC. We also carried out an analogous series of calculations with each participant's V1 activity patterns. The similarity of spatiotemporal patterns has an interesting connection to computing distance correlation [102,103], a measure of general (i.e. possibly non-linear) statistical dependence between two signals of arbitrary dimensionality.

Functional Alignment

We performed functional alignment on Study 1 using the shared response model (SRM) [57]. The deterministic SRM algorithm performs a joint PCA and can identify a reduced set of functional response components that maximally align across participants. We chose an arbitrarily selected 1000 dimensional embedding for each ROI. This algorithm learns a separate transformation matrix that projects each participant into this common space [57]. Estimating a lower dimensional common model space can potentially aid in filtering out measurement noise that is assumed to be independent across individuals. To minimize bias, and ensure ISC analyses were estimated independent of computing the SRM [55]; we learned transformation matrices using data from episode 2 and subsequently applied these transformations to data from episode 1. This allowed us to project each ROI's voxel responses into the common 1000 dimensional space.

Temporal Recurrence of Spatial Patterns

We created a spatial recurrence matrix by taking the pairwise Pearson correlation of the vectorized vmPFC activity for each TR collected during the 45-minute episode. This results in a number-of-TRs by number-of-TRs matrix (n=1364 TRs) reflecting the degree to which the spatial activity patterns in the vmPFC are correlated at each pair of TRs. Sequential TRs where the spatial configuration of the vmPFC persists create a block diagonal structure. Recurring activity patterns at distant timepoints are reflected in the off-diagonal similarity.

Autocorrelation

To compute the autocorrelation function at each voxel, we calculated the average correlation of the voxel's activation timeseries with itself shifted by a lag of t timesteps, for $t \in \{1, 2, \ldots, 50\}$.. This resulted in a correlation coefficient for each lag, for each voxel. We summarized the overall shape of this function within each ROI by taking the median across all voxels within the ROI and fitting a 3 parameter exponential function using a robust estimator. The parameter estimation was performed by minimizing the residual sum of squared error using the curve_fit() function from the scipy toolbox [104]. We used the median of the distribution of autocorrelation across participants to ensure robust parameter estimation.

$$\underset{\alpha,\beta,c}{\operatorname{argmin}} A - \alpha e^{-\beta x} + c \qquad \text{(eq 7)}$$

where A is the median autocorrelation function.

We used a similar procedure to calculate autocorrelation in spatial patterns. For each ROI, we computed a spatial recurrence matrix by correlating the activity patterns of each time t with the pattern at time $t + lag$, for $lag \in \{1, 2, \ldots, 50\}$, and then averaging the matrices created for each timepoint. We fit an exponential function to the diagonal of the across-participants average spatial recurrence matrix..

To compare autocorrelations across ROIs, we calculated the number of TRs the estimated exponential function took to reach an arbitrary correlation threshold of 0.1. This metric provides an intuitive and interpretable value for comparing the activity drift rates across regions and modalities (e.g. voxel versus spatial patterns).

Hidden Markov Model

To identify state changes reflected in the neural dynamics within the vmPFC, we fit Hidden Markov Models (HMMs) to the voxel timeseries [105,106]. HMMs are generative probabilistic models that propose that a sequence of observations is generated by a sequence of latent states. The transitions between these latent states are assumed to form a first-order Markov chain, such that the likelihood of transitioning to a new state from the current state can be predicted solely from the most recent data. We also assume that the observations are drawn from a multivariate Gaussian distribution with a diagonal covariance matrix. Prior to fitting an HMM to each participant's vmPFC voxel pattern timeseries (3591 voxels), we first applied PCA and retained enough components to explain 90% of the variance in the original signal. In Study 2, the vmPFC (mean=44.43, sd=14.18) had a higher number of components than V1 (mean=17.0, sd=4.61). We then fit the HMM using the GaussianHMM function from the Python hmmlearn package version 0.2.1, which estimates the model parameters using expectation-maximization. We used the fitted models to compute the MAP sequence of vmPFC states for each participant.

In implementing the above procedure, we fit independent HMMs to each participant's PCA-reduced vmPFC voxel timeseries. This approach allows both the spatial patterns and the sequences of state changes to vary across participants. We fit the HMMs using the number of states, *k*, that most frequently (across participants) resulted in the largest increase in individual model fitness.

As described in the main text, we aligned the latent states across participants by maximizing the spatial similarity between the estimated feature models using the Hungarian algorithm [63].

Study 3 (Face Expression)

Subjects

Twenty participants (mean [sd] age=18.9 [0.91] years; 13 female) were recruited from the Department of Psychological Brain sciences at Dartmouth College for course credit.

Procedure

Participants watched the first four episodes of the first season of Friday Night Lights over two separate two-hour sessions. Here we report results from episode one. Facial expressions during the experiment were monitored using GoPro HERO 4 cameras recording at 120 frames per second at 1920 x 1080 resolution. Each camera was positioned using a custom facecam headsets developed by our group [64]. This approach is invariant to head motion and minimizes many types of facial occlusions. Recorded videos were then temporally aligned to the episodes by minimizing differences in audio intensity using our open source Python FaceSync toolbox version 0.0.8 [64]. Facial behavior features consisting of 20 facial action units (AU), a standard for measuring facial muscle movement based on the Facial Action Coding System [FACS, 65] were extracted using the FACET algorithm [66] accessed through the iMotions biometric research platform [107]. Data were downsampled to 0.5 hz.

Dimensionality Reduction

We performed dimensionality reduction of the face expression timeseries data (Study 3; Fig. S5) and affective rating timeseries data (Study 4; Fig. 75) using Principal Components Analysis (PCA) implemented in scikit-learn [108]. Each feature for each participant was *z*-scored. Next we concatenated each participant's data into an observations by features matrix. Each component time series was correlated with the participant state concordance timeseries from Studies 1 and 2 (Tables S2 and S3). We report the results for the component with the highest temporal correlation across both studies that exceeded FDR correction (q < 0.05). Figure 6 depicts the feature loadings onto each component for face expressions (Figure 6C) and emotion ratings (Figure 6D).

Face Model Visualization

We used our Python Facial Expression Analysis Toolbox version 0.0.1 [109] to visualize how specific features from our model correspond to changes in facial morphometry data collected during Study 3. In brief, we learned a mapping between 20 facial action unit intensities and 68 landmarks comprising a 2-dimensional face using partial least squares implemented in scikit-learn [108]. We used 10,708 images corresponding to each frame from the extended cohn-kanade facial expression database [CK+, 110] and extracted landmarks using OpenFace [111] and action units from iMotions FACET engine [107]. Next, we performed an affine transformation to a canonical face prior to fitting the model, and added pitch, roll, and yaw head rotation parameters as covariates with the 20 action unit features. We then fit our model using a 3-fold cross-validation procedure. We achieved an overall training model fit of r^2 = .61 and a cross-validated mean of r^2 = .53. We used this facial morphometry model to visualize the degree to which each AU loaded on a specific principal component. Action Unit intensities are overlaid onto corresponding facets of the face landmark model (Fig. S6).

Study 4 (Emotion Ratings)

Subjects

We recruited 192 participants from the Amazon Mechanical Turk workplace to participate in a 60-minute study (mean [sd] age=37.11 [10.71] years; 105 female), during which they viewed the first episode of Friday Night Lights. All participants provided informed consent in accordance with the guidelines set by the Committee for the Protection of Human Subjects at Dartmouth College and were compensated for their participation ($9). Four subjects were excluded for not having enough data to perform collaborative filtering (less than 25% of samples).

Procedure

The experimental paradigm was delivered via a custom open source web application built using Flask (http://flask.pocoo.org/), jsPsych (https://www.jspsych.org/), psiTurk (https://psiturk.org/), and MySQL (https://www.mysql.com/) and served using an NGNIX (https://www.nginx.com/) web server hosted in our laboratory (https://github.com/cosanlab/moth_app, https://github.com/cosanlab/moth_turkframe). For each participant, the episode was paused over the course of its 45-minute runtime at random intervals of 200 - 280 seconds sampled from a uniform distribution (following an initial uninterrupted viewing period of 30 - 270 seconds). At each pause, participants were instructed to rate the intensity, if any, of 16 different dimensions: contempt, surprise, relief, anger, envy, shame, interest, elation, satisfaction, guilt, sadness, hope, pride, fear, joy, and disgust. The set of dimensions rated by each participant were presented in random order across participants. On average, participants completed 11.45 ratings (average sd=1.93 across emotion categories). The average intertrial interval for making a rating was 229.65 sec (average std = 53.33). Ratings were downsampled to 1 Hz and used to create a subject-by-time sparse matrix (188 x 2702). A collaborative filtering technique (non-negative matrix factorization with stochastic gradient descent) was applied to resulting matrices in order to infer each participant's rating for every moment in time [112].

Collaborative Filtering

We used collaborative filtering (CF) to create a timeseries for each affective dimension for every Study 4 participant [113]. CF is a machine-learning technique that is commonly used by industrial recommender systems (e.g., Netflix, Amazon) to infer user preferences on many items (e.g., movies, music, products) when their preferences are only known for a few items [114]. The key assumption driving CF is that, to the extent that participants *A* and *B* have similar ratings on *n* samples, they will rate other items similarly. To extend this general approach to the timeseries affective rating data we collected in Study 4, we convolved each participant's (sparse) affective ratings (i.e., on a subset of the dimensions we sought to estimate) with a 60-second boxcar function kernel. We then used non-negative matrix factorization with stochastic gradient descent (with 100 iterations) implemented in our Python emotionCF toolbox (https://github.com/ljchang/emotionCF) to perform the matrix completion [112]. This technique allows us to reconstruct a time series of each participant's emotional ratings at every second during the show, while attempting to minimize the overall impact of the rating procedure on participants' emotional experiences. This approach makes no assumptions about the underlying distribution of ratings across participants and allows us to maintain individual differences in rating timecourses.

Software

Unless otherwise noted, all of our imaging analyses were performed with our open source Python nltools toolbox version 0.3.10 (https://github.com/ljchang/nltools). This package wraps many scientific Python tools including: Nibabel 2.1 [115] , nilearn 0.4 [116], numpy [117], pandas [118], scipy [104], and scikit-learn [108] . Our plots were generated using matplotlib [119], seaborn [120], and custom plotting functions https://github.com/cosanlab/nltools. All code used to perform the analyses in this paper will be available on our lab github page pending publication of this manuscript (https://github.com/cosanlab/vmPFC_dynamics).

Acknowledgments

The authors would like to acknowledge Sushmita Sadhuka, Zainab Molani, and Antonia Hoidal who assisted in the data collection. This work was supported by funding from the National Institute of Mental Health R01MH116026 and R56MH080716.

Author Contributions

All authors were involved in conceptualizing this project. EJ & JC collected data for Study 1 & 2, JC collected data for Study 3, and NG designed and collected data for Study 4. LC & EJ analyzed the data. All authors wrote the manuscript.

Competing Interests

The authors declare no competing interests.

References

1. Goodale, M. A. & Milner, A. D. Separate visual pathways for perception and action. *Trends Neurosci.* **15,** 20–25 (1992).
2. Mishkin, M., Ungerleider, L. G. & Macko, K. A. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* **6,** 414–417 (1983).
3. Ashar, Y. K., Chang, L. J. & Wager, T. D. Brain Mechanisms of the Placebo Effect: An Affective Appraisal Account. *Annu. Rev. Clin. Psychol.* **13,** 73–98 (2017).
4. Roy, M., Shohamy, D. & Wager, T. D. Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends Cogn. Sci.* **16,** 147–156 (2012).
5. Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 12574–12579 (2016).
6. Huntenburg, J. M., Bazin, P.-L. & Margulies, D. S. Large-scale gradients in human cortical organization. *Trends Cogn. Sci.* **22,** 21–31 (2018).
7. Raichle, M. E. *et al.* A default mode of brain function. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 676–682 (2001).
8. Greicius, M. D., Krasnow, B., Reiss, A. L. & Menon, V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 253–258 (2003).
9. Shulman, G. L. *et al.* Common blood flow changes across visual tasks: II. decreases in cerebral cortex. *J. Cogn. Neurosci.* **9,** 648–663 (1997).
10. Gusnard, D. A., Akbudak, E., Shulman, G. L. & Raichle, M. E. Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 4259–4264 (2001).
11. Fox, M. D. *et al.* The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 9673–9678 (2005).
12. Andrews-Hanna, J. R., Smallwood, J. & Spreng, R. N. The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Ann. N. Y. Acad. Sci.* **1316,** 29–52 (2014).
13. Mason, M. F. *et al.* Wandering minds: the default network and stimulus-independent thought. *Science* **315,** 393–395 (2007).
14. Christoff, K., Gordon, A. M., Smallwood, J., Smith, R. & Schooler, J. W. Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 8719–8724 (2009).
15. Kelley, W. M. *et al.* Finding the self? An event-related fMRI study. *J. Cogn. Neurosci.* **14,** 785–794 (2002).
16. Jenkins, A. C., Macrae, C. N. & Mitchell, J. P. Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 4507–4512 (2008).
17. Schacter, D. L., Addis, D. R. & Buckner, R. L. Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.* **8,** 657–661 (2007).
18. Buckner, R. L. & Carroll, D. C. Self-projection and the brain. *Trends Cogn. Sci.* **11,** 49–57 (2007).
19. Cabeza, R. & St Jacques, P. Functional neuroimaging of autobiographical memory. *Trends Cogn. Sci.* **11,** 219–227 (2007).
20. Spreng, R. N., Mar, R. A. & Kim, A. S. N. The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative

meta-analysis. *J. Cogn. Neurosci.* **21,** 489–510 (2009).

21. Rissman, J., Greely, H. T. & Wagner, A. D. Detecting individual memories through the neural decoding of memory states and past experience. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 9849–9854 (2010).
22. Lerner, Y., Honey, C. J., Silbert, L. J. & Hasson, U. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* **31,** 2906–2915 (2011).
23. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject synchronization of cortical activity during natural vision. *Science* **303,** 1634–1640 (2004).
24. Chen, J. *et al.* Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* **20,** 115–125 (2017).
25. Bhandari, A., Gagne, C. & Badre, D. Just above Chance: Is It Harder to Decode Information from Human Prefrontal Cortex Blood Oxygenation Level-dependent Signals? *J. Cogn. Neurosci.* 1–26 (2018).
26. Mueller, S., Wang, D., Fox, M. D., Yeo, B. & Sepulcre, J. Individual variability in functional connectivity architecture of the human brain. *Neuron* (2013).
27. Gordon, E. M. *et al.* Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* **26,** 288–303 (2016).
28. Haber, S. N., Kunishio, K., Mizobuchi, M. & Lynd-Balta, E. The orbital and medial prefrontal circuit through the primate basal ganglia. *J. Neurosci.* **15,** 4851–4867 (1995).
29. Kahnt, T., Chang, L. J., Park, S. Q., Heinzle, J. & Haynes, J.-D. Connectivity-based parcellation of the human orbitofrontal cortex. *J. Neurosci.* **32,** 6240–6250 (2012).
30. Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R. & Buckner, R. L. Functional-anatomic fractionation of the brain's default network. *Neuron* **65,** 550–562 (2010).
31. Damasio, A. R. *Descartes' error*. (Random House, 2006).
32. Damasio, A. R. *et al.* Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat. Neurosci.* **3,** 1049–1056 (2000).
33. Chikazoe, J., Lee, D. H., Kriegeskorte, N. & Anderson, A. K. Population coding of affect across stimuli, modalities and individuals. *Nat. Neurosci.* **17,** 1114–1122 (2014).
34. Fox, K. C. R. *et al.* Changes in subjective experience elicited by direct stimulation of the human orbitofrontal cortex. *Neurology* **91,** 1–9 (2018).
35. Rangel, A., Camerer, C. & Montague, P. R. A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* **9,** 545–556 (2008).
36. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441,** 223–226 (2006).
37. Plassmann, H., O'Doherty, J., Shiv, B. & Rangel, A. Marketing actions can modulate neural representations of experienced pleasantness. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 1050–1054 (2008).
38. Kahnt, T., Heinzle, J., Park, S. Q. & Haynes, J.-D. The neural code of reward anticipation in human orbitofrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 6010–6015 (2010).
39. O'Doherty, J. P., Kringelbach, M. L., Rolls, E. T., Hornak, J. & Andrews, C. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat. Neurosci.* **4,** 95–102 (2001).
40. Ferstl, E. C., Neumann, J., Bogler, C. & Von Cramon, D. Y. The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Hum. Brain Mapp.* **29,** 581–593 (2008).
41. Simony, E. *et al.* Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.* **7,** 12141 (2016).

42. Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D. & Hasson, U. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E4687–96 (2014).
43. Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532,** 453–458 (2016).
44. Baldassano, C., Hasson, U. & Norman, K. A. Representation of real-world event schemas during narrative perception. *J. Neurosci.* **38,** 9689–9699 (2018).
45. Yeshurun, Y. *et al.* Same Story, Different Story. *Psychol. Sci.* **28,** 307–319 (2017).
46. Weiskopf, N., Hutton, C., Josephs, O., Turner, R. & Deichmann, R. Optimized EPI for fMRI studies of the orbitofrontal cortex: compensation of susceptibility-induced gradients in the readout direction. *Magnetic Resonance Materials in Physics, Biology and Medicine* **20,** 39–49 (2007).
47. Du, Y. P., Dalwani, M., Wylie, K., Claus, E. & Tregellas, J. R. Reducing susceptibility artifacts in fMRI using volume-selective z-shim compensation. *Magn. Reson. Med.* **57,** 396–404 (2007).
48. Glover, G. H. & Law, C. S. Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Magn. Reson. Med.* **46,** 515–522 (2001).
49. Hasson, U., Chen, J. & Honey, C. J. Hierarchical process memory: memory as an integral component of information processing. *Trends Cogn. Sci.* **19,** 304–313 (2015).
50. Honey, C. J. *et al.* Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* **76,** 423–434 (2012).
51. O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H. & Dolan, R. J. Temporal difference models and reward-related learning in the human brain. *Neuron* **38,** 329–337 (2003).
52. Kable, J. W. & Glimcher, P. W. The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* **10,** 1625–1633 (2007).
53. McClure, S. M. *et al.* Neural correlates of behavioral preference for culturally familiar drinks. *Neuron* **44,** 379–387 (2004).
54. O'Doherty, J. P., Hampton, A. & Kim, H. Model-based fMRI and Its application to reward learning and decision making. *Ann. N. Y. Acad. Sci.* **1104,** 35–53 (2007).
55. Haxby, J. V. *et al.* A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72,** 404–416 (2011).
56. Guntupalli, J. S. *et al.* A model of representational spaces in human cortex. *Cereb. Cortex* **26,** 2919–2934 (2016).
57. Chen, P.-H. *et al.* A Reduced-Dimension fMRI Shared Response Model. in *Advances in Neural Information Processing Systems 28* (eds. Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R.) 460–468 (Curran Associates, Inc., 2015).
58. Ma, F., Nastase, S. A., Guntupalli, J. S. & Haxby, J. V. Reliable individual differences in fine-grained cortical functional architecture. *Neuroimage* **183,** 375–386 (2018).
59. Yarkoni, T., de la Vega, A. & Chang, L. J. Fully automated meta-analytic clustering and decoding of human brain activity.
60. de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D. & Yarkoni, T. Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. *J. Neurosci.* **36,** 6553–6562 (2016).
61. Choe, A. S. *et al.* Comparing test-retest reliability of dynamic functional connectivity methods. *Neuroimage* **158,** 155–175 (2017).
62. Glerean, E., Salmi, J., Lahnakoski, J. M., Jääskeläinen, I. P. & Sams, M. Functional magnetic resonance imaging phase synchronization as a measure of dynamic functional

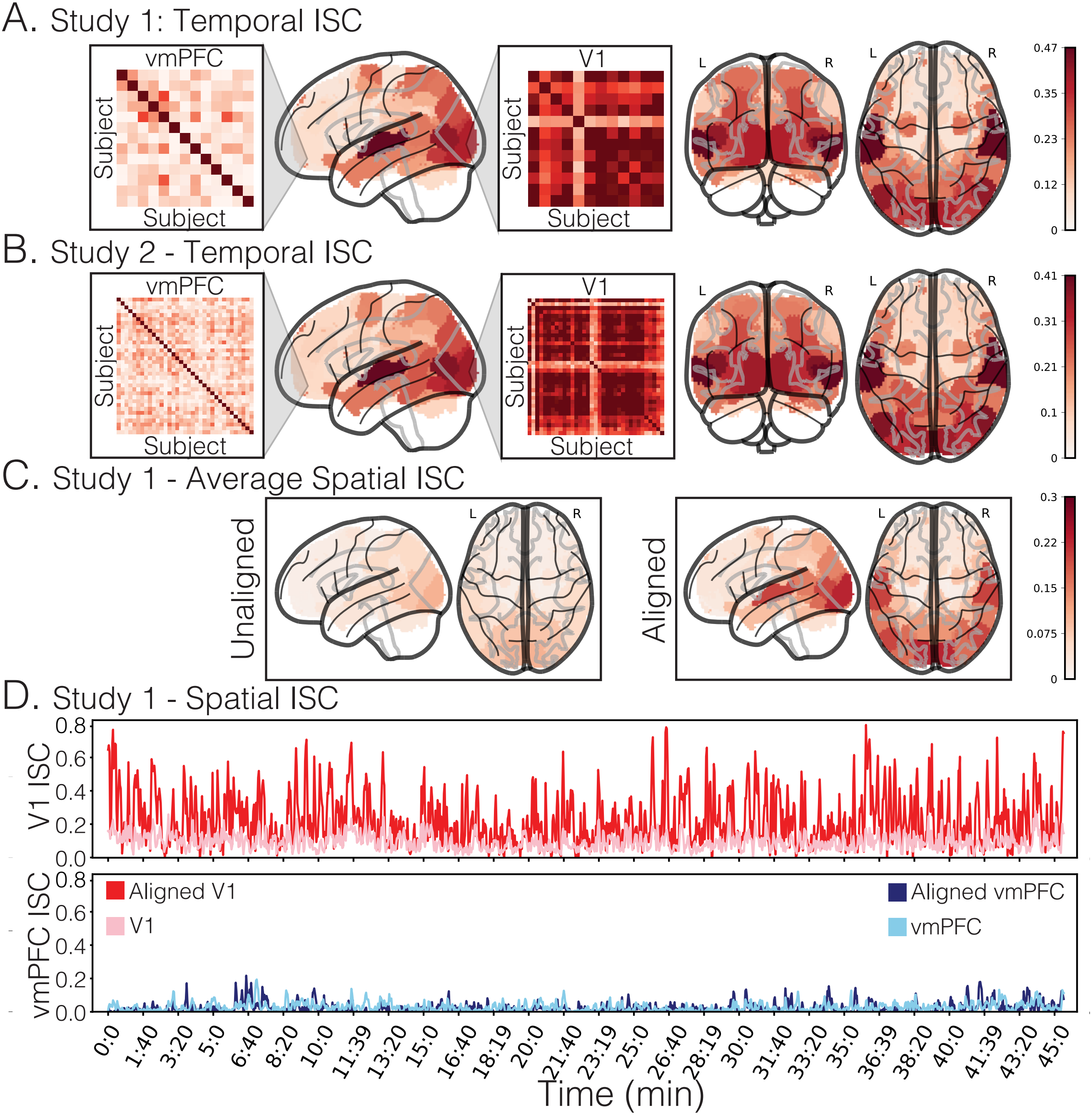
connectivity. *Brain Connect.* **2,** 91–101 (2012).

63. Kuhn, H. W. The Hungarian method for the assignment problem. *Naval research logistics quarterly* (1955).
64. Cheong, J. H., Brooks, S. & Chang, L. J. FaceSync: Open source framework for recording facial expressions with head-mounted cameras. (2017).
65. Ekman, P. & Friesen, W. V. Measuring facial movement. *J. Nonverbal Behav.* **1,** 56–75 (1976).
66. Littlewort, G. *et al.* The computer expression recognition toolbox (CERT). in *Face and Gesture 2011* 298–305 (2011).
67. Cohn, J. F., Ambadar, Z. & Ekman, P. Observer-based measurement of facial expression with the Facial Action Coding System. *The handbook of emotion elicitation and assessment* 203–221 (2007).
68. Kang, O. & Wheatley, T. Pupil dilation patterns spontaneously synchronize across individuals during shared attention. *J. Exp. Psychol. Gen.* **146,** 569–576 (2017).
69. Doeller, C. F., Barry, C. & Burgess, N. Evidence for grid cells in a human memory network. *Nature* **463,** 657–661 (2010).
70. Constantinescu, A. O., O'Reilly, J. X. & Behrens, T. E. J. Organizing conceptual knowledge in humans with a gridlike code. *Science* **352,** 1464–1468 (2016).
71. Schuck, N. W., Cai, M. B., Wilson, R. C. & Niv, Y. Human orbitofrontal cortex represents a cognitive map of state space. *Neuron* **91,** 1402–1412 (2016).
72. Stephens, G. J. & Silbert, L. J. Speaker–listener neural coupling underlies successful communication. *Proceedings of the* **107,** 14425–14430 (2010).
73. Suzuki, S., Cross, L. & O'Doherty, J. P. Elucidating the underlying components of food valuation in the human orbitofrontal cortex. *Nat. Neurosci.* **20,** 1780 (2017).
74. Reber, J. *et al.* Selective impairment of goal-directed decision-making following lesions to the human ventromedial prefrontal cortex. *Brain* **140,** 1743–1756 (2017).
75. Robinson, M. J. F. & Berridge, K. C. Instant Transformation of Learned Repulsion into Motivational 'Wanting'. *Curr. Biol.* **23,** 282–289 (2013).
76. Skerry, A. E. & Saxe, R. Neural representations of emotion are organized around abstract event features. *Curr. Biol.* **25,** 1945–1954 (2015).
77. Saarimäki, H. *et al.* Distributed affective space represents multiple emotion categories across the human brain. *Soc. Cogn. Affect. Neurosci.* **13,** 471–482 (2018).
78. Anderson, A. K. *et al.* Dissociated neural representations of intensity and valence in human olfaction. *Nat. Neurosci.* **6,** 196–202 (2003).
79. Gottfried, J. A., Deichmann, R., Winston, J. S. & Dolan, R. J. Functional heterogeneity in human olfactory cortex: an event-related functional magnetic resonance imaging study. *J. Neurosci.* **22,** 10819–10828 (2002).
80. Baldassano, C. *et al.* Discovering event structure in continuous narrative perception and memory. *Neuron* **95,** 709–721.e5 (2017).
81. Lositsky, O. *et al.* Neural pattern change during encoding of a narrative predicts retrospective duration estimates. *Elife* **5,** e16070 (2016).
82. Heusser, A. C., Fitzpatrick, P. C. & Manning, J. R. How is experience transformed into memory? *bioRxiv* (2018). doi:10.1101/409987
83. Friston, K. J., Holmes, A. P. & Worsley, K. J. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* (1994).
84. Fox, M. D. & Raichle, M. E. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* **8,** 700–711 (2007).
85. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping.

Proc. Natl. Acad. Sci. U. S. A. **103,** 3863–3868 (2006).

86. Haynes, J.-D. & Rees, G. Neuroimaging: decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* **7,** 523 (2006).
87. Mukamel, R. *et al.* Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science* **309,** 951–954 (2005).
88. Heller, A. S., Lapate, R. C., Mayer, K. E. & Davidson, R. J. The face of negative affect: trial-by-trial corrugator responses to negative pictures are positively associated with amygdala and negatively associated with ventromedial prefrontal cortex activity. *J. Cogn. Neurosci.* **26,** 2102–2110 (2014).
89. Chang, L. J., Yarkoni, T., Khaw, M. W. & Sanfey, A. G. Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. *Cereb. Cortex* **23,** 739–749 (2013).
90. Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A. & Wager, T. D. A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. *PLoS Biol.* **13,** e1002180 (2015).
91. Krishnan, A. *et al.* Somatic and vicarious pain are represented by dissociable multivariate brain patterns. *Elife* **5,** e15166 (2016).
92. Woo, C.-W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging. *Nat. Neurosci.* **20,** 365–377 (2017).
93. Eisenbarth, H., Chang, L. J. & Wager, T. D. Multivariate Brain Prediction of Heart Rate and Skin Conductance Responses to Social Threat. *J. Neurosci.* **36,** 11987–11998 (2016).
94. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data* **3,** 160044 (2016).
95. Visconti di Oleggio Castello, M. *et al. ReproNim/reproin: 0.1.1*. doi:10.5281/zenodo.1207118
96. Gorgolewski, K. J. *et al.* Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinform.* **5,** 13 (2011).
97. Millman, K. J. & Brett, M. Analysis of functional magnetic resonance imaging in python. *Comput. Sci. Eng.* **9,** 52–55 (2007).
98. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17,** 825–841 (2002).
99. Avants, B. B., Tustison, N. & Song, G. Advanced normalization tools (ANTS). *Insight J.* **2,** 1–35 (2009).
100. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage* **62,** 782–790 (2012).
101. Whitfield-Gabrieli, S. Artifact detection and QA manual. *Massachusetts: Massachusetts Institute of Technology* (2009).
102. Székely, G. J. & Rizzo, M. L. The Energy of Data. *Annu. Rev. Stat. Appl.* **4,** 447–479 (2017).
103. Székely, G. J., Rizzo, M. L. & Bakirov, N. K. Measuring and testing dependence by correlation of distances. *Ann. Stat.* **35,** 2769–2794 (2007).
104. Jones, E., Oliphant, T. & Peterson, P. {SciPy}: Open source scientific tools for {Python}. (2001--).
105. Rabiner, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77,** 257–286 (1989).
106. Bilmes, J. A. & Others. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer*

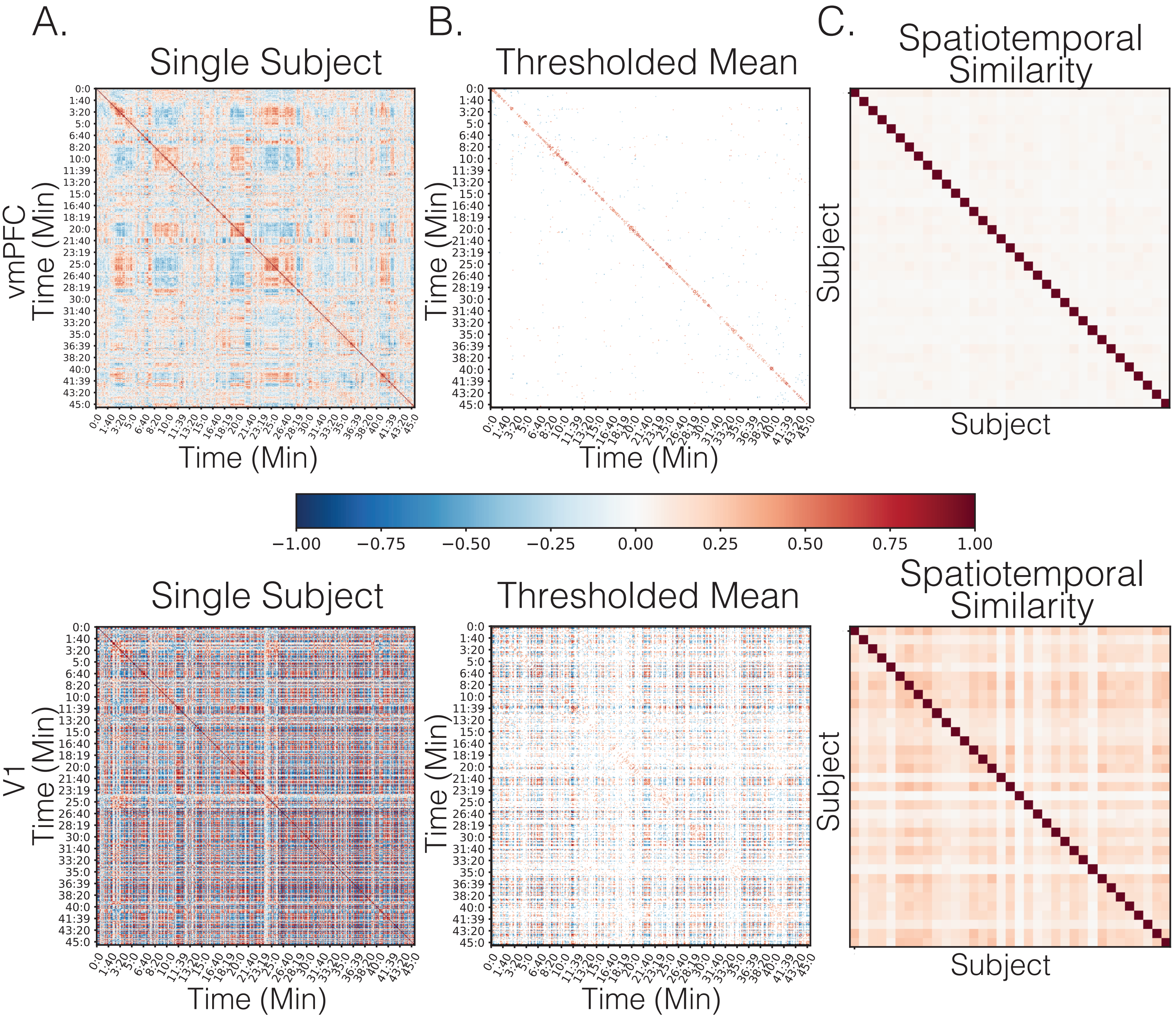
Science Institute **4,** 126 (1998).
107. *iMotions Biometric Research Platform 6.0*. (iMotions A/S, Copenhagen, Denmark, 2016).
108. Pedregosa, F. *et al.* Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12,** 2825–2830 (2011).
109. Cheong, J. H., Byrnes, S. & Chang, L. J. Facial expression analysis toolbox.
110. Lucey, P. *et al.* The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops* 94–101 (ieeexplore.ieee.org, 2010).
111. Baltrušaitis, T., Robinson, P. & Morency, L. P. OpenFace: An open source facial behavior analysis toolkit. in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* 1–10 (ieeexplore.ieee.org, 2016).
112. Chang, L. J., Eisenbarth, H., Reddan, M., Andrews, E. & Wager, T. D. Recovering individual sparse emotion ratings using collaborative filtering techniques.
113. Linden, G., Smith, B. & York, J. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Comput.* **7,** 76–80 (2003).
114. Su, X. & Khoshgoftaar, T. M. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence* **2009,** 2 (2009).
115. Brett, M. *et al.* nibabel: 2.1. 0. *Zenodo* (2016).
116. Abraham, A. *et al.* Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* **8,** 14 (2014).
117. Oliphant, T. E. *A guide to NumPy*. **1,** (Trelgol Publishing USA, 2006).
118. McKinney, W. & Others. Data structures for statistical computing in python. in *Proceedings of the 9th Python in Science Conference* **445,** 51–56 (Austin, TX, 2010).
119. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering* **9,** 90–95 (2007).
120. Waskom, M., Botvinnik, O., Hobson, P. & Warmenhoven, J. Seaborn: statistical data visualization. (2014).


A. Study 1: Temporal ISC

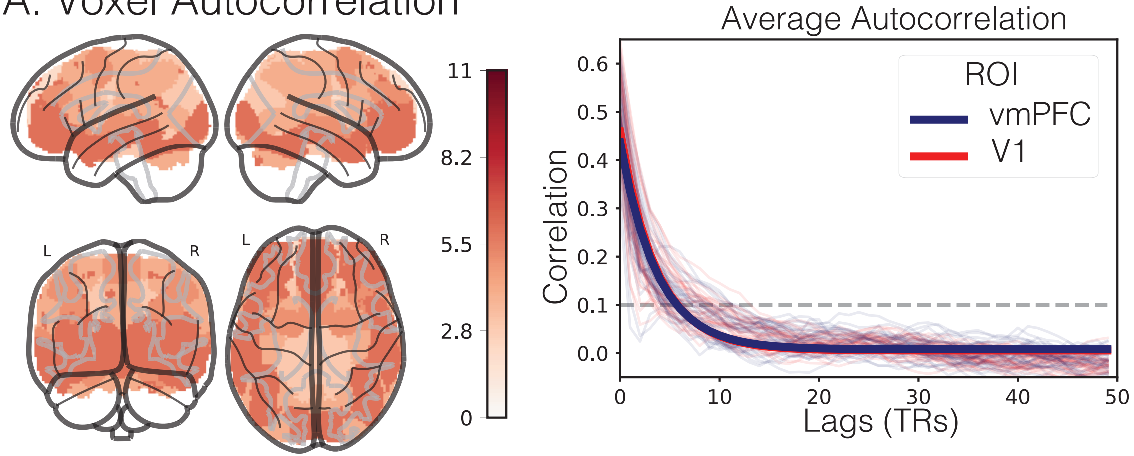
B. Study 2 - Temporal ISC

C. Study 1 - Average Spatial ISC

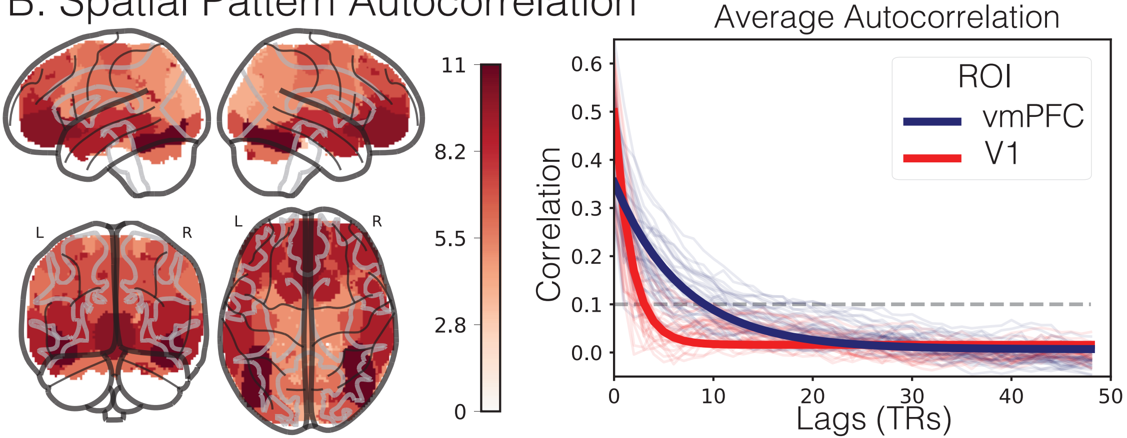
D. Study 1 - Spatial ISC


A.
B.
C.
Single Subject
Thresholded Mean
Spatiotemporal Similarity
vmPFC
V1
Time (Min)
Subject
0:0
1:40
3:20
5:0
6:40
8:20
10:0
11:39
13:20
15:0
16:40
18:19
20:0
21:40
23:19
25:0
26:40
28:19
30:0
31:40
33:20
35:0
36:39
38:20
40:0
41:39
43:20
45:0
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

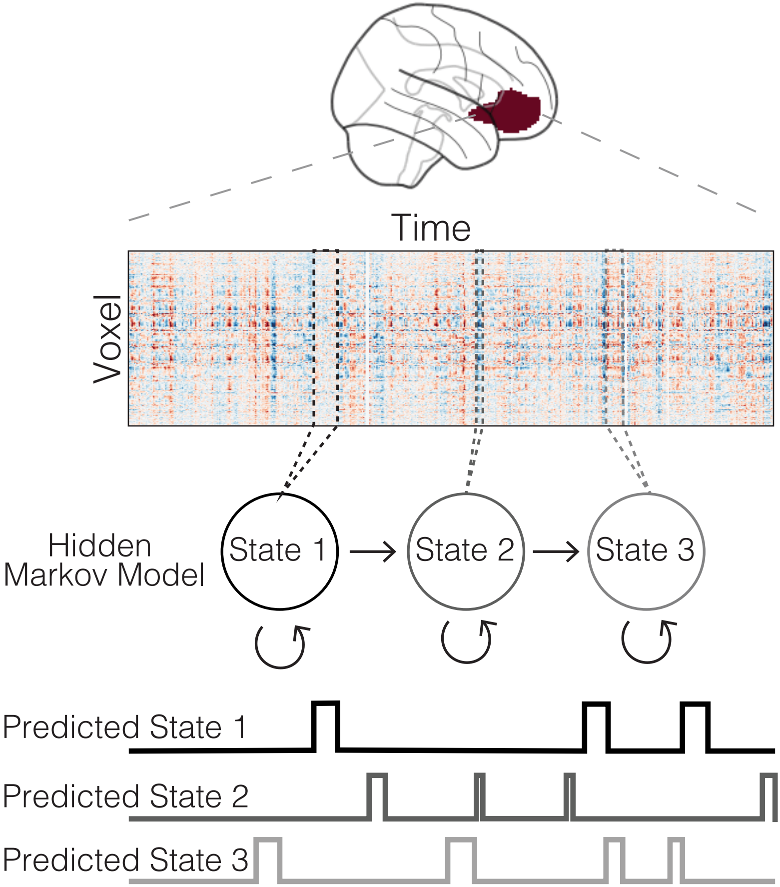
A. Voxel Autocorrelation



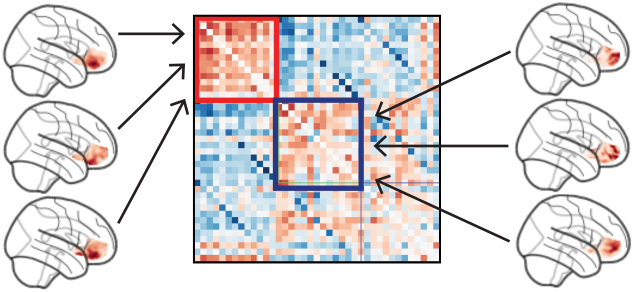
B. Spatial Pattern Autocorrelation

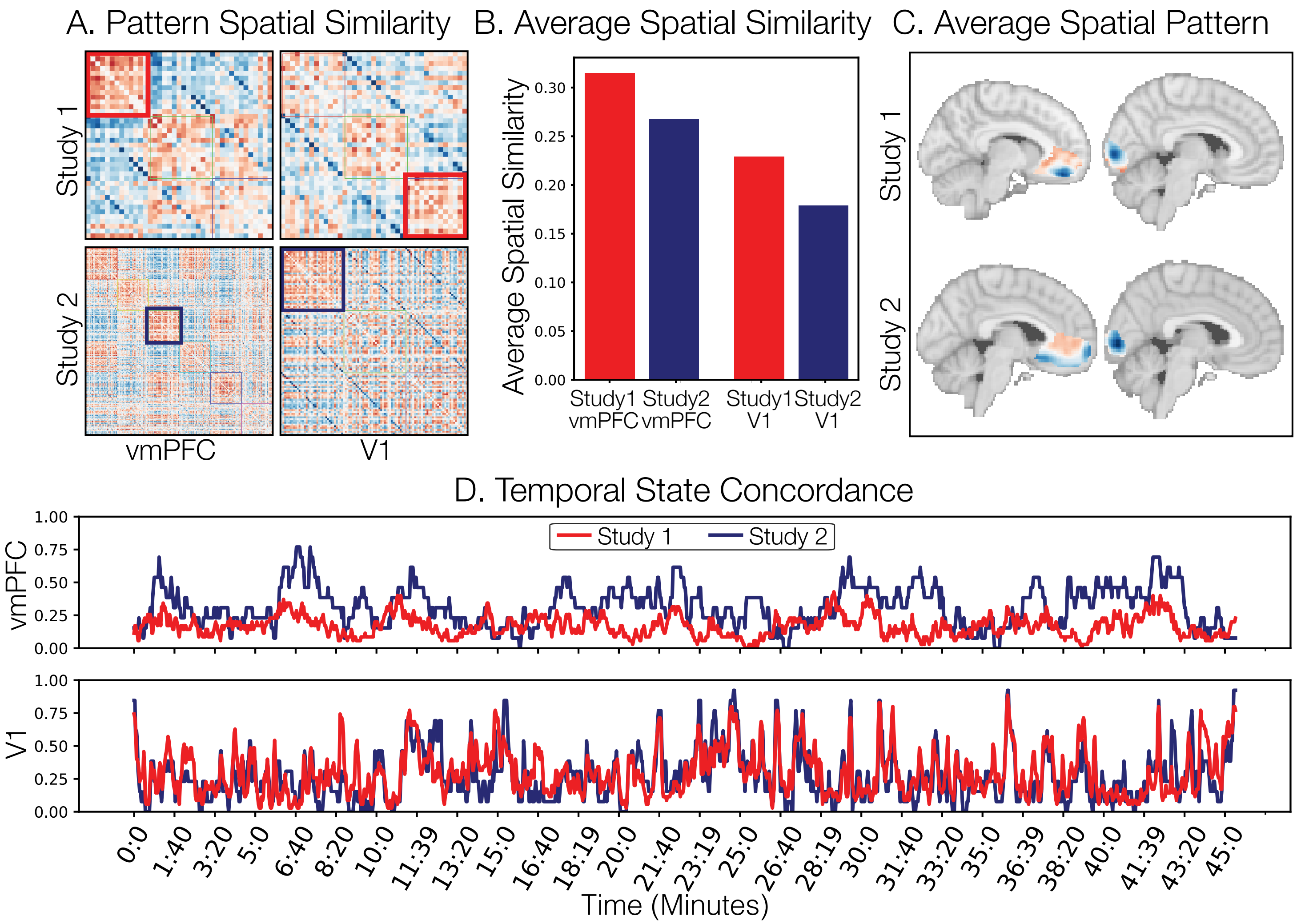


A. Fit Hidden Markov Model to vmPFC

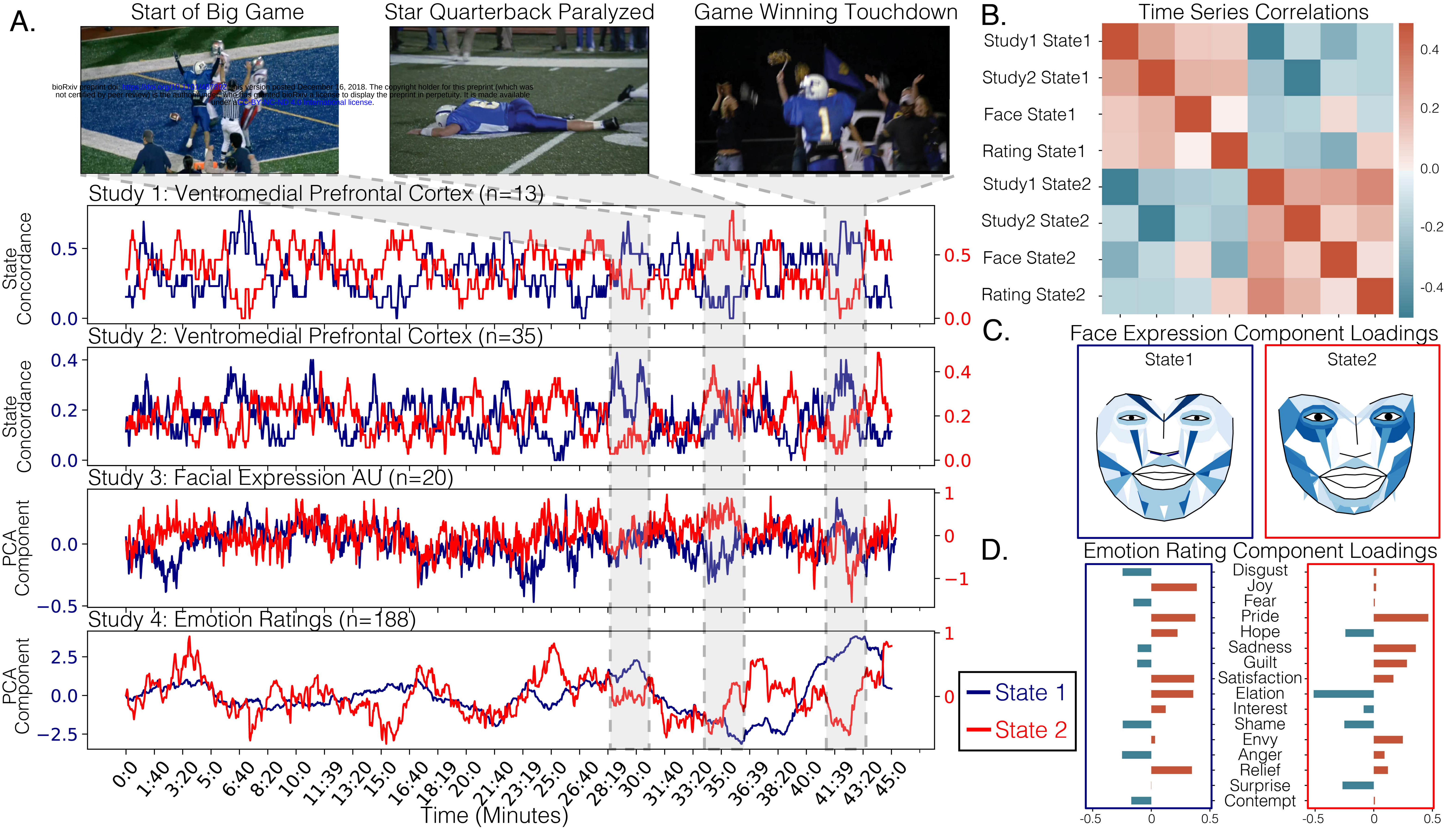


B. Spatial Alignment





A. Pattern Spatial Similarity
Study 1
Study 2
vmPFC
V1
B. Average Spatial Similarity
Average Spatial Similarity
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Study1 vmPFC
Study2 vmPFC
Study1 V1
Study2 V1
C. Average Spatial Pattern
Study 1
Study 2
D. Temporal State Concordance
vmPFC
V1
Study 1
Study 2
1.00
0.75
0.50
0.25
0.00
0:0
1:40
3:20
5:0
6:40
8:20
10:0
11:39
13:20
15:0
16:40
18:19
20:0
21:40
23:19
25:0
26:40
28:19
30:0
31:40
33:20
35:0
36:39
38:20
40:0
41:39
43:20
45:0
Time (Minutes)


A.
Start of Big Game
Star Quarterback Paralyzed
Game Winning Touchdown
Study 1: Ventromedial Prefrontal Cortex (n=13)
State Concordance
0.5
0.0
Study 2: Ventromedial Prefrontal Cortex (n=35)
0.4
0.2
0.0
Study 3: Facial Expression AU (n=20)
PCA Component
0.0
−0.5
1
0
−1
Study 4: Emotion Ratings (n=188)
2.5
0.0
−2.5
0:0 1:40 3:20 5:0 6:40 8:20 10:0 11:39 13:20 15:0 16:40 18:19 20:0 21:40 23:19 25:0 26:40 28:19 30:0 31:40 33:20 35:0 36:39 38:20 40:0 41:39 43:20 45:0
Time (Minutes)
State 1
State 2
B.
Time Series Correlations
Study1 State1
Study2 State1
Face State1
Rating State1
Study1 State2
Study2 State2
Face State2
Rating State2
0.4
0.2
0.0
-0.2
-0.4
C.
Face Expression Component Loadings
State1
State2
D.
Emotion Rating Component Loadings
Disgust
Joy
Fear
Pride
Hope
Sadness
Guilt
Satisfaction
Elation
Interest
Shame
Envy
Anger
Relief
Surprise
Contempt
-0.5
0
0.5