Generalizable Neural Models of Emotional Engagement and Disengagement

Melanni Nanni-Zepeda[1,2], Travis C. Evans[13,15], Audreyana Jagger-Rickels[12,14,16], 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, 07743 Jena, Germany

[2]Department of Psychiatry and Psychotherapy, University of Tübingen, 72076 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 reactivity and regulation are essential for mental well-being, as they are critical for managing emotional responses, fostering resilience, and maintaining overall emotional balance. Despite their importance, little is known about the underlying neural mechanisms that support these processes, which is needed to develop more effective interventions for emotional dysregulation and related mental health disorders. Previous research has predominantly relied on discrete trials and experimental task manipulations to isolate emotional processes. However, the use of emotional movies offers an ecologically valid method, allowing for the investigation of emotional processing in a context that simulates dynamic real-world experiences. Using functional magnetic resonance imaging (fMRI), during the viewing of neutral and sad movies, along with subsequent real-time emotional annotations of participants' experiences, we differentiated distinct phases of emotional engagement and disengagement. Functional connectivity across large-scale networks dynamically covaried with the viewer's engagement and disengagement patterns. A common pattern emerged across films: emotional engagement was primarily associated with connections within the Visual and between the Visual and Dorsal Attention Network. Only for highly emotional movies, emotional engagement was associated with connections between the Visual and Default Mode Network. In contrast, emotional disengagement was associated with connections mostly within the Visual and connections between the Sensorimotor Network and Salience Network.

We then applied predictive modeling to test whether these neural connectivity markers of engagement and disengagement generalized to other independent movie-watching datasets. Our findings revealed that disengagement patterns generalized specifically across intense clips, while engagement patterns generalized across all movie clips, including neutral contexts. Together this work helps to better understand cognitive and neural mechanisms underpinning engagement *in* and disengagement *from* emotionally evocative narratives, offering potential pathways for identifying generalizable neural

patterns that can inform future affective research and clinical applications.

Keywords: movie, emotion, emotional intensity, engagement, disengagement, fMRI

Generalizable Neural Models of Emotional Engagement and Disengagement

Significance statement:

How emotions unfold over time has been a subject of significant theoretical debate. In real life, negative emotional reactions often trigger a subsequent phase of regulation aimed at returning to a state of emotional equilibrium. However, the neural processes underlying these phases are rarely studied in their natural temporal progression. Functional MRI experiments using movies offer an opportunity to simulate these real-life scenarios, enabling individuals to empathize with characters and immerse themselves in hypothetical real-life situations. These experiments have revealed that emotional responses to negative movie content, along with the regulation of these responses, lead to significant reorganizations in the brain's functional architecture. Notably, these reorganizations differ when individuals react emotionally compared to when the emotional experience wanes.

Introduction

Our well-being profoundly depends on the way we engage *in* and disengage *from* emotions when interacting with others and the world. Theories of affect have widely acknowledged that there is a temporal progression of an emotional response and its regulation (e.g. (Gross et al., 2014; Scherer, 2009; Sheppes & Gross, 2011)pp. 16-17). However, empirically the neural mechanisms underlying these processes have traditionally been studied in isolation, disregarding their dynamic nature throughout time. In real life, emotions evolve temporally, and naturalistic scenarios like movies have been shown to mirror this phenomenon. Movies evoke a rich dynamic affective experience (Morgenroth et al., 2023; Saarimäki, 2021). This experience involves periods of both engagement and disengagement, mirroring the natural fluctuations that occur over time. However, when gauging emotional responses through movies, there is a diverse range of movie-related content, each contributing differently to the overall emotional experience. For instance, involvement in sad emotions could be induced by an emotionally charged dialogue where the protagonist discovers the loss of a significant other. In contrast factors such as the color

of the actor's costumes, people passing by in the background, the camera perspective, or the sentence structure might have little relevance to the emotional experience. 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. This implies that the findings may not extend uniformly across various naturalistic contexts (Nastase et al., 2020). 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. In the limited scope of fMRI studies employing dynamic predictive modeling to assess neural mechanisms underlying affect and attention, findings indicate a central role for the Default Mode Network (DMN). For instance, in a study using short (1-minute) audio narratives one study was able to predict six emotions of interest based on within-DMN connectivity (Saarimäki et al., 2022). Further, DMN connectivity was predictive of continuous attentional engagement in a narrative across a movie and audiobook (Song, Finn, et al., 2021). This suggests that the DMN may serve as a key hub in processing and predicting emotional experiences during naturalistic stimuli involving narratives. In a separate study, researchers achieved the prediction of dynamic fear responses across both an entire fearful movie and a short fearful movie excerpt (Zhou et al., 2023). This was accomplished by leveraging redundant widespread large-scale brain network connectivity, without relying on specific network subsystems. Together this work establishes a foundation for linking brain activity with varying levels of subjective attention or affect. Nonetheless, there is still a significant gap in our understanding of the neural mechanisms that initiate shifts in the subjective emotional experience. Here, we aim to assess the mechanisms involved in two transitional stages of dynamic adaptation to emotional stimuli, specifically focusing on when individuals recognize and respond to an

incremental shift in emotional intensity and the subsequent shift towards a decrease in intensity. The incremental transitional stage draws audiences deeper into the narrative with each nuanced shift (Visch et al., 2010). Throughout this dynamic progression, viewers become immersed in the nuanced shifts of emotional intensity, fostering a more profound connection with the narrative. The emphasis is on the shifts rather than distinct peaks, as viewers steadily navigate the evolving emotional landscape. The subsequent phase, marked by a shift towards a decrease of emotional intensity involves a winding down of the emotional journey. As the emotional journey recedes, viewers process the preceding intense emotional experience, e.g. through regulation strategies such as cognitive re-appraisal or expressive suppression (Koval, Butler, et al., 2015; Kuppens & Verduyn, 2015). The dynamics of these alternations play a pivotal role in shaping the emotional experience itself: the manner in which we engage in an emotion influences subsequent disengagement and, consequently, re-engagement in emotions (Koval, Brose, et al., 2015). Thus, it is critical to elucidate the temporal dynamics of engagement and disengagement as they naturally evolve throughout time, along with their corresponding neural signatures.

To achieve this objective, we employed two independent datasets, where participants retrospectively provided 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) during functional magnetic resonance imaging (fMRI) scanning. Within the temporal fluctuations of individual annotations of affect, we identified moments of positive and negative shifts of subjective emotional intensity, here called emotional engagement (EE) and disengagement (ED). Our initial inquiry was to assess dynamic functional connectivity patterns supporting collective group changes of EE and ED patterns throughout movie viewing. To this aim, we used support vector regression (SVR) to predict EE and ED across independent participant groups, movie clips of either high or low emotional intensity. Based on the evidence described above, we expected that both within and between DMN connectivity could play a central role in predicting EE and ED. However, given the lack of predictive studies of affect in naturalistic scenarios, we did not constrain our analysis to specific regions of interest. In summary, our analysis revealed that dynamic Functional Connectivity (dynFC) associated to ED, including connections within Visual, SAL and SMN-SAL, successfully generalized to negative narratives and participants when trained to predict ED in other narratives, but not to neutral narratives. This indicates that these patterns are specific to high emotional intensity situations, primarily marked by increases in between-network connections. Conversely, dynFC models associated to EE showed connections within Visual, Visual-DAN and Visual-DMN. The generalization to all contexts, including neutral narrative included only the connections within Visual and visual-DAN, indicating no specificity to high emotional intensity. However, the connections between the Visual-DMN were exclusive to negative narratives. These patterns were characterized by widespread decreases in between-network connectivity. This study marks initial attempts to differentiate the dynamic neural mechanisms underlying EE and ED during emotional cinematic experiences.

Material and Methods

Data Sources

We analyzed both fMRI and behavioral data from two distinct independent studies that employed similar paradigms and data acquisition methods. It is important to note that the research questions addressed in these previous studies were entirely unrelated to the focus of our current investigation. For the sake of clarity, we will herein refer to these studies as Study 1 or data set 1 (Borchardt et al., 2018) and Study 2 or data set 2 (Raz et al., 2016), respectively.

Participants

Study 1 comprised twenty-two healthy female volunteers aged 20-49 years (mean age 28.1 $\mp$ 6.5), participants underwent screening for the absence of neurological or psychiatric disorders using the short version of the Structured Clinical Interview for DSM-IV (SCID; Wittchen et al. (1997). The study protocols were in accordance with the latest version of the Declaration of Helsinki and approved by the institutional review board of the Charité, Germany. Study 2 comprised data from forty-four participants (25 females, 19 males) aged 21-37 years (mean age 26.73 $\mp$ 4.69). None of the participants had a known history of neurological or psychiatric disorders. All participants provided written informed consent, and the study received approval from the ethics committees of the Tel Aviv Sourasky Medical Center.

Naturalistic viewing paradigm

In Study 1, participants watched an excerpt from the movie "21 Grams" (Iñárritu, 2003) which presents a mother learning about the death of her two daughters in a car accident (total duration: 4.45 minutes). As a neutral condition, an excerpt from the movie "Son's Room" (Moretti, 2001) was shown where scenes of a family are presented in daily life (e.g., having a casual conversation at the dinner table, reading the paper; total duration: 4.54

minutes). This movie was selected as a comparison condition to include similar lower-level features such as the appearance of human faces and domestic scenes (4.54 minutes). Movie clips were shown in the dubbed German version. In Study 1, subjects rated the negative clip significantly of higher negative valence and arousal as compared to the neutral clip (see Borchardt et al., 2018). The two film clips have been used in previous studies for negative and neutral emotion induction, respectively (2018; Gaviria et al., 2021; Hanich et al., 2014).

In Study 2, participants viewed an excerpt from the movie *Sophie's Choice* (Pakula, 1983), depicting a mother facing the harrowing decision of choosing which of her two children to save from certain death. (total duration: 10.00 minutes). This film clip has been used in previous studies to induce sadness (Innes-Ker, 2015; Raz et al., 2016).

Subjective continuous emotional intensity annotations

Participants in Study 1 viewed the neutral and the negative movie clips in a counterbalanced order during fMRI scanning. Subsequently, approximately 15 minutes post-scanning, subjects rewatched both movie clips outside the scanner environment in an experimental room while concurrently providing continuous emotional intensity annotations for their emotional experiences during the initial fMRI viewing. The continuous emotional intensity annotations were collected using a trackball-operated mouse interface. Adjacent to the video playback, a visual analog scale (VAS) with a range from 0 ("not at all") to 250 ("very much") was displayed. Participants manipulated the trackball to adjust VAS values according to their emotional experience. The instruction provided was as follows: "Following, are 2 short ratings of the movie clips that you previously watched. You are going to see both movie clips again on the PC screen and are subsequently asked to rate how emotional you felt while watching" (German original in supplements). Emotionality annotations were sampled at a frequency rate of 30Hz and subsequently downsampled to match the fMRI sampling rate of 0.5Hz for data compatibility.

In Study 2, participants underwent a similar experimental procedure as in Study 1. Initially, participants viewed the movie clip while undergoing fMRI scanning. Approximately 15 minutes following the scanning session, participants were instructed to watch the same movie clip outside the scanner environment (experimental room). During this period, they provided continuous emotional intensity ratings, following the guidelines as in Study 1. To gather these subjective annotations, participants employed a 7-point Likert scale, with values ranging from 0 (representing "neutral") to 7 (indicating "very intense"). This Likert scale was finely divided, comprising three sublevels within each main level. Consequently, participants had to register their responses three times to progress to the subsequent main level of emotional intensity (for details, see Raz et al., 2016). As in Study 1, we sampled emotional intensity ratings at a frequency rate of 10Hz. Subsequently, to ensure alignment with the fMRI data, we downsampled the annotations to match the fMRI sampling rate of 0.33Hz.

Calculation of Engagement and Disengagement

Engagement and disengagement moments in time were extracted by using an algorithm previously employed to separate four phases of a temporal signal (trough, rise, peak, fall) (Dessu et al., 2020; Kato et al., 2015; Shine et al., 2019). Accordingly, 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 a certain range (the 33rd and 67th percentile) and (2) exhibit a positive and a negative change between 2 subsequent time points for EE and ED, respectively.

Group Response of Emotional Engagement and Disengagement. Our objective was to examine the functional architecture associated with the frequency of simultaneous reports of EE or ED across individuals at a given moment in time. In simpler terms, we aimed to create a group time course where each time point represented the number of moments when individuals reported EE or ED. To achieve this, we utilized the algorithm

mentioned above to analyze individual continuous annotations, resulting in two binary time courses for each individual. One of these time courses indicated the presence of EE (yes/no), while the other indicated the presence of ED at every moment in time (yes/no). Using a sliding window approach (window size of 40s; time step of 1 TR) we counted the number of EE moments present across all individuals within one sliding window in time. The same procedure was repeated for ED, resulting in one group signature time course for EE, and one group signature time course for ED. The group EE and ED time course was then convolved with the HRF. These convolved EE and ED time courses were then employed in conjunction with fMRI data to train and test SVR models. This approach of calculating fluctuations of an emotional group response enables us to extract information on both the similarity between individuals' subjective experiencing *and* directionality (higher and lower frequency of EE/ED reports).

Image acquisition

In Study 1, functional MRI images were acquired on a Siemens Trio 3-T scanner (Siemens, Erlangen, Germany) 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 located at the Wohl Institute for Advanced Imaging at the Tel Aviv Sourasky Medical Center. 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. (Taken from 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 sample 1 were discarded to avoid T1 saturation effect. The last four volumes of fMRI during the negative movie clip (samples one and two) were removed, given that the clip length was shorter than the entire scan session. One subject from sample one 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 of emotional cinematic experience. Our main research aim was to identify dynamic functional connectivity patterns predictive of group signatures of EE *in* and ED *from* emotionally intense negative content that generalize across independent movie data sets. 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), using a tapered sliding window with a length of 40 seconds and a tapering factor of 1 TR.

In the next step, we aimed to investigate if the number of inter-regional connections can predict a dynamic group response of EE and ED, respectively. First, we selected neural features based on significant correlations between ROI-pairs connections and the group response of either EE or ED. The significance of this correlation was determined through a one-sample t-test ($p<0.01$). We then used SVR with a Radial Basis Function (RBF) kernel to predict the group response of either EE or ED based on these multivariate features in each subject's functional architecture. Next, we did a leave-one-out (LOO) cross-validation with a maximum iteration set to 1,000 to fit a model on the train data and predict on test data (sklearn.svm.SVR in Python). For every LOO run, the subject's test data was excluded for the calculation of the group EE and ED response time courses (see methods). We then evaluated the predictive performance of the cross-validated models by using Pearson correlation between the predicted and the empirical scores for each participant. To determine the mean prediction accuracy, we utilized Mean Squared Error (MSE) and the coefficient of determination (r-squared). This allowed us to estimate three metrics for the sake of comparison: MSE, r-squared, and r. Treating each subject prediction as an independent sample within the model, we computed these three metrics individually for each subject. To represent the group collectively, we calculated the mean value for each metric. These analyses were done by using the publicly shared codes by Song, Finn, et al. (2021).

Between Data-Set Prediction of Engagement and Disengagement. To assess the generalizability of neural patterns related to EE and ED in emotionally intense narratives, independently of specific movie characteristics, we examined whether the neural features that predicted group EE responses in one negative movie clip could predict group EE

responses in a different negative movie clip from an independent study. To further test generalizability in various emotional contexts, successful predictions would need to extend across both negative and neutral narratives, as well as across participants. Specifically, this would involve predictions between the neutral narrative and the emotionally intense narratives (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 dynamic predicting modeling code initially developed by Song et al. (Song, Park, et al., 2021).

Engagement and Disengagement Networks. We further investigated which canonical networks were primarily involved in predicting emotional responses. To this aim, we grouped ROI-pairs that significantly contributed to the model into seven canonical functional networks based on a parcellation by Yeo et al. (2011). We then calculated the proportion of all participants' connections within a particular network, specifically those connections responsible for predicting emotional responses. First, we divided the number of significant connections in one specific network by the total number of connections within that network. To evaluate the relevance of one network in predicting emotional responses (EE and ED frequency), we generated null matrices through the phase-randomization of empirical emotional response time courses. This process involved transforming the original data into the frequency domain, randomizing the phases, and subsequently converting the data back into the time domain, thus generating surrogate data with equivalent second-order properties as the original time series (Gias, 2023). These randomized data sets were used as input for the SVR model to generate simulated predictions, thus forming the null distribution. To evaluate the significance of empirical mean r values, we calculated the percentage of network pairs from the randomized data that surpassed the observed proportion of network pairs in the empirical data. This comparison ensured that the

observed proportion did not exceed 99%, signifying statistical significance ($p < 0.01$); Fig. 4A.

To assess effects that were not reliant on the particular movie context, we superimposed the significant network pairs from all the movie clips. This was done to pinpoint consistent patterns across emotionally intense movie clips and the neutral/low-intensity movie clip (Fig. 4B). Our second objective was to identify the region pairs predictive of EE and ED across movie clips according to the following criteria: 1) The connections that manifested in over 50% of the subjects, 2) The ROI-pairs connections that were significant for at least 2 movie clips. For visualization, we selected the two nodes with the greatest number of ROI-pairs connections (Fig. 4C).

Results

Distinctive Functional Connectivity Patterns of Emotional Engagement and Disengagement

We analyzed both fMRI and behavioral data from two independent studies. Time points of EE were identified as moments when subjective emotional intensity rose, conversely, time points of ED were identified as moments in time when subjective emotional intensity decreased.

Distinctive patterns in behavior. In line with previous research demonstrating that emotion induction tends to induce some degree of agreement between subjective affective responses (Li et al., 2015; Malandrakis et al., 2011), we initially explored whether our data showed a similar pattern and if so, whether the increased alignment was indeed owed to the higher emotional intensity in one clip as compared to the neutral one. Hence, we tested if (1) the inter-subject correlation (ISC) of emotional shifts was higher for emotional clips as compared to the neutral clip and (2) if ISC was different for EE and ED. To investigate this, we employed a mixed-effects model predicting ISC by the interaction of condition (negative/neutral) and emotional shift (EE and ED). The interaction effect was significant

(β = -0.03; 95% CI: [-0.03, -0.02], $p < .001$). A post-hoc test revealed that ISC (1) was higher during emotionally intense as compared to neutral content for EE (estimate[neutral-negative]= -0.233 ∓0.001; 95% CI: [-0.237, -0.230], $p < .001$) and ED (estimate[ED-EE]= -0.262 ∓0.001; 95% CI: [-0.265, -0.258], $p < .001$) and (2) ISC was higher for ED as compared to EE in the negative condition (estimate[ED-EE]= 0.26 ∓0.001; 95% CI: [0.024, 0.027], $p < .001$). Together these findings suggest that emotionally intense negative movie material (1) aligns subjective emotional responses and (2) is enhanced for ED patterns as compared to EE patterns and (3) neither of the two is influenced by the degree to which subjects resemble in the overall number of emotional shifts. We further tested if the degree to which subjects align is influenced by how similar they are with each other in terms of the number of shifts. This was calculated by the normalized difference between subject scores, with normalization performed relative to the maximum absolute difference between subject scores. The inclusion of this similarity score did not improve the model fit, neither as 1) an additive effect, 2) interaction with either emotion condition or movie clip, nor 3) an interaction with both movie clip and shift type.

We then built a mixed-effects model to predict ISC of subjective emotional intensity by the interaction between the factors movie clip (21 grams, Son's room, Sophie's Choice) and emotional shift direction (EE/ED). This interaction effect between movie clip and emotional shift direction was significant (β = 0.03; 95% CI: [0.03, 0.04], $p < .001$). A post-hoc test revealed that ISC of ED was significantly higher than ISC of EE in negative movie of Study 1 (estimate[ED-EE]= 0.051 ∓0.001; 95% CI: [0.048, 0.054],$p < .001$), as well for the negative movie of Study 2 (estimate[ED-EE]= 0.020 ∓0.001; 95% CI: [0.018, 0.021],$p < .001$). While for the neutral movie, there was no significant difference between ISC for EE and ED (estimate[ED-EE]= -0.002 ∓0.002; 95% CI: [-0.006, 0.001], $p = .279$); see Fig. 1.

Generalizable Dynamic Functional Connectivity Patterns of Engagement and Disengagement. As a next step, we aimed to capture a group signature that reflects the

number of collective shifts of subjective emotional intensity (EE/ED) that subjects experience throughout the course of movie viewing. To do so, we used a sliding window of 40s with 1s time steps to count the number of emotional shifts across all individuals. This analysis resulted in two group time courses (EE and ED), here called group EE and ED dynamics index, indicating the group frequency of EE and ED throughout time (Fig.3). For each of the three movie clips individually, we then used a leave-one-subject-out (LOO) cross-validation approach to predict a group response of EE and ED based on each subject's dynFC patterns using nonlinear SVR models with a Radial Basis Function (RBF) kernel.

For all three movie clips (negative and neutral in Study 1 and negative in Study 2), SVR models predicted emotional the dynamic group index 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: 21 Grams: $r = 0.689, p < 0.001; MSE = 2.673; R^2 = 0.331$; negative movie of Study 2: $r = 0.593, p < 0.001; MSE = 8.444; R^2 = 0.302$; neutral movie of Study 1: $r = 0.516, p < 0.005; MSE = 0.778, R^2 = 0.222$).

The primary goal of this study was to uncover dynFC patterns capable of predicting the dynamic group index of EE and ED during viewing of movies that induce high levels of emotional intensity. To ensure the robustness of our findings, we investigated whether the neural features predicting emotional shifts were generalizable and not specific to the idiosyncrasies of a particular movie clip. If the predictions prove successful across negative contexts whilst failing in predictions between negative and neutral stimuli, it would indicate a potential specificity to contexts with high emotional intensity (Fig. **??**).

First, we tested predictions across negative clips. The model trained on dynFC of subjects watching the negative clip of Study 2 successfully predicted the EE group dynamics index 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 dynamics index 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):
$r = -0.028; p = 0.791, MSE = 401.26, R^2 = -0.272$. Together these findings imply that EE tends to generalize across contexts of varying emotional intensity.

The model trained on dynFC patterns of subjects watching the negative clip of Study 2 successfully predicted the group dynamics index for 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:
$r = 0.167, p = 0.005; MSE = 13.673; R^2 = -0.128$). However, 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:
$r = -0.009, p = 0.42; MSE = 15.243; R^2 = -0.257$).

Together these findings demonstrate successful generalization across different narratives and participants for clips inducing high levels of emotional intensity, but not for neutral contexts.

Next, we sought to characterize ROI-pairs that significantly predicted the EE and ED group dynamics index with regard to their assignment to canonical brain networks. To do

so, we divided the number of significant connections by the total number of possible connections within each network. This was done for each movie clip and emotional shift type (EE vs. ED) separately (see Fig. 4A).

For the intense negative clip of Study 1, the number of connections *positively* correlated with the EE group dynamics index mainly belonged to Visual-Visual (52%) and the DAN-Visual (10%) network pairs. The number of connections *negatively* correlated with the EE group dynamics index mainly belonged to the Salience Network (SAL)-Visual (21%) and the (DMN)-Visual pairs (18%). For the intense negative clip of Study 2, the number of connections *positively* correlated with the EE group dynamics index belonged to Visual-Visual (29%), DAN-DAN (22%), and Visual-DAN (19%) pairs. The number of connections *negatively* correlated with the EE group dynamics index mainly belonged to Visual-SAL (14%) and Somatomotor Network (SMN)-Visual (13%) pairs. For the neutral movie clip (Son's room), the number of connections *positively* correlated with the EE group dynamics index belonged to Visual-Visual (23%) and DAN-Visual (5%). The number of connections *negatively* correlated with the EE group dynamics index mainly belonged to Visual-Visual (17%) and Subcortical-Subcortical pairs (10%).

For *ED*, we found a broader distribution of connections across network pairs. For the negative clip of Study 1, the majority of *positive* connections *positively* correlated with ED, belonged to Visual-Visual (26%) and SMN-SMN (8%) regions, whereas the majority of connections *negatively* correlated with ED, belonged to SAL-Visual (8%) and Visual-CEN regions (5%). For the negative clip of Study 2, *positively* correlated connections with ED mainly belonged to SAL-SAL (17%) and SMN-SAL regions (15%), and *negatively* correlated connections with ED were in DMN-Visual (8%) and Visual-Visual regions (5%). In neutral movie clip *positively* correlated connections with ED mainly belonged to Visual-Visual (23%) and Visual-DAN (5%).

Next, we investigated which network patterns generalize across movie clips (1) within negative contexts and (2) those across negative *and* neutral contexts (Fig. 4). Exclusively

within negative contexts, the number of connections that positively correlated with the EE group dynamics index were primarily concentrated in Visual-DMN connections (8%), and the number of connections that negatively correlated with the EE group dynamics index were primarily concentrated in Visual-SAL connections (8%). Across all narratives (negative and neutral), the number of connections that positively correlated with the EE group dynamics index were primarily concentrated in Visual-Visual (54%) and Visual-DAN (20%) network pairs, and the number of connections that negatively correlated with the EE group dynamics index were primarily concentrated in Visual-DMN (13%) and Vis-CEN (10%) network pairs. For ED, there were no significant effects when considering negative and neutral contexts together. In contrast, exclusively for negative movie clips, the number of connections positively correlated with ED belonged to SAL-SAL (13%) and SMN-SAL (12%) pairs, and the number of connections negatively correlated with ED belonged to SAL-SAL (13%) and SMN-SAL (12%) pairs.

Discussion

This study investigated how neural response patterns support emotional engagement (EE) and disengagement (ED) as individuals experience emotions during movie viewing, with a focus on identifying generalizable neural mechanisms across independent data sets and movie clips. By analyzing temporal shifts in subjective emotional intensity, we uncovered distinct dynamic functional connectivity (dynFC) patterns associated with high-intensity negative and neutral contexts. In intense negative movie clips, more simultaneous ED reports were linked to increased between-network connections, while more simultaneous EE reports were linked to reduced between-network connections. This opposing pattern may suggest that ED involves regulatory cognitive processes needing extensive network communication (Cohen & D'Esposito, 2016; Morawetz et al., 2020), while EE may relate to reduced cognitive control and enhanced immersion in emotionally salient movie material. A model trained to predict ED based on neural features in one negative movie

clip successfully predicted ED in another negative movie clip, even when the movie clips were from independent samples with different narratives. However, this predictive ability did not extend to neutral contexts, implying some specificity to negative content. In summary, our results indicate that emotional movie viewing elicits consistent brain states that can be traced with subjective assessments of emotional experiences.

Our main objective was to assess whether (1) a model trained on dynFC to predict a group EE and ED dynamics index in one movie clip was predictive of a group EE and ED dynamics index in another movie clip and if so, (2) if the predictability is specific to negative and fails for neutral content. Given that Study 1 included both a negative and a neutral clip from the same individuals, and Study 2 featured a negative movie clip from an entirely independent sample with a different narrative, we hypothesized that meeting condition (2) would be adequate to infer some degree of specificity for emotional, as opposed to neutral, movie clip content. Our results indicate that both conditions were met for ED, supporting emotional specificity in this aspect. However, these conditions failed to hold for EE, suggesting a divergence in the mechanisms underlying EE and ED across different movie contexts. In our previous study, we used the same behavioral measures of emotional engagement (EE) and disengagement (ED) to differentiate brain-behavior association patterns based on inter-subject synchronization similarities through Representation Similarity Analysis (RSA) during naturalistic narratives (Nanni-Zepeda et al., 2024). In contrast, the current study extends this work by shifting the focus from inter-subject synchronization to a group manifold approach, incorporating a different narrative with comparable emotional elicitation. This shift allows us to examine the generalizability of brain functional architecture and behavior within the same emotional context at the group level through predictive modeling.

Disengagement. In general, group ED dynamics were predominantly predicted by widespread increases in between-network connections. The most substantial effects were observed with a greater number of connections within the SAL and between SAL and the

SMN. Previous research has linked the strength of SAL-SMN coupling to somatic arousal related to anxiety (Wu et al., 2019) and the attention directed toward interoceptive markers of arousal, such as heart rate (Critchley et al., 2004). In our study, the recruitment of connections between SAL and SMN during negative emotional movies could indicate an intensified focus on individuals' bodily sensations induced by intense emotional experiences. This heightened interoception may be a relevant factor influencing ED. Specifically, as individuals direct their attention away from emotional stimuli. In this state, they may become more internally focused paying attention to their bodily sensations as they process the emotional experience. Notably, a key hub that exhibited the highest number of connections in predicting group ED dynamics was the ventromedial prefrontal cortex (vmPFC). This region has been implicated in both the generation and regulation of negative emotions, with converging evidence suggesting it may function as a relay for generalizable representations of negative emotions (Kragel et al., 2018). Since the highly connected nodes were exclusive to negative movie clips, one could speculate that the involvement of the vmPFC may support the regulation of negative emotions induced by the negative movie clips.

Engagement. While cross-movie predictions for ED exhibited some specificity to negative movie content, successful predictions for EE were not limited to negative content but extended to both negative and neutral movie clips. This suggests that EE is associated with network connections that are more generalizable for engagement with stimuli, regardless of the degree of emotional intensity.

A greater number of connections within the visual network, as well as between the visual network and the DAN were predictive of the group EE index. Further, work linking networks like DAN and VN with goal-directed, stimulus-driven attention (see review (Corbetta & Shulman, 2002)), as well as the voluntary control of visuospatial attention (**tosoni2023neuroimaging**; **kincade2005event**; 2002). This may indicate that EE is linked to attentional processes, reflecting attentional engagement with emotional content

presented in the movie. In our study, we observed a subset of Visual-to-DMN connections that predicted both patterns aligned with EE and opposing patterns, suggesting that these networks are implicated not only in promoting EE but also in its reduction. Prior research has also demonstrated that the coupling between the visual cortex and various networks varies depending on specific task requirements (Chadick & Gazzaley, 2011).In some cognitive processes, the visual cortex and components of the DMN show positive coupling, while in others, activation of the DMN is associated with concurrent suppression in the visual cortex (2011; Daselaar et al., 2004; Shulman et al., 2007; Smallwood et al., 2012). Previous findings have indicated that differential connectivity patterns between the DMN and the visual network predict emotional states such as sadness or happiness during movie viewing (Xu et al., 2023). The interplay between the DMN and the visual network has been linked to their involvement in attention processes, where the DMN plays a crucial role in facilitating connectivity between the DAN and visual networks, which is essential for directing attention toward external salient stimuli (Anticevic et al., 2012). In sustained attention tasks, the DAN has been associated with effortful attention states that are more prone to errors. In contrast, sub-threshold activity in the DMN has been linked to more stable performance, suggesting a state of effortless attention (**esterman2013zone**; **esterman2014intrinsic**). This suggests a dynamic interaction between the DMN and attention networks guided by emotional content, critical in driving attentional engagement. However, further research is needed to investigate the specific brain regions implicated and the directionality of connectivity strength.

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. However, the inclusion of a neutral movie clip in the analysis enabled the assessment of neural patterns underlying emotional content processing.

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.

Ethical statement

Participants from both studies gave written informed consent. For study 1, protocols followed the latest version of the Declaration of Helsinki and were approved by the institutional review board of the Charité (Borchardt et al., 2018). For study 2, protocols were approved by the ethics committees of the Tel Aviv Sourasky Medical Center (Raz et al., 2016).

Data availability

The data used in this study is confidential. Upon publication, all code will be available on our Open Science Framework page for this study (https://osf.io/4938k/).

Author Contributions

M.NZ: Investigation, data curation, formal analysis, methodology, software, writing-original draft, review, and editing. T.C.E: Writing-reviewing and Editing. A.JR: Writing-reviewing and editing. G.R: Data source. T.H: Data source. Y.F: Data source. S.G: Data source. M.W: Resources, Supervision. M.E: Writing-reviewing and editing. A.Z: Conceptualization, supervision, writing- reviewing and editing.

Funding

M.NZ was supported by the National Council on Science and Technology CONACYT (Grant/Award Number: 486354).

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments We would like to express our gratitude to Simone Grimm and colleagues from Charité-Universitätsmedizin Berlin, for sharing their data of Study 1, Gal Raz and colleagues from Tel Aviv University for the data of Study 2. Our work was supported by a doctoral fellowship from the National Council of Science and Technology (CONACYT 486354) and the resources provided by the Jena University Hospital.

References

Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X.-J., & Krystal, J. H. (2012). The role of default network deactivation in cognition and disease. *Trends in cognitive sciences*, *16*(12), 584–592.

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.

Daselaar, S. M., Prince, S. E., & Cabeza, R. (2004). When less means more: Deactivations during encoding that predict subsequent memory. *Neuroimage*, *23*(3), 921–927.

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.

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.

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

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.

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., Brose, A., Pe, M. L., Houben, M., Erbas, Y., Champagne, D., & Kuppens, P. (2015). Emotional inertia and external events: The roles of exposure, reactivity, and recovery. *Emotion*, *15*(5), 625.

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.

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.

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.

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

Saarimäki, H., Glerean, E., Smirnov, D., Mynttinen, H., Jääskeläinen, I. P., Sams, M., & Nummenmaa, L. (2022). Classification of emotion categories based on functional connectivity patterns of the human brain. *NeuroImage*, *247*, 118800.

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.

Shulman, G. L., Astafiev, S. V., McAvoy, M. P., d'Avossa, G., & Corbetta, M. (2007). Right tpj deactivation during visual search: Functional significance and support for a filter hypothesis. *Cerebral cortex*, *17*(11), 2625–2633.

Smallwood, J., Brown, K., Baird, B., & Schooler, J. W. (2012). Cooperation between the default mode network and the frontal–parietal network in the production of an internal train of thought. *Brain research*, *1428*, 60–70.

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.

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.

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.

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.

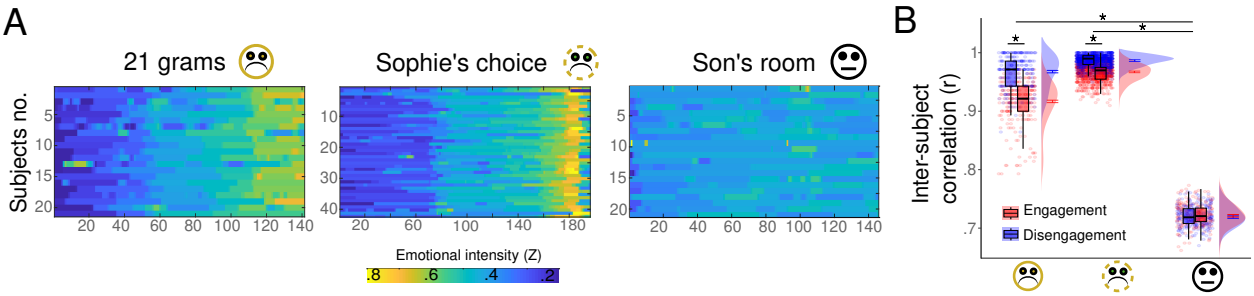

Figure 1. (A) Individual subjective emotional intensity annotations over time for each movie clip, z-scored and scaled between 0 and 1 for visualization purposes. (B) Estimated ISC of EE and ED periods, inferred from fluctuations in subjective emotional intensity annotations across the three movie clips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

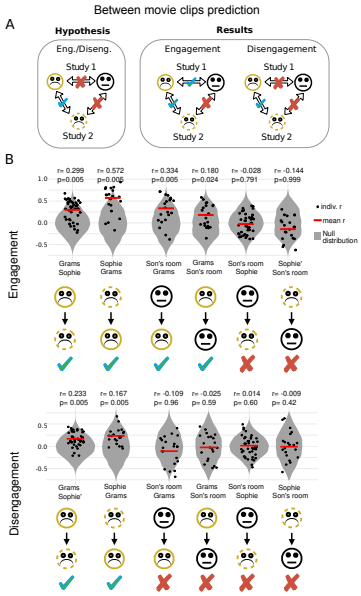

Figure 2. Generalizability of dynFC markers of the EE and ED group dynamics index. (A) Schematic representation of the conditions for dynFC markers of the EE and ED group dynamics index. To infer specificity to intense clips, dynFC patterns trained to predict the EE and ED group dynamics index in one negative movie clip (Study 1) must be (1) successful in predicting the EE and ED group dynamics index in the other negative movie clip (Study 2) and vice versa, implying both independence from idiosyncratic features of the participant group and narratives, *and* (2) fail for predictions between any of the emotional narratives and the neutral narrative. (B) Results: Conditions (1) and (2) were only met for ED. For EE, only condition (1) was fulfilled, while condition (2) failed. The black dots in the scatter plots represent Pearson's correlations between predicted and observed EE (top) or ED (bottom) based on each participant's dynFC features. 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 of models predicting the EE and ED group dynamics index after phase-randomization. The significance of empirical r was computed based on the null distribution (one-tailed tests). *DynFC*, Dynamic Functional Connectivity, *EE*, Emotional Engagement, *ED*, Emotional Disengagement.

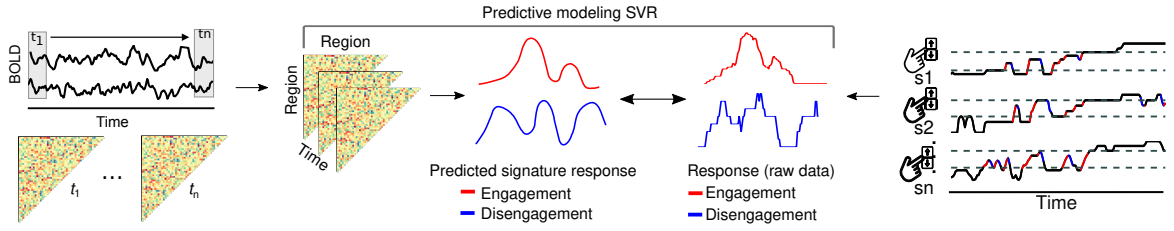

Figure 3. DynFC and Emotional Dynamics Index for EE and ED. (A) DynFC and the Emotional Dynamics Index were calculated using a sliding window approach. DynFC represents the Pearson correlation of BOLD time courses within a window (left panel), while the Emotional Dynamics Index for EE and ED reflects the number of EE and ED phases within each window across all participants. An EE and ED phase was determined by positive or negative shifts within a subject's continuous subjective emotional intensity annotations, within predefined intra-individual thresholds (right panel). *DynFC*, Dynamic Functional Connectivity, *EE*, Emotional Engagement, *ED*, Emotional Disengagement.

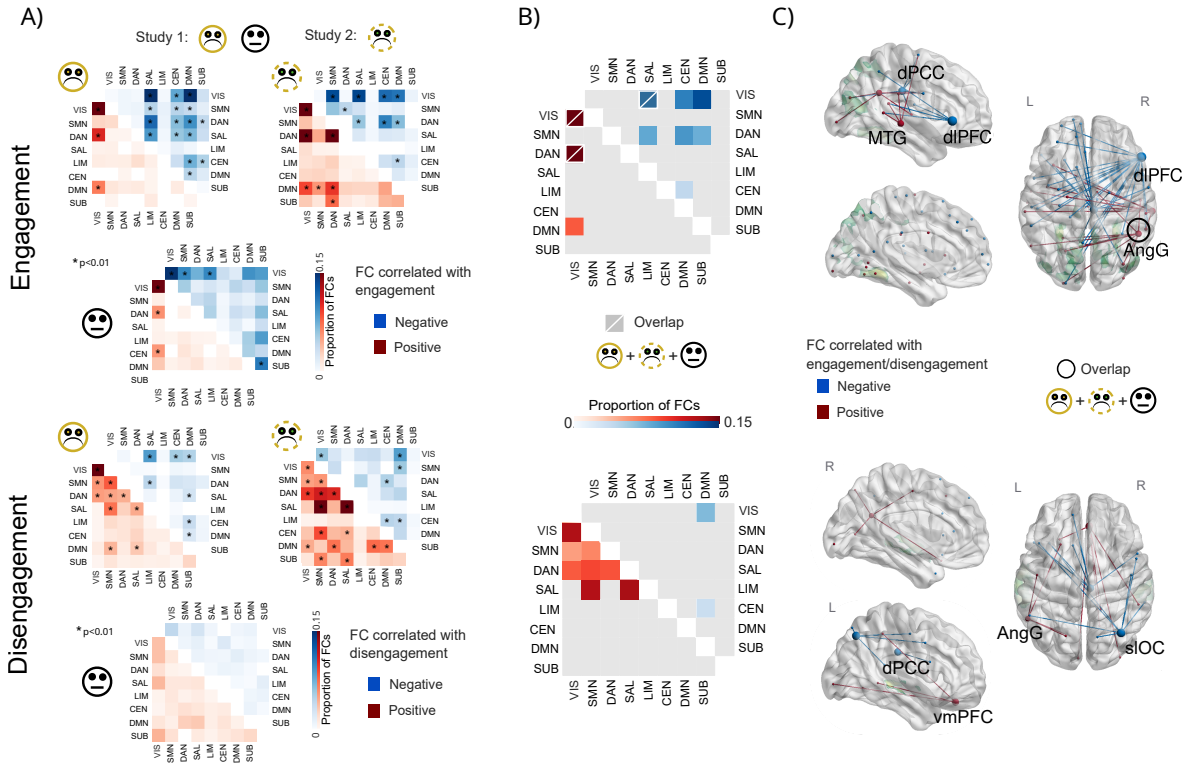

Figure 4. DynFC predicts Emotional Dynamics Index of EE and ED. (A) Within each cross-validation fold, dynFCs were grouped into predefined canonical functional networks (Schaefer et al., 2018). The lower triangle matrix (red) illustrates the proportion of FCs positively correlated with the Emotional Dynamics Index for EE or ED, and the upper triangle matrix (blue) illustrates the proportion of FCs negatively correlated with the Emotional Dynamics Index for EE or ED, respectively. Asterisks indicate network pairs selected above chance (one-tailed tests, FDR-P < 0.001). The intensity of colors indicates the proportion of selected FCs from all possible network-to-network connections. (B) Shared significant FCs across negative movie clips. The crossed gray square indicates FCs shared among negative movie clips that also overlapped with the neutral clip. (C) Top two highest connected nodes in the predictive model. FCs overlapping across all three movie clips are indicated by black circles, whereas those specific to negative clips do not show black circles. *DynFC*, Dynamic Functional Connectivity, *EE*, Emotional Engagement, *ED*, Emotional Disengagement.