Generalizable Neural Models of Emotional Engagement and Disengagement

Melanni Nanni-Zepeda[1,2], Travis C. Evans[13,15], Audreyana Jagger-Rickels[12,14,15], Gal Raz[3,4,5], Talma Hendler[4,5,6,7], Yan Fan[8], Simone Grimm[9,10,11], Martin Walter[1,2,16,17], Michael Esterman[12, 15], and Agnieszka Zuberer[1,2]

[1]Department of Psychiatry and Psychotherapy, Jena University Hospital, Jena, Germany

[2]Attention and Affect Laboratory, Department of Psychiatry and Psychotherapy, University of Tübingen, Tübingen, Germany

[3]Steve Tisch School of Film and Television, Tel Aviv University, Tel Aviv, Israel

[4]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

[5]Sagol Brain Institute, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel

[6]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

[7]The School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel

[8]Department Psychology and Neurosciences, Leibniz Research Centre for Working Environment and Human Factors at the TU Dortmund (IfADo), Dortmund, Germany

[9]Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric Hospital, University of Zurich, Zurich, Switzerland

[10]Department of Psychology, MSB Medical School Berlin, Berlin, Germany

[11]Berlin Institute of Health, Campus Benjamin Franklin, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany

[12]National Center for PTSD, VA Boston Healthcare System, Boston, MA, USA

[13]Boston Attention and Learning Laboratory, VA Boston Healthcare System, Boston, MA, 02130, USA

[14]Department of Psychiatry, Harvard Medical School, Boston, MA, USA

[15]Department of Psychiatry, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, USA

[16]Center for Intervention and Research on adaptive and maladaptive brain Circuits underlying mental health (C-I-R-C), Halle-Jena-Magdeburg, Germany

[17]German Center for Mental Health (DZPG), partner site Halle-Jena-Magdeburg, Germany

Author Note

Correspondence concerning this article should be addressed to Melanni Nanni-Zepeda, Department of Psychiatry and Psychotherapy, Attention and Affect Laboratory, Eberhard Karls University, Tübingen, Germany. E-mail: nanni.melanni@gmail.com and azuberer@gmail.com

Abstract

Emotional experiences are never static but continuously evolve in response to internal and external contexts. Little is known about how neural patterns change as a function of these experiences, particularly in response to complex, real-world stimuli. This study aimed to identify generalizable neural patterns as individuals collectively engage and disengage from emotions dynamically. To do so, we analyzed functional magnetic resonance imaging (fMRI) along with subjective emotional annotations from two independent studies as individuals watched negative and neutral movie clips. We used predictive modeling to test if a model trained to predict a group emotional signature response in one study generalizes to the other study and vice versa. Disengagement patterns generalized specifically across intense clips. They were supported by connections within and between the sensorimotor and salience networks, maybe reflecting the processing of feeling states as individuals regulate their emotions. Prediction success for the engagement signature was mixed, primarily linked to connections within the visual and between the visual and dorsal attention networks, supporting visual attention orienting as emotions intensify. This work offers potential pathways for identifying generalizable neural patterns contributing to future affective research and clinical applications aiming to better understand dynamic emotional responses to naturalistic stimuli.

Keywords: movie, film, emotional intensity, engagement, disengagement, fMRI, arousal, subjective

Generalizable Neural Models of Emotional Engagement and Disengagement

Introduction

Our well-being profoundly depends on how we engage *in* and disengage *from* emotions when interacting with others and the world. While theories of affect emphasize the temporal progression of emotional responses and their regulation (e.g. Gross et al. (2014); Scherer (2009); Sheppes and Gross (2011)pp. 16-17), these processes have typically been studied in isolation (Goldin et al., 2009; Gruber et al., 2011; McRae et al., 2010), disregarding their dynamic nature throughout time. In real life, emotions evolve, and naturalistic scenarios like movies may preserve this temporal alternation of reactivity and recovery, offering rich, dynamic affective experiences (Morgenroth et al., 2023; Saarimäki, 2021). Although movies have the potential to mimic real-life scenarios, there is a challenge for affective movie fMRI in terms of ecological validity not necessarily directly translating to ecological generalizability, in that affective correlates may not extend uniformly across various naturalistic contexts (Nastase et al., 2020), and instead may be partially driven by contextual factors (jump cuts, actor's costumes, camera perspective, background (Hasson et al., 2008; Zacks & Magliano, 2011). Predictive modeling of a dynamic affective experience across various movies and participant groups holds the potential to unveil more generalized neural mechanisms relevant to the actual emotional experience, independent from the idiosyncratic nature of each movie. This way, common patterns, and neural responses can be distilled that transcend specific cinematic content and individual differences.

Few studies have shown that subjective reports of attention and emotion state can be predicted across movies and participant groups. One study showed that continuous subjective experience of fear during movie viewing was predicted by widespread large-scale brain network connectivity patterns, and that these predictions were successful between independent subject groups and movies (Zhou et al., 2023). In one study it was found that functional connectivity strength within the DMN was predictive of continuous attentional

engagement across a movie and audiobook from independent studies (Song, Finn, et al., 2021). Other work linked DMN connectivity to both intensity and polarity of continuous affective experiences across independent movie data sets (Lettieri et al., 2022). In contrast, another study found that even though stronger dynamic DMN recruitment was linked to moment-to-moment movie comprehension in separate movies, a DMN signature obtained in one movie failed to predict movie comprehension in the other movie and vice versa (Song, Park, et al., 2021). Together, these findings suggest that both dynamic changes in attentional engagement levels and affective processing, or an interplay of both may be tracked by dynamic functional brain connectivity (dynFC) across large-scale networks. Rather than following a continuum of their temporal changes of intensifying and waning of attention and affect may reflect disparate processing phases where individuals engage in and disengage from an emotional experience as they emotionally respond to a movie (Koval et al., 2015; Kuppens & Verduyn, 2015). Specifically, engagement phases may reflect intensified emotional response paired with heightened attention to direct focus on the narrative (Visch et al., 2010). In contrast, disengagement phases may capture shifts towards internally directed attention and temporary emotional detachment as viewers may or may not regulate their emotional state. Thus, a refined understanding of these transitions from low to high and high to low emotional intensity during movie viewing may be a way to study brain signatures supporting common patterns of engagement and disengagement patterns across individuals. To tackle this, we analyzed data from two independent studies, both using retroactive subjective emotional intensity annotations in response to either one intense negative and one neutral movie clip (study 1) or one intense negative movie clip (study 2) together with fMRI. Positive and negative shifts in subjects' annotation timecourses were labeled as EE and ED, respectively. We used support vector regression (SVR) to test the predictive nature of dynFC patterns across independent data linked to moment-to-moment changes in the degree to which a participant group exhibited EE and ED phases throughout movie viewing. Based on previous work linking subjective

attention states during movie viewing to the DMN, we wondered if DMN connectivity, both within and between networks, might play a central role in predicting group signature of EE or ED. However, given the limited predictive studies in naturalistic scenarios, we did not constrain our analysis to specific regions or networks of interest. Our analysis showed that a model trained on one negative movie clip successfully predicted an ED group signature in different participant groups across other negative clips, but not in neutral/low intensity clips, suggesting specificity to high intensity/negative content. This model used dynFC features across wide-spread large scale networks possibly supporting cognitive regulation of emotional responses. Conversely, a model predicting an EE group signature in one clip did not consistently succeed in forecasting EE signatures across emotional contexts, but predominantly involved dynFC features within the visual network and between the visual, dorsal attention network (DAN), and default mode network (DMN) potentially indicating heightened attentional and emotional involvement as emotions intensify. This study marks initial attempts to differentiate neural mechanisms behind the ebb and flow of emotional experiences during movie viewing.

Material and Methods

Data Sources

We analyzed fMRI and behavioral data from two independent studies with similar paradigms and acquisition methods. These prior studies addressed unrelated research questions. For clarity, we refer to them as study 1 (Borchardt et al., 2018) and study 2 (Raz et al., 2016). We previously published work based on study 1 (Nanni-Zepeda et al., 2024) with another research question.

Participants

Study 1 included 22 female volunteers aged 20–49 years (mean age 28.1 ∓ 6.5), screened for neurological/psychiatric disorders using the short SCID (Wittchen et al., 1997). Study 2

comprised 44 participants (25 female, 19 male) aged 21–37 years (mean age 26.73 ∓ 4.69), with no known history of neurological/psychiatric disorders.

Naturalistic Viewing Paradigm

In study 1, participants watched an excerpt from *21 Grams* (Iñárritu, 2003) (4.45 min) where a mother learns of her daughters' deaths. The neutral condition featured an excerpt from *Son's Room* (Moretti, 2001) (4.54 min), showing everyday family life. This clip matched lower-level features like human faces and domestic settings. Films were shown in dubbed German. The negative clip elicited significantly higher negative arousal and valence (Borchardt et al., 2018). In study 2, participants viewed an excerpt from *Sophie's Choice* (Pakula, 1983) (10.00 min), where a mother must choose which child to save. All clips have previously been used for emotion induction during fMRI recordings (Borchardt et al., 2018; Gaviria et al., 2021; Hanich et al., 2014; Innes-Ker, 2015; Raz et al., 2016).

Subjective Continuous Emotional Intensity Annotations

Participants in study 1 watched a neutral and negative movie clip in a counterbalanced order during fMRI scanning. About 15 minutes post-scan, they re-watched the clips while providing continuous emotional intensity annotations via a trackball-operated mouse. A visual analog scale (VAS) from 0 ("not at all") to 250 ("very much") was displayed alongside the video for real-time adjustments (instructions in supplements). Annotations were sampled at 30Hz and downsampled to the fMRI sampling rate. Study 2 followed a similar procedure. Participants viewed the clip during fMRI scanning and re-watched it 15 minutes later, providing continuous emotional intensity annotations using a 7-point Likert scale (0 = "neutral" to 7 = "very intense"). For details, see 2016. Annotations were sampled at 10Hz and downsampled to match the fMRI sampling rate. In both studies, individuals were instructed to annotate their emotional intensity with regard to the first time they watched the movie.

Inter-subject Correlation of Engagement and Disengagement

We hypothesized that individuals exhibit more synchronized EE and ED in the negative clips as compared to the neutral one (Li et al., 2015; Malandrakis et al., 2011). To measure between subject similarity, we used inter-subject-correlation (ISC) with hamming distance for dichotomous variables (Ayala et al., 2013). Specifically, we computed pairwise correlations between each subject binary EE and ED timecourses (see methods for EE and ED calculations section).

Group Signature of Engagement and Disengagement

We next sought to assess the degree to which individuals collectively exhibited EE and ED phases throughout movie viewing. To this aim, we obtained a dynamic group signature reflecting the dynamic change in frequency of simultaneous reports of EE or ED across individuals. As a first step, we extracted time points of EE and ED from individual subjective annotations with an algorithm previously used to separate four phases of a temporal signal into rise, high, fall and low magnitude(Dessu et al., 2020; Kato et al., 2015; Shine et al., 2019). EE and ED periods were defined as rises and falls within the temporal fluctuations of subjective emotional intensity. To count as an EE or ED period, an emotional intensity score had to be within the 33rd and 67th percentile and exhibit a positive (EE) or a negative change (ED) between 2 subsequent time points. This way we obtained two binary time courses per participant: one indicating the presence of EE (yes/no), and the other indicating ED (yes/no) at each moment. Using a sliding window approach (window size of 40s; time step of 1 TR) we counted the number of EE or ED moments present across all individuals within each sliding window, resulting in two group signature time courses, i.e. one for EE, and one for ED. These binary time courses were then convolved with the hemodynamic response function (HRF) and used in combination with fMRI timecourses to train and test SVR models.

Image Acquisition

In study 1, functional MRI images were acquired on a Siemens Trio 3-T scanner with a twelve-channel radiofrequency head coil, using a T2*-weighted Echo Planar Imaging (EPI) sequence (repetition time (TR)/error time(TE) = 2,000/30 ms, flip angle = 70°, 64 × 64 matrix, field of view (FOV) = 192 × 192 mm^2, in-plane resolution = 3 mm^2, slice thickness = 3mm, 37 axial slices). The natural-viewing sessions consisted of 147 volumes. T1-weighted anatomical reference images were acquired using 3D-MPRAGE sequence (176 sagittal slices covering the whole brain, 1 mm^3 isotropic resolution, TR/TE = 1900/2.52 ms, 9° flip angle, 256 × 256 matrix). In study 2, functional MRI images were obtained by a GE 3-T Signa Excite echo speed scanner with an eight-channel head coil. Functional whole-brain scans were performed in interleaved order with a T2*-weighted gradient EPI pulse sequence (TR/TE = 3,000/35 ms, flip angle = 90°, pixel size = 1.56 mm, FOV = 200 × 200 mm, slice thickness = 3 mm, 39 slices per volume). The structural scans included a T1-weighted 3-D axial spoiled gradient echo (SPGR) pulse sequence (TR/TE = 7.92/2.98 ms, 150 slices, slice thickness = 1 mm, flip angle = 15°, pixel size = 1 mm, FOV = 256 × 256 mm). Active noise-cancelling headphones (Optoacoustic) were used (Raz et al., 2016).

Image Preprocessing

Both data sets were preprocessed using fMRIprep 1.3.2 (Esteban et al., 2019), which is based on Nipype 1.1.9 (K. Gorgolewski et al., 2011; K. J. Gorgolewski et al., 2019). The pipeline includes slice time and motion correction, susceptibility distortion correction, realignment, smoothing, and registration to common space. ICA-AROMA was performed to identify and remove motion-related components (Pruim et al., 2015-05-15). Physiological nuisance was corrected by extracting the mean time course of cerebrospinal fluid (CSF) and white matter (WM) calculated from the preprocessed images. The first two volumes of fMRI sequence of movie clips in study 1 were discarded to avoid T1 saturation effect. The last four volumes of fMRI during the negative movie clip in study 1 were removed, given

that the clip length was shorter than the entire scan session. One subject from study 1 was discarded due to excessive head movement, identified with mean frame-wise displacement (threshold = 0.20) resulting in 21 participants. Applying the same cut-off to study 2, three subjects were discarded resulting in forty-four participants in the final sample.

Image Analysis

Dynamic Predictive Modeling. To extract dynFC, for every participant, we extracted BOLD time courses using a 200 region parcellation by Schaefer et al. (2018), the Harvard-Oxford structural subcortical atlas (8 regions) Kennedy et al. (2016) and insula sub-regions (6 regions) Deen et al. (2011) resulting in a total of 214 brain regions). Next, we calculated dynFC of BOLD time series, computed as the Pearson correlation between pairs of regions (214 x 214 ROIs). We used a tapered sliding window with a length of 40 seconds and a tapering factor of 1 TR as done previously in naturalistic studies linking dynFC to subjective annotations (Petrican et al., 2021; Song, Finn, et al., 2021; Song, Park, et al., 2021).

In a next step, we used SVR to investigate whether the number of inter-regional connections could predict dynamic changes in the group signature response of EE and ED. First, to reduce dimensionality and enhance model accuracy, we identified neural features by selecting ROI-pairs connections that significantly correlated with the group signature response of either EE or ED (assessed using a one-sample t-test $p<0.01$). Next, we performed SVR with a Radial Basis Function (RBF) kernel to predict the group signature response of EE and ED based on these multivariate features in each subject's dynFC time courses. We performed a leave-one-out (LOO) cross-validation with a maximum iteration limit of 1,000 (sklearn.svm.SVR in Python). For every LOO iteration, the test subject's data was excluded for the calculation of the group EE and ED response time courses (see methods section). The predictive performance of the cross-validated models was evaluated using Pearson correlation between the predicted and the empirical scores for each

participant. For the sake of comparison, we also assessed model accuracy with two other metrics, i.e. mean Squared Error (MSE) and the coefficient of determination (r-squared). These metrics were computed individually for each subject and averaged across subjects to represent group-level performance. For these analyses, we used publicly shared codes by Song, Finn, et al. (2021) to ensure consistency and reproducibility (https://github.com/hyssong/NarrativeEngagement). To assess the generalizability of dynFC patterns related to EE and ED in emotionally intense narratives, independently of specific movie characteristics, we examined whether the neural features that predicted group signature response of EE in the negative movie clip in study 1 could predict group EE responses in the negative movie clip in study 2, and vice versa. This analysis was repeated for ED. We further tested if predictions were successful between the neutral clip of study 1 and negative clips of study 1 and 2, to investigate if there is some degree specificity to negative/high intense movie material (Fig. 2A). To assess the significance of the empirical mean correlation values, we computed the percentage of simulated correlation values resulting from the SVR phase-randomized EE and ED. The threshold was set to ensure that this percentage did not surpass 95% ($p < 0.05$); Fig. 2B. This iterative procedure was repeated 200 times. Throughout this section's analyses, we employed a modified version of the publicly available dynamic predicting modeling code initially developed by Song et al. (2021).

Networks Supporting Engagement and Disengagement. We further explored the involvement of canonical networks in predicting group signature response of EE and ED. We did this by grouping connections that significantly contributed to the model into canonical functional networks based on the network parcellation by Yeo et al. (2011). We then calculated the proportion of all participants' significant connections against all possible connections within a particular network. To evaluate the relevance of each network in predicting a group signature response of EE and ED, we generated null matrices by phase-randomizing the empirical emotional response time courses. This process involved

transforming the original data into the frequency domain, randomizing the phases, and then transforming the data back into the time domain to generate surrogate data with equivalent second-order properties as the original time series (Gias, 2023). These randomized data sets were input into the SVR model to generate simulated predictions, forming the null distribution. The significance of empirical mean r values was assessed by comparing the observed proportion of network pairs in the empirical data to the proportion of network pairs from the randomized data that surpassed the observed value. We inferred significance if more than 99% of the randomized pairs exceeded the empirical proportion ($p < 0.01$); Fig. 4A. We then grouped the significant network pairs into networks (Fig. 4B) and identified the region pairs predictive of EE and ED between movie clips in over 50% of the subjects. For visualization, we selected the two top nodes that exhibited most connections to other regions across the connectome (Fig. 4C).

Results

Negative Clips Synchronize Emotional Responses

We built a mixed-effects model predicting the ISC of binary EE and ED timecourses based on the interaction between condition (negative/neutral) and emotional shift (EE and ED). The interaction effect was significant (β = -0.03; 95% CI: [-0.03, -0.02], $p < .001$), and a Tukey post-hoc test revealed that the ISC (1) was higher during the negative as compared to neutral clip for EE (estimate[neutral-negative]= -0.233 $\mp$0.001; 95% CI: [-0.237, -0.230], $p < .001$) and ED (estimate[ED-EE]= -0.262 $\mp$0.001; 95% CI: [-0.265, -0.258], $p < .001$) and (2) the ISC was higher for ED as compared to EE in the negative clip (estimate[ED-EE]= 0.26 $\mp$0.001; 95% CI: [0.024, 0.027], $p < .001$) (Fig. 1). We also observed a higher prevalence of EE during negative clips compared to neutral clips (Fig. S1).

Functional Connectivity Patterns underlying Engagement and Disengagement Generalize across Studies

Next, we sought to assess if features of individuals' functional architecture are capable to predict EE and ED group signature responses (see methods section). Within each of the three movie clips, SVR models predicted the corresponding group signature response of EE and ED at levels above chance (EE: negative movie of study 1: $r = 0.823, p < 0.001; MSE = 0.454, R^2 = 0.546$; negative movie of study 2: $r = 0.722, p < 0.001; MSE = 0.552; R^2 = 0.448$; neutral movie of study 1: $r = 0.618, p < 0.005; MSE = 0.663, R^2 = 0.337$; ED: negative movie of study 1: $r = 0.689, p < 0.001; MSE = 2.673; R^2 = 0.331$; negative movie of study 2: $r = 0.593, p < 0.001; MSE = 8.444; R^2 = 0.302$; neutral movie of study 1: $r = 0.516, p < 0.005; MSE = 0.778, R^2 = 0.222$) Fig. S2.

In a next step, we investigated whether the neural features predicting the group signature response of EE and ED were generalizable across movie clips. To this aim, we tested if a model trained to predict a group signature response in one study is able to predict the group signature response in the other study and vice versa (Fig. 2). First, we tested predictions across negative clips of study 1 and 2. The model trained on dynFC of subjects watching the negative clip of study 2 successfully predicted the EE group signature in the negative clip of study 1 ($r = 0.572; p < 0.005, MSE = 50.84, R^2 = -1.30$), as did predictions in the opposite direction (negative clip of study 1 to negative clip of study 2: $r = 0.299; p = 0.005, MSE = 296.03, R^2 = 0.061$). The model trained on dynFC of subjects watching the neutral clip of study 1 successfully predicted the EE group signature in the negative clip of study 1 ($r = 0.334; p = 0.005, MSE = 30.8, R^2 = -0.389$), as did predictions in the opposite direction (negative clip of study 1 to neutral clip of study 1: $r = 0.18; p = 0.024, MSE = 14.553, R^2 = -45.167$). The model trained with the negative clip of study 2 did not predict the neutral clip ($r = -0.144; p = 0.999, MSE = 99.432, R^2 = -314.827$), similarly in the opposite direction

(neutral clip of study 1 to negative movie of study 2): $r = -0.028; p = 0.791, MSE = 401.26, R^2 = -0.272$. Taken together, predictions failed only between the neutral clip of study 1 and the negative clip of study 2 but succeeded between the neutral and negative clip of study 1.

The model trained on dynFC patterns to predict the group signature of ED in subjects watching the negative clip of study 2 successfully predicted the group signature of ED in the negative clip of study 1 ($r = 0.233, p = 0.005; MSE = 4.497; R^2 = -0.357$), as did predictions in the opposite direction (negative clip of study 1 to negative clip of study 2: $r = 0.167, p = 0.005; MSE = 13.673; R^2 = -0.128$). Conversely, predictions between negative movie clips (study 1 and 2) with the neutral movie clip of study 1 failed (from the negative clip of study 1 to the neutral clip of study 1: $r = -0.025, p = 0.59; MSE = 1.482; R^2 = -0.273$; from the negative clip of study 2 to the neutral clip of study 1: $r = 0.014, p = 0.60; MSE = 5.891; R^2 = -4.184$; neutral movie clip of study 1 to negative clip of study 1:$r = -0.109, p = 0.96; MSE = 1.344; R^2 = -0.324$), neutral movie clip of study 1 to negative clip of study 2: $r = -0.009, p = 0.42; MSE = 15.243; R^2 = -0.257$). Thus these findings suggest that the neural features predicting the group signature response for ED generalize across different narratives and participants for clips inducing high levels of emotional intensity, but not for neutral contexts.

Next, we sought to understand which canonical brain networks contributed to the group signature responses. To achieve this, we calculated the proportion of significant ROI connections within each network, divided by the total number of possible connections within that network for each movie and EE and ED signature responses separately (Fig. 4A). Across all movies from study 1 and 2, within visual network connections predominated (40%) among those positively correlated with the EE signature response, while only connections within Visual-SAL (14%) negatively correlated with EE. In contrast, for ED, no significant effects were observed across all movies combined. Across

the negative clips only from study 1 and 2, EE correlated positively with Visual-DMN (8%) and negatively mainly with Visual-DMN (13%) connections. In contrast, most of the connections correlating positively with ED in negative clips, were found within SAL connections (13%), while negatively correlated mainly with Visual-DMN (7%) connections (Table 2). The results for individual movies are summarized in Table 1.

Discussion

This study aimed at identifying generalizable functional connectivity (dynFC) patterns that underlie moment-to-moment changes in collective EE and ED during movie viewing. To do so, we used predictive modeling across movie fMRI data and continuous subjective annotations of affect from two independent studies (Borchardt et al., 2018; Raz et al., 2016). Specifically, we trained a dynFC-based model to predict group signatures of EE and ED in one data set and then tested whether the model could generalize to an independent study with a different clip and participant group and vice versa. We further aimed to test whether predictive dynFC patterns are specific to emotionally intense clips or if they also generalize to a neutral context.

Disengagement

A model trained to predict an ED group signature based on neural features in one negative movie clip successfully predicted an ED group signature in another negative movie clip, from independent samples with different narratives. However, since this predictive ability did not extend to the neutral clip, this may imply some level of specificity to negative or high emotional intensity contexts. We found that ED was linked to widespread positive associations with between-network connections, specifically for negative movie clips, . We hypothesized that ED may reflect an internally directed shift of processing involving the regulation of emotional responses. Thus, the positive association with wide-spread network connections may reflect enhanced cognitive regulation of emotions that relies on the involvement of large scale network communication (Cohen & D'Esposito, 2016; Morawetz

et al., 2020). Within these wide-spread connections, the strongest effects were found within the SMN and SAL and between them. Previous research has linked stronger recruitment of the sensorimotor cortex and regions of the salience network to attention directed toward interoceptive markers of arousal (Critchley et al., 2004; Wu et al., 2019). Thus, in our findings, the connections between SAL and SMN during negative emotional movies could indicate an intensified focus on bodily sensations induced by intense emotional experiences. This heightened interoception could play a role in ED, shifting attention away from external emotional stimuli towards internally directed processing and regulation of feeling states. We found that moment-to-moment fluctuations in the ED signature correspond to hub-like behavior within the ventromedial prefrontal cortex (vmPFC). The vmPFC has been implicated in both generating and regulating negative emotions through the formation of generalizable representations of negative emotions (Kragel et al., 2018). Since the highly connected nodes were exclusive to negative movie clips, the involvement of the vmPFC may facilitate the moment-to-moment regulation of negative emotions elicited by negative clips. In our previous study, we observed that inter-individual similarity in static vmPFC activations supported synchronized responses in both EE and ED in the negative clip of study 1 \citep{nanni2024neural}. In contrast, adopting a dynamic perspective, the current study reveals that moment-to-moment connectivity of the vmPFC may also reflect fluid adjustments in regulatory processes as emotional contexts shift throughout the movie. Taken together our findings suggest generalizable dynFC patterns across wide-spread large scale networks predicting group collective moment-to-moment changes of ED confined to emotionally intense contexts.

Engagement

In contrast to ED, prediction success for EE varied: predictions were successful across negative clips in both study 1 and study 2, and also across the neutral and negative contexts in study 1. However, predictions did not succeed between the negative clip of

study 2 and the neutral clip of study 1. Thus, idiosyncratic features within the participant group may have driven the successful cross-predictions across clips within one and the same study. Another or complementary interpretation could be that EE relies on network connections that support engagement regardless of emotional intensity levels. However, given that the predictions did not succeed between the neutral clip of study 1 and the negative clip of study 2 the latter interpretation may be more reasonable. Connections within the visual network and between the visual and the DAN were predictive of The EE group signature response. These networks are critical for goal-directed and stimulus-driven attention, aligning with their established roles in the voluntary control of visuo-spatial attention (Corbetta & Shulman, 2002; Kelley et al., 2008; Tosoni et al., 2023). Specifically, the connectivity in these networks may reflect attentional engagement with the emotional content of emotionally salient stimuli, potentially through mechanisms of visuospatial attention prioritizing emotionally salient stimuli. This prioritization could involve the DAN initiated selective top-down enhancement of visual processing areas linked to the perception and interpretation of emotionally relevant cues (Mohanty & Sussman, 2013). Interestingly, some Visual-to-DMN connections predicted EE in negative and positive directions, highlighting their dual role in promoting and reducing engagement. Prior research has shown that connectivity between the visual cortex and DMN changes depending on task demands, with positive coupling in some contexts like processing task irrelevant information (Chadick & Gazzaley, 2011) or automated information processing (Vatansever et al., 2017), and negative coupling during low cognitive effort (Weber et al., 2022). Further, during movie viewing, FC profiles of visual-DMN have shown to contribute to the classification of emotional states such as sadness or happiness (Xu et al., 2023). Together these findings may reflect the DMN's role in facilitating coordination between the DAN and visual networks to direct attention on emotionally salient stimuli.

Limitations

The use of multiple independent data sets for a common research objective bears limitations. The data sets differed in terms of cultural background and functional image acquisition methodologies and subjective annotation schemes. Specifically, one group utilized continuous pointer movements on a bar, while the other used double button presses to indicate a subjective change of emotional intensity. Nevertheless, we found a considerable consistency of results across data sets. Additionally, the study's scope was confined to movie clips with a sad affective tone, which may raise concerns about generalizability to other emotions. This limitation highlights the need for further research incorporating a broader spectrum of emotional stimuli to fully understand the dynamics of emotional processing across various settings.

Conclusion

Our findings reveal that emotional disengagement is characterized by neural patterns specific to emotionally intense situations, while emotional engagement shows broader neural responses that generalize across both high- and low-intensity contexts. This suggests that the neural mechanisms underlying engagement are more adaptable across diverse intensity contexts, whereas disengagement involves more processes specific to high-intensity emotions. These results provide insights into the neural responses driving shifts in emotional experiences and highlight generalizable neural patterns that can guide future research on emotional transitions and regulation in real-world scenarios.

Acknowledgment

We thank our collaborators Simone Grimm from Charité - Universitätsmedizin Berlin and Gal Raz from Tel Aviv University for sharing their data. We used ChatGPT and Perplexity AI for spell checks and help with phrasing and literature search assistance.

Author contributions

Melanni Nanni-Zepeda: Investigation, Data Curation, Formal analysis, Methodology, Software, Writing- Original Draft, review & editing. *Travis C. Evans:* Writing- Reviewing and Editing. *Audreyana Jagger-Rickels:* Writing- Reviewing and Editing. *Gal Raz*: Data source, *Talma Hendler*: Data source, *Yan Fan*: Data source, *Simone Grimm*: Data source, *Martin Walter*: Resources, Supervision. *Michael Esterman:* Writing- Reviewing and Editing. *Agnieszka Zuberer:* Conceptualization, Supervision, Writing- Reviewing and Editing.

References

Ayala, J., Shang, D., & Yakovlev, A. (2013). *Proc. integrated circuit and system design. power and timing modeling, optimization and simulation*. Springer.

Borchardt, V., Fan, Y., Dietz, M., Melendez, A. L. H., Bajbouj, M., Gärtner, M., Li, M., Walter, M., & Grimm, S. (2018). Echoes of affective stimulation in brain connectivity networks. *Cerebral Cortex (New York, N.Y.: 1991)*, *28*. https://doi.org/10.1093/cercor/bhx290

Chadick, J. Z., & Gazzaley, A. (2011). Differential coupling of visual cortex with default or frontal-parietal network based on goals. *Nature neuroscience*, *14*(7), 830–832.

Cohen, J. R., & D'Esposito, M. (2016). The segregation and integration of distinct brain networks and their relationship to cognition. *Journal of Neuroscience*, *36*(48), 12083–12094.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, *3*(3), 201–215.

Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature neuroscience*, *7*(2), 189–195.

Deen, B., Pitskel, N. B., & Pelphrey, K. A. (2011). Three systems of insular functional connectivity identified with cluster analysis. *Cerebral cortex*, *21*(7), 1498–1506.

Dessu, S. B., Price, R. M., Kominoski, J. S., Davis, S. E., Wymore, A. S., McDowell, W. H., & Gaiser, E. E. (2020). Percentile-range indexed mapping and evaluation (prime): A new tool for long-term data discovery and application. *Environmental Modelling & Software*, *124*, 104580.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Gaviria, J., Rey, G., Bolton, T., Delgado, J., Van De Ville, D., & Vuilleumier, P. (2021). Brain functional connectivity dynamics at rest in the aftermath of affective and cognitive challenges. *Human brain mapping*, *42*(4), 1054–1069.

Gias, C. (2023). CVX: Phase randomization: Simultaneous phase randomization of a set of time series, version 1.0.0.0.

Goldin, P. R., Manber, T., Hakimi, S., Canli, T., & Gross, J. J. (2009). Neural bases of social anxiety disorder: Emotional reactivity and cognitive regulation during social and physical threat. *Archives of general psychiatry*, *66*(2), 170–180.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python [Publisher: Frontiers]. *Frontiers in Neuroinformatics*, *5*. https://doi.org/10.3389/fninf.2011.00013

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Jarecka, D., Johnson, H., Notter, M. P., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Goncalves, M., Jordan, K., Waskom, M., Wong, J., Modat, M., Clark, D., … Ghosh, S. (2019). *Nipy/nipype: 1.2.2*. Zenodo. https://doi.org/10.5281/zenodo.3402255

Gross, J. J. et al. (2014). Emotion regulation: Conceptual and empirical foundations. *Handbook of emotion regulation*, *2*, 3–20.

Gruber, J., Harvey, A. G., & Purcell, A. (2011). What goes up can come down? a preliminary investigation of emotion reactivity and emotion recovery in bipolar disorder. *Journal of affective disorders*, *133*(3), 457–466.

Hanich, J., Wagner, V., Shah, M., Jacobsen, T., & Menninghaus, W. (2014). Why we like to watch sad films. the pleasure of being moved in aesthetic experiences [Place: US Publisher: Educational Publishing Foundation]. *Psychology of Aesthetics, Creativity, and the Arts*, *8*(2), 130–143. https://doi.org/10.1037/a0035690

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of film. *Projections*, *2*(1), 1–26.

Hox, J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

Iñárritu, A. G. (2003). 21 grams.

Innes-Ker, Å. (2015). Film and music in laboratory experiments: Emotion induction. *Music and the Moving Image*, *8*(2), 58–68.

Kato, S., Kaplan, H. S., Schrödel, T., Skora, S., Lindsay, T. H., Yemini, E., Lockery, S., & Zimmer, M. (2015). Global brain dynamics embed the motor command sequence of caenorhabditis elegans. *Cell*, *163*(3), 656–669.

Kelley, T. A., Serences, J. T., Giesbrecht, B., & Yantis, S. (2008). Cortical mechanisms for shifting and holding visuospatial attention. *Cerebral cortex*, *18*(1), 114–125.

Kennedy, D., Haselgrove, C., Fischl, B., Breeze, J., Frazier, J., Seidman, L., Goldstein, J., & Kosofsky, B. (2016). Harvard-oxford cortical and subcortical structural atlases. *Harvard Center for Morphometric Analysis*.

Koval, P., Butler, E. A., Hollenstein, T., Lanteigne, D., & Kuppens, P. (2015). Emotion regulation and the temporal dynamics of emotions: Effects of cognitive reappraisal and expressive suppression on emotional inertia. *Cognition and Emotion*, *29*(5), 831–851.

Kragel, P. A., Kano, M., Van Oudenhove, L., Ly, H. G., Dupont, P., Rubio, A., Delon-Martin, C., Bonaz, B. L., Manuck, S. B., Gianaros, P. J., et al. (2018). Generalizable representations of pain, cognitive control, and negative emotion in medial frontal cortex. *Nature neuroscience*, *21*(2), 283–289.

Kuppens, P., & Verduyn, P. (2015). Looking at emotion regulation through the window of emotion dynamics. *Psychological Inquiry*, *26*(1), 72–79.

Lettieri, G., Handjaras, G., Setti, F., Cappello, E. M., Bruno, V., Diano, M., Leo, A., Ricciardi, E., Pietrini, P., & Cecchetti, L. (2022). Default and control network

connectivity dynamics track the stream of affect at multiple timescales. *Social cognitive and affective neuroscience*, *17*(5), 461–469.

Li, T., Baveye, Y., Chamaret, C., Dellandréa, E., & Chen, L. (2015). Continuous arousal self-assessments validation using real-time physiological responses. *Proceedings of the 1st International Workshop on Affect & Sentiment in Multimedia*, 39–44.

Malandrakis, N., Potamianos, A., Evangelopoulos, G., & Zlatintsi, A. (2011). A supervised approach to movie emotion tracking. *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2376–2379.

McRae, K., Hughes, B., Chopra, S., Gabrieli, J. D., Gross, J. J., & Ochsner, K. N. (2010). The neural bases of distraction and reappraisal. *Journal of cognitive neuroscience*, *22*(2), 248–262.

Mohanty, A., & Sussman, T. J. (2013). Top-down modulation of attention by emotion.

Morawetz, C., Riedel, M. C., Salo, T., Berboth, S., Eickhoff, S. B., Laird, A. R., & Kohn, N. (2020). Multiple large-scale neural networks underlying emotion regulation. *Neuroscience & Biobehavioral Reviews*, *116*, 382–395.

Moretti, N. (2001). La stanza del figlio [Translated title: The Son's Room , event-location:].

Morgenroth, E., Vilaclara, L., Muszynski, M., Gaviria, J., Vuilleumier, P., & Van De Ville, D. (2023). Probing neurodynamics of experienced emotions—a hitchhiker's guide to film fmri. *Social Cognitive and Affective Neuroscience*, *18*(1), nsad063.

Nanni-Zepeda, M., DeGutis, J., Wu, C., Rothlein, D., Fan, Y., Grimm, S., Walter, M., Esterman, M., & Zuberer, A. (2024). Neural signatures of shared subjective affective engagement and disengagement during movie viewing. *Human brain mapping*, *45*(4), e26622.

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*, 117254.

Pakula, A. J. (1983). Sophie's choice.

Petrican, R., Graham, K. S., & Lawrence, A. D. (2021). Brain-environment alignment during movie watching predicts fluid intelligence and affective function in adulthood. *NeuroImage*, *238*, 118177.

Pruim, R. H. R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015-05-15). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, *112*, 267–277. https://doi.org/10.1016/j.neuroimage.2015.02.064

Raz, G., Touroutoglou, A., Wilson-Mendenhall, C., Gilam, G., Lin, T., Gonen, T., Jacob, Y., Atzil, S., Admon, R., & Bleich-Cohen, M. (2016). Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. *Cognitive, Affective, & Behavioral Neuroscience*, *16*, 709–723.

Saarimäki, H. (2021). Naturalistic stimuli in affective neuroimaging: A review [Publisher: Frontiers]. *Frontiers in Human Neuroscience*, *0*. https://doi.org/10.3389/fnhum.2021.675068

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, *28*(9), 3095–3114.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and emotion*, *23*(7), 1307–1351.

Sheppes, G., & Gross, J. J. (2011). Is timing everything? temporal considerations in emotion regulation. *Personality and Social Psychology Review*, *15*(4), 319–331.

Shine, J. M., Breakspear, M., Bell, P. T., Ehgoetz Martens, K. A., Shine, R., Koyejo, O., Sporns, O., & Poldrack, R. A. (2019). Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nature neuroscience*, *22*(2), 289–296.

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

Song, H., Park, B.-y., Park, H., & Shim, W. M. (2021). Cognitive and neural state dynamics of narrative comprehension. *Journal of Neuroscience*, *41*(43), 8972–8990.

Tosoni, A., Capotosto, P., Baldassarre, A., Spadone, S., & Sestieri, C. (2023). Neuroimaging evidence supporting a dual-network architecture for the control of visuospatial attention in the human brain: A mini review. *Frontiers in Human Neuroscience*, *17*.

Vatansever, D., Menon, D. K., & Stamatakis, E. A. (2017). Default mode contributions to automated information processing. *Proceedings of the National Academy of Sciences*, *114*(48), 12821–12826.

Visch, V. T., Tan, E. S., & Molenaar, D. (2010). The emotional and cognitive effect of immersion in film viewing. *Cognition and Emotion*, *24*(8), 1439–1445.

Weber, S., Aleman, A., & Hugdahl, K. (2022). Involvement of the default mode network under varying levels of cognitive effort. *Scientific reports*, *12*(1), 6303.

Wittchen, H.-U., Wunderlich, U., Gruschwitz, S., & Zaudig, M. (1997). Skid i. strukturiertes klinisches interview für dsm-iv. achse i: Psychische störungen. interviewheft und beurteilungsheft. eine deutschsprachige, erweiterte bearb. d. amerikanischen originalversion des skid i.

Wu, X., Shi, L., Wei, D., & Qiu, J. (2019). Brain connection pattern under interoceptive attention state predict interoceptive intensity and subjective anxiety feeling. *Human brain mapping*, *40*(6), 1760–1773.

Xu, S., Zhang, Z., Li, L., Zhou, Y., Lin, D., Zhang, M., Zhang, L., Huang, G., Liu, X., Becker, B., et al. (2023). Functional connectivity profiles of the default mode and visual networks reflect temporal accumulative effects of sustained naturalistic emotional experience. *NeuroImage*, *269*, 119941.

Yeo, B. T., Krienen, F., Sepulcre, J., Sabuncu, M., & Lashkari, D. (2011). 551 m. hollinshead, jl roffman, jw smoller, l. zöllei, jr polimeni, 552 et al., the organization of the human cerebral cortex estimated by intrinsic 553 functional connectivity. *Journal of neurophysiology*, *106*(3), 1125–554.

Zacks, J. M., & Magliano, J. P. (2011). Film, narrative, and cognitive neuroscience. *Art and the senses*, *435*, 454.

Zhou, F., Zhang, R., Yao, S., Dong, D., Feng, P., Kranz, G., Feng, T., & Becker, B. (2023). Capturing dynamic fear experiences in naturalistic contexts: An ecologically valid fmri signature integrating brain activation and connectivity. *bioRxiv*, 2023–08.

Table 1

Percentage of ROI-pairs within canonical networks connections significantly predicting the group signature responses for EE and ED listed for each clip separately.

		EE		**ED**	
study	**Movie-clip**	Positive corr.	Negative corr.	Positive corr.	Negative corr.
study 1	Negative clip of study 1 (21 grams)	Visual-Visual (52%) Visual-DAN (10%)	Visual-SAL (21%) Visual-DMN (18%)	Visual-Visual (26%) SMN-SMN (8%)	Visual-SAL (8%) Visual-CEN (5%)
	Neutral clip of study 1 (Son's room)	Visual-Visual (23%) Visual-DAN (5%)	Visual-Visual (17%) SubC-SubC (10%)	VIsual-Visual (23%)	Visual-DAN (5%)
study 2	Sophie's choice	Visual-Visual (29%) DAN-DAN (22%)	Visual-SAL (14%) Visual-SMN (13%)	SAL-SAL (17%) SMN-SAL (15%)	Visual-DMN (8%) VIsual-Visual (5%)

Note: Percentage of ROI-pairs within a canonical network significantly predicting the group signature responses for EE and ED. The two most prominent network pairs are shown.

Table 2

Percentage of ROI-pairs within a canonical network significantly predicting the group signature responses for EE and ED across studies.

	EE		**ED**	
Context	Positive corr.	Negative corr.	Positive corr.	Negative corr.
Negative movie-clips	Visual-DMN (8%)	Visual-DMN (13%) VIsual-CEN (10%) DAN-CEN (8%) DAN-SAL (7%) DAN-DMN (7%) DMN-CEN (3%)	SAL-SAL (13%) SMN-SAL (12%) Visual-Visual (12%) DAN-SMN (9%) DAN-VIS (8%) DAN-DAN (8%) SMN-SMN (6%) Visual-SMN (5%)	Visual-DMN (7%) DMN-CEN (4%)
All movie-clips	Visual-Visual (40%) Visual-DAN (14%)	Visual-SAL (14%)	–	–

Note: Percentage of ROI-pairs within a canonical network significantly predicting the group signature responses for EE and ED across studies. No significant pairs were found in relation to ED considering negative and neutral contexts together.

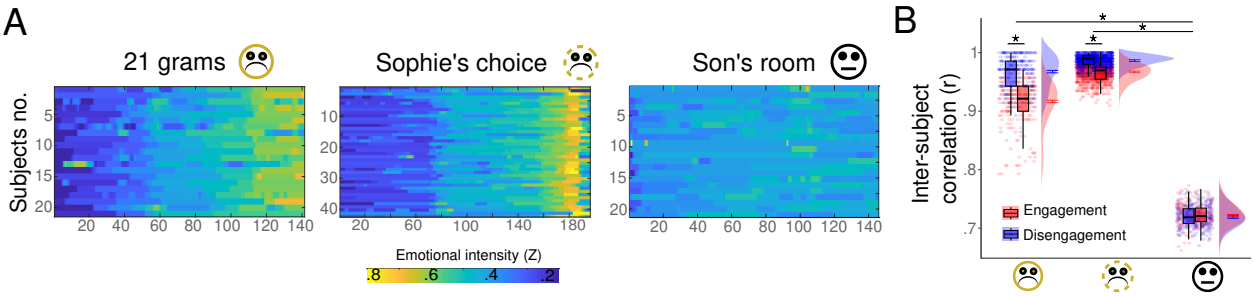

Figure 1. (A) Individual subjective emotional intensity annotations over time for each movie clip, z-scored and scaled between 0 and 1 for visualization purposes. (B) Estimated ISC of EE and ED periods derived from positive and negative shifts in subjective emotional intensity annotations across the three movie clips (negative clips from study 1 and 2 and neutral clip from study 1; (see methods section). *ISC*), Inter-Subject Correlation. *EE*, Emotional Engagement, *ED*, Emotional Disengagement.

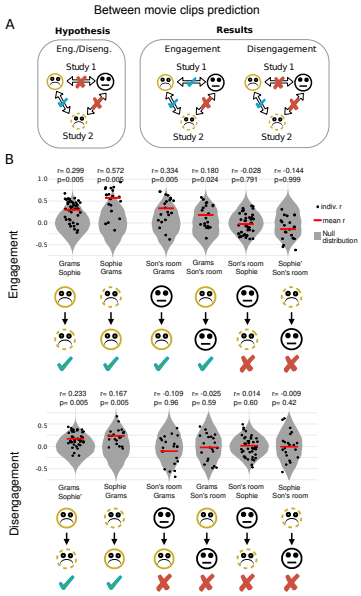

Figure 2. Functional connectivity patterns underlying engagement and disengagement generalize across independent studies (A) left panel: To infer specificity to negative clips, models trained to predict a group signature response of EE and ED in one negative movie clip in one study (e.g. study 1) must succeed in predicting the group signature response in another negative movie clip (e.g. study 2) and vice versa, but fail for predictions across negative and neutral clips. Based on these assumptions we infer generalization due to both independence from idiosyncratic features of the participant group and narratives in each study. Right panel: Both conditions were met for the group signature response of ED. For EE, only the first condition (between negative clip prediction) was met. (B) Predictive performance of dynFC features shown for each clip separately. The black dots in the scatter plots represent Pearson's correlations between predicted and observed group signature response for EE (top) and ED (bottom). Horizontal lines indicate the prediction quality reflected in the mean r across cross-validation folds. Violin plots illustrate the null distributions of mean prediction performance after phase-randomization. The significance of empirical r was computed based on the null distribution (one-tailed t-test). *DynFC*, Dynamic Functional Connectivity, *EE*, Emotional Engagement, *ED*, Emotional Disengagement.

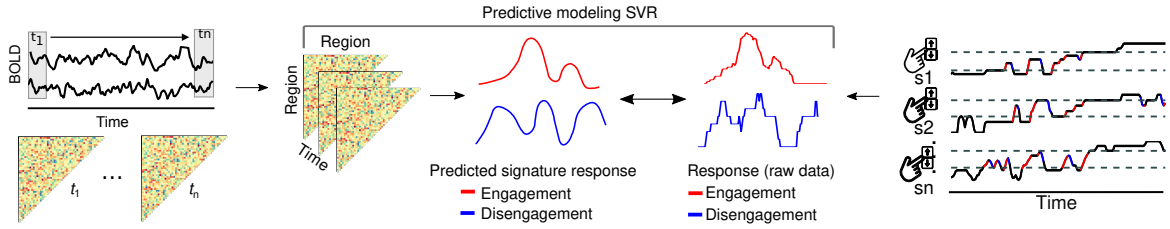

Figure 3. Visualization of how dynFC and group signature responses for EE and ED were derived. (A) DynFC and the group signature responses were calculated using a sliding window approach. From left to right: DynFC represents the Pearson correlation of BOLD time courses within each sliding window, while the group signature responses for EE and ED reflects the number of EE and ED phases within each sliding window aggregated across all participants. An EE and ED phase was identified as positive or negative shifts within a subject's continuous subjective emotional intensity annotation (see methods section). *DynFC*, Dynamic Functional Connectivity, *EE*, Emotional Engagement, *ED*, Emotional Disengagement.

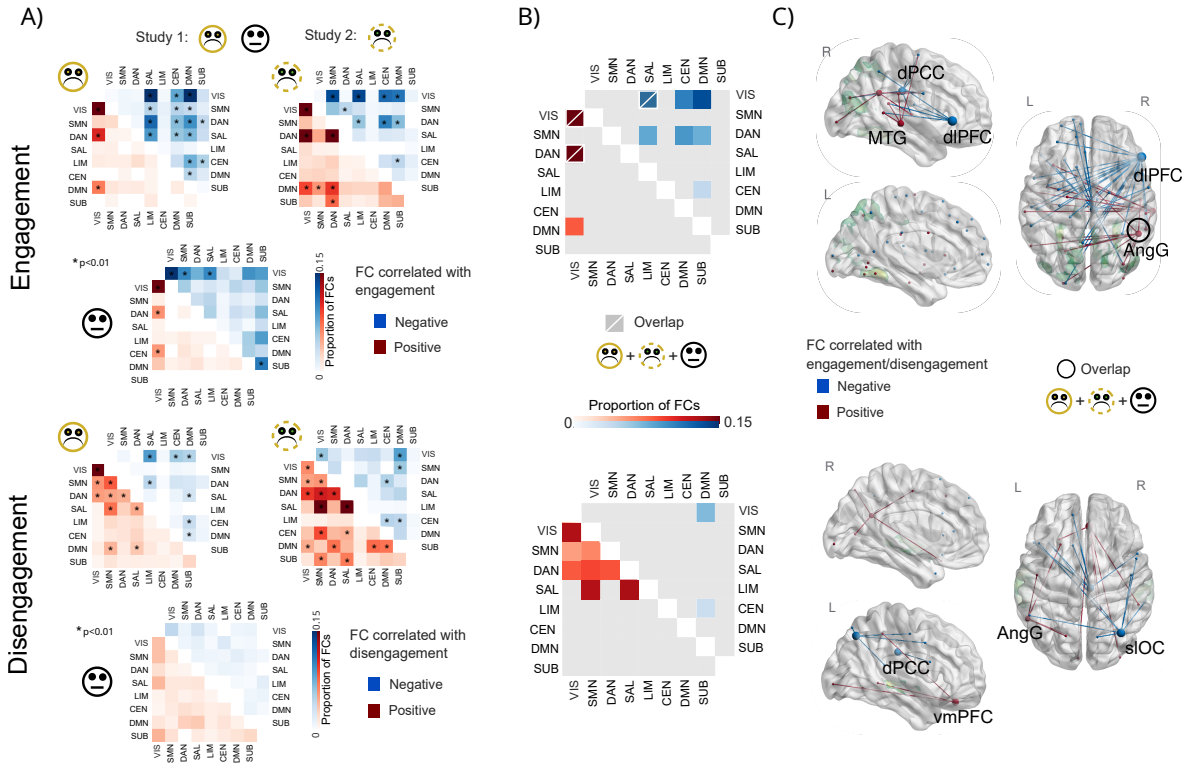

Figure 4. Functional connectivity within and across canonical networks predicts a group signature response of EE and ED. (A) predictive between-region connections were grouped into canonical functional networks (Schaefer et al., 2018). The lower triangle matrix (red) illustrates the proportion of FCs positively correlated with the group signature response for EE or ED, and the upper triangle matrix (blue) illustrates the proportion of FCs negatively correlated with the group signature response for EE or ED. Asterisks indicate network pairs selected above chance (one-tailed t tests, $p_{\mathrm{FDR}} < 0.001$). The intensity of colors indicates the proportion of selected connections from all possible ROI-to-ROI connections with regard to its network assignment. (B) Shared significant ROI-to-ROI connections across negative movie clips. The crossed gray square indicates connections shared among negative movie clips that also overlapped with the neutral clip. (C) Top two strongest ROI-to-ROI connections predicting a group signature response for EE and ED. Connections overlapping across all three movie clips are indicated by black circles, whereas those specific to negative clips do not show black circles. *EE*, Emotional Engagement, *ED*, Emotional Disengagement.