NeuroImage 216 (2020) 116618

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Decoding dynamic affective responses to naturalistic videos with shared neural patterns

Hang-Yee Chan [a,*], Ale Smidts [a], Vincent C. Schoots [a], Alan G. Sanfey [b], Maarten A.S. Boksem [a]

[a] Department of Marketing Management, Rotterdam School of Management, Erasmus University Rotterdam, the Netherlands
[b] Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

A B S T R A C T

This study explored the feasibility of using shared neural patterns from brief affective episodes (viewing affective pictures) to decode extended, dynamic affective sequences in a naturalistic experience (watching movie-trailers). Twenty-eight participants viewed pictures from the International Affective Picture System (IAPS) and, in a separate session, watched various movie-trailers. We first located voxels at bilateral occipital cortex (LOC) responsive to affective picture categories by GLM analysis, then performed between-subject hyperalignment on the LOC voxels based on their responses during movie-trailer watching. After hyperalignment, we trained between-subject machine learning classifiers on the affective pictures, and used the classifiers to decode affective states of an out-of-sample participant both during picture viewing and during movie-trailer watching. Within participants, neural classifiers identified valence and arousal categories of pictures, and tracked self-reported valence and arousal during video watching. In aggregate, neural classifiers produced valence and arousal time series that tracked the dynamic ratings of the movie-trailers obtained from a separate sample. Our findings provide further support for the possibility of using pre-trained neural representations to decode dynamic affective responses during a naturalistic experience.

1. Introduction

In recent years, numerous studies have attempted to use signals recorded from the brain to infer affective states (Chikazoe et al., 2014; Kassam et al., 2013; Kim et al., 2015; Klasen et al., 2011; Knutson et al., 2014; Kragel and LaBar, 2015; Peelen et al., 2010; Saarimäki et al., 2018, 2016). Using either a circumplex affect model (Russell, 1980) or discrete emotion classes (Panksepp, 1982), these functional magnetic resonance imaging (fMRI) studies have found that neural representations – i.e., blood oxygenation level dependent (BOLD) signal patterns – observed in various brain regions such as subcortical structures (amygdala and thalamus), precuneus, medial prefrontal, cingulate, temporal and primary somatosensory cortices are associated with certain emotional experiences (Kragel and LaBar, 2015; Saarimäki et al., 2016).

What distinguishes more recent studies from the preceding decade of fMRI research on neural processing of affect (Kober et al., 2008; Lindquist et al., 2012) is three-fold. First, the increasing popularity of multivariate pattern analysis (MVPA) (Haxby, 2012; Haxby et al., 2001; Norman et al., 2006; Nummenmaa and Saarimäki, 2019) allows for attempts to extract information from neural activities of distributed brain regions instead of individual voxels. Second, rather than locating neural substrates associated with mental processes, there is a shift in focus towards making predictions of mental states based on neural patterns (Poldrack, 2011). Lastly, there has been an increase in the use of naturalistic, complex stimuli in social and affective neuroscientific studies, which stems from the recognition that more reliable neural responses are observed with these stimuli as compared to their well-controlled yet simplified counterparts (Adolphs et al., 2016).

In experimental studies, emotions are often evoked by various stimuli such as video, music, or pictures. An important question is whether any neural representations associated with these affective stimuli are specific within or common across modalities. Recent evidence suggests that neural representations, at least to some extent, are modality-general and also individual-invariant. For example, common affective neural patterns have been found between words and pictures (Kassam et al., 2013), movies and imagery (Saarimäki et al., 2016), and faces and situations (Skerry and Saxe, 2014). Moreover, Chikazoe et al. (2014) found that participants shared neural patterns of valence evoked by images and taste, while Kragel and LaBar (2015) observed distinct neural patterns for various emotions that were shared among participants and across modalities (music and movies). Recently, Kragel et al. (2019) observed that neural patterns at the occipital cortex could be used to classify emotional experience while viewing affective visual stimuli. In that study, they conducted separate analyses on affective pictures and movie clips, and

* Corresponding author. Department of Marketing Management, Rotterdam School of Management, Burgemeester Oudlaan 50, 3062 PA, Rotterdam, the Netherlands.
E-mail address: chan@rsm.nl (H.-Y. Chan).

https://doi.org/10.1016/j.neuroimage.2020.116618
Received 20 June 2019; Received in revised form 21 January 2020; Accepted 5 February 2020
Available online 7 February 2020

found that in each case there were distinct neural patterns in the occipital cortex that could be used to identify emotional categories of the stimuli and that these patterns were shared among participants.

The studies discussed above have provided evidence that individuals share neural representations of affect and that these representations are common across modalities. However, emotions are essentially momentary experiences, as evidenced by past studies examining the temporal changes of affect in an extended period of emotionally varying experience (Brans and Verduyn, 2014; Frijda, 2009; Mesquita, 2010; Scherer, 2009), and there have also been attempts to map out these dynamics at the neural level. These earlier studies, however, adopted the classic univariate approach to the analysis of brain data. For example, an early approach compared modelling brain responses to affective movie clips on block regressors to individual continuous ratings, and found that regressing on individual continuous ratings uncovered more frontal, temporal and subcortical areas associated with positive arousal than block contrasts did (Goldin et al., 2005). Similarly, Chaplin and colleagues found that dynamic changes in arousal elicited by music are associated with neural activation in the inferior frontal gyrus (Chapin et al., 2010). More recently, there have been attempts to apply multivariate approaches to investigate the temporal patterns of affect. For example, functional connectivity of the salience network (anterior cingulate cortex, insula, amygdala) tracked negative arousal during watching a stressful movie clip (Young et al., 2017). In multiple studies (Raz et al., 2016b, 2016a, 2012), moment-to-moment functional connectivity of salience and limbic networks has been linked to continuous intensity ratings of negative emotions such as fear, anger and sadness. Lastly, Nummenmaa et al. (2012) investigated how inter-subject neural synchronization tracked continuous changes in valence and arousal while watching short movie clips, and found that negative valence was associated with inter-subject synchronization at subcortical regions (thalamus, nucleus accumbens) and default mode network, while high arousal was associated with synchronization at somatosensory cortices.

Thus, extant studies suggest that (1) activation of certain brain areas may track dynamic changes of affect over time, (2) that individuals share common spatial neural patterns reflecting affective states elicited by brief exposure to univalent stimuli (ones designed to evoke a specific emotion) and (3) that these patterns show significant overlap between modalities (e.g., emotions elicited by pictures, videos, or music). However, whether these cross-modal, shared affective patterns generalize to time-variable emotional experiences remains to be examined. Two follow-up questions would be: (1) Do neural patterns extracted during sustained episodes of affect generalize to settings where affective states are changing continuously? (2) Are these neural patterns common across individuals, also in dynamic settings? In other words, can we decode an individual's dynamic changes of affective states, using a model derived from neural responses elicited by static experiences of affect in different individuals?

In this study we aim to answer these outstanding empirical questions. In addition, we provide a more methodological contribution by testing whether the application of functional alignment of fMRI data may improve the extraction of shared affective neural patterns. Past findings have shown that functional alignment as a preprocessing step improved performance of pattern analysis in visual processing (Haxby et al., 2011; Nishimoto and Nishida, 2016), although its application on higher order mental processes is less common. Conventionally, inter-subject alignment is typically first carried out based on anatomical features, such that each individual is registered into a common coordinate space, a procedure commonly referred to as spatial normalization (Gholipour et al., 2007). However, this procedure alone does not account for individual variations in the neural encoding of fine-scaled information (Conroy et al., 2013). To this end, more recent attempts focus on functional alignment, i.e., maximizing inter-subject alignment based on local brain functions (Conroy et al., 2013; Haxby et al., 2011; Sabuncu et al., 2010). For example, Haxby et al. (2011) developed a procedure called hyperalignment, which derives inter-subject functional correspondence by aligning different individuals' regional neural responses to complex, dynamic stimuli, such as movies. The procedure produces transformation matrices that project the same brain regions from different individuals into a shared high-dimensional space based on their common functional responses. This method has been found to be beneficial to improving inter-subject classification accuracy of neural patterns in visual processing (Guntupalli et al., 2016; Haxby et al., 2011). The application of functional alignment to uncover shared affective neural patterns, however, has been limited so far.

In summary, the aim of the present study is to test directly whether affective neural representations, based on brief and isolated episodes of affect (evoked by affective pictures), could be used to decode a time-variable emotional experience (here, watching movie-trailers). We investigate whether the use of functional alignment improved the extraction of affective neural patterns. We examine the validity of affective neural patterns in three steps. First, we verify whether different individuals indeed shared neural representations when they viewed affective pictures. Second, we examine whether the decoded affective time series during movie-trailer watching tracked the retrospective summary ratings of valence and arousal by the participants. Third, we test whether these affective time series of movie-trailers, in aggregate, tracked the continuous ratings of affective experience by a separate sample.

2. Materials and methods

2.1. Overview of the study design and analysis

We report findings from a study on video watching during fMRI scanning, part of which has been previously reported (Chan et al., 2019), with previously unreported fMRI data of affective picture viewing and self-report ratings of affect. Whereas the previous report focused on neural features that predicted out-of-sample, market-level, popularity of naturalistic stimuli, here we explore individual affective responses to movie-trailers. In addition, we conducted an online study in which a separate group of participants provided continuous ratings of the affective experience of the videos.

The overview of the study design and analysis is as follows (Fig. 1). In the fMRI study, participants viewed affective pictures, then watched the movie-trailers in random order. They also provided valence and arousal ratings for each of the trailers. We first determined brain regions sensitive to affective picture categories by univariate analysis (feature selection). Then, we built across-participants valence and arousal classifiers based on these neural responses. Before doing so, we performed a hyperalignment procedure to improve functional alignment between participants (Haxby et al., 2014). Specifically, we obtained hyperalignment parameters based on neural responses to movie-trailers, then applied the transformation to both movie-trailer and affective picture neural data. Valence and arousal classifiers were built with the hyperaligned affective picture data from all-but-one participants. Based on these two classifiers, the remaining out-of-sample participant's neural responses to affective pictures and movie-trailers were decoded.

To examine the validity of the classifiers, three tests were conducted. First, we examined whether classifiers correctly identified the affective categories of the pictures viewed by the participants. Second, valence and arousal classifications based on the individual's neural responses to the movie-trailers were compared to their self-reported valence and arousal ratings of the corresponding videos. Finally, to determine whether the neural valence and arousal time series tracked the dynamic affective profiles of the movie trailers, we asked a separate online sample to continuously rate the trailers in both valence and arousal dimensions. The valence and arousal time series from the fMRI and online samples were then compared.

2.2. FMRI study

2.2.1. Participants

We recruited 31 healthy volunteers via the recruitment system of the

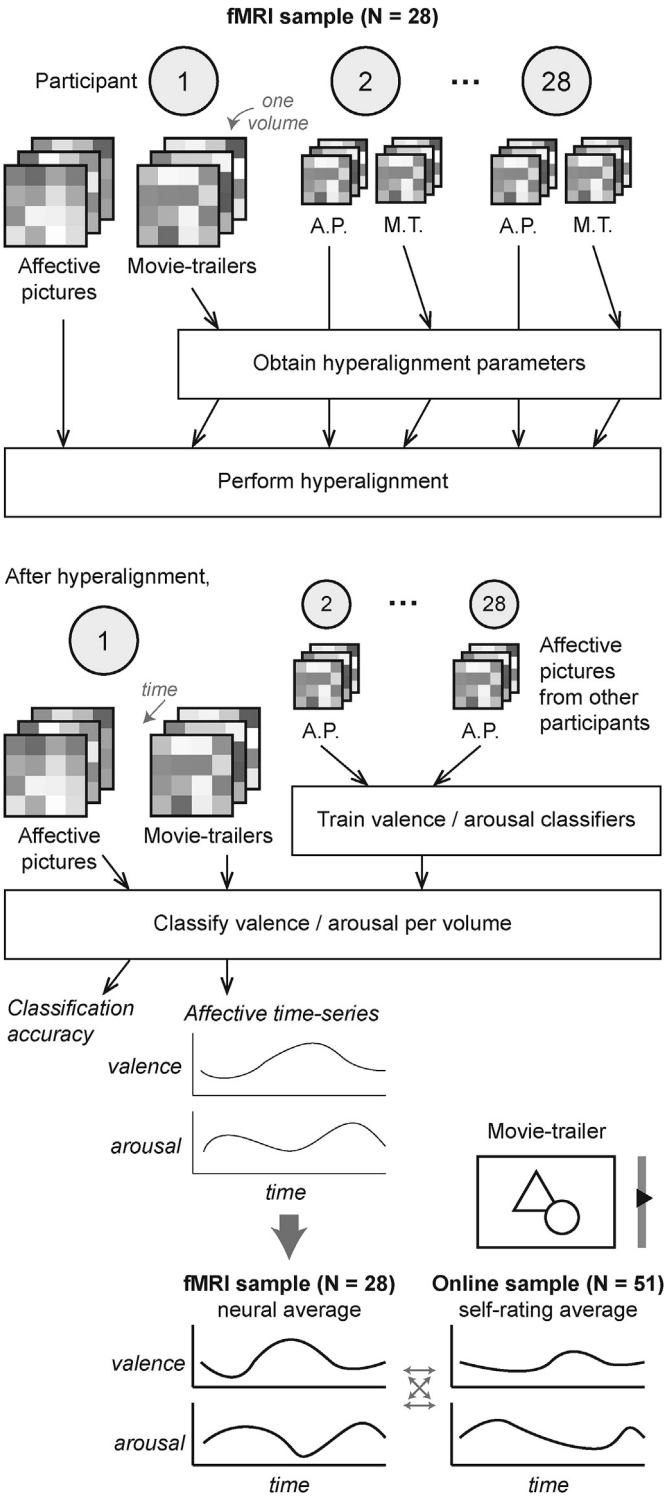

Fig. 1. Overview of the study design and analysis.

university. One participant had to be excluded due to falling asleep in the scanner, and two more due to excessive head movements (>5 mm movement in any direction between successive images in one or both scanning sessions). These are omitted from all behavioral and neural analyses. The final sample thus consisted of 28 participants (14 women, mean age 20.9). The study was approved by the local ethics committee, in line with the Declaration of Helsinki. All participants signed informed consent prior to participation.

Table 1
Movie trailers used in the study (movie information from the International Movie Database). Genre legend: A = action; B = biography; C = crime; D = drama; F = fantasy; Mu = musical; My = mystery; R = romance; Sf = sci-fi; Sp = sport; T = thriller; W = western.

Title	Code	Year released	Genre	Length (s)
The Namesake	M1	2006	D	150
Pollock	M2	2000	B,D	141
The Town	M3	2010	C,D,T	145
Quills	M4	2000	B,D	143
Love & Basketball	M5	2000	D,R,SP	142
Takers	M6	2010	A,C,D	141
Northfork	M7	2000	D,F	143
Waist Deep	M8	2006	A,C,D	133
Michael Clayton	M9	2007	C,D,M	134
The Debt	M10	2010	D,T	142
Idlewild	M11	2006	C,D,M	130
To Save a Life	M12	2009	D	137
Confidence	M13	2003	C,T	125
Extraordinary Measures	M14	2010	D	147
Impostor	M15	2001	D,SF,T	66
Harsh Times	M16	2005	A,C,D	131
The Warrior's Way	M17	2010	A,F,W	136
Gracie	M18	2007	B,D,SP	131

2.2.2. Materials and procedures

Participants received a fixed fee of €10 per hour. After signing informed consent, participants were taken to the MRI scanner, and performed a number of tasks in separate scanning sessions. The two tasks presented in this report were movie-trailer watching and picture viewing.

2.2.2.1. Movie-trailer watching. Participants viewed 18 unedited movie trailers (see Table 1), chosen with the aim to represent a wide range of genres and commercial successfulness (as measured by box office returns) while avoiding extreme content (e.g., horror). (See supplementary material S1 for more details on movie selection.) Participants were screened prior to inclusion to ensure that they had not seen any of our selected movies already. The 18 movie trailers were presented to the participants in random order. (One participant did not watch the movie trailer M16 due to a technical error.). Presentation of the trailers was preceded and followed by a picture of the cover (3s) of the DVD to make clear which movie the trailer belonged to. Immediately following each movie-trailer, participants were asked to state, via button presses, (1) their expected liking of the movie on a five-star-scale (zero stars possible with half-star increments); (2) willingness to pay (WTP) to obtain a DVD of the movie (€0 - €2.5, with 25 cents increments) under a Becker-DeGroot-Marschak (BDM) auction procedure (Becker et al., 1964); (3) valence with the self-assessment manikins (SAM), a nine-point visual analog scale (Hodes et al., 1985); and (4) arousal with SAM. All of the ratings were self-paced without time limit. We focused on the self--reported valence and arousal ratings for this analysis.

Before and after all movie-trailers were shown, we showed emotionally neutral movie excerpts (scenes from *Comment j'ai tué mon père* [2001], each 90s in length) for participants to return to baseline emotional state. In addition, after the fourth and the twelfth trailer, we showed a scrambled version of trailer of either M5 or M13 by random shuffling of pixels within each frame and replacing the original soundtrack with white noise of the same dynamic amplitude. (The order was counterbalanced across participants.) These were not used in the analysis.

2.2.2.2. Picture viewing. Before movie-trailer watching, in a separate session, we showed 156 pictures, each for 2s successively, drawn from the International Affective Picture System (IAPS) database (Lang et al., 2008), grouped in blocks of the six affective categories in the valence-by-arousal space: (1) positive valence/high arousal; (2) positive valence/low arousal; (3) negative valence/high arousal; (4) negative valence/low arousal; (5) neutral valence/medium arousal; and (6)

neutral valence/low arousal. They were matched for content of bodies, faces, animals and plants. (See supplementary material S2 for more description of the pictures.) Pictures from each of the first five categories were shown in two 12-picture (24s) blocks (i.e., 5 blocks × 2 runs), presented in random order within each run. In between blocks, a 4-picture (8s) block from the last category (neutral valence/low arousal) were shown to allow the emotional response to return to baseline. Of interest for the present study were the first four picture categories.[1]

2.2.3. FMRI acquisition and preprocessing

Subjects were scanned using a Siemens (Erlangen, Germany) Skyra 3 T MRI scanner, with a 32-channel head coil. Subjects could respond using a right-hand 4-button response device. A mirror mounted on the head coil ensured that participants could view the projector screen mounted at the back of the scanner. Functional data was acquired with a T2*-sensitized parallel imaging multi-echo sequence, with echo times (TE) at 9, 19.3, 30 and 40 ms. Thirty-four horizontal slices were acquired in ascending order (3.0 mm slice thickness, 0.5 mm slice gap, 3.5 × 3.5 mm in-plane resolution, 64 × 64 voxels per slice, flip angle 90°, total repetition time (TR) 2.07 s). The first 30 vol (before the start of each session's task) were used for echo-weighting. Prior to preprocessing, the four readouts acquired via the multi-echo sequence were combined and realigned by using standard procedures described by Poser et al. (2006). A T1-weighted image was acquired for anatomical reference (1.0 × 1.0 × 1.0 mm resolution, 192 sagittal slices, TE 3.03 ms, TR 2300 ms).

Data pre-processing was carried out using SPM12 (Statistical Parametric Mapping, Wellcome Department of Imaging Neuroscience, University College London, London, UK). Rigid-body transformations were applied to realign the volumes to the first echo of the first volume. Images were then corrected for differences in slice acquisition time. The anatomical image was co-registered with the mean functional image for each participant. Functional and anatomic images were then normalised to Montreal Neurological Institute (MNI) space. Finally, the normalised functional images were smoothed using a 3 mm full-width-at-half-maximum (FWHM) Gaussian kernel. The relatively small size of the smoothing kernel is informed by past literature that showed excessive smoothing reduced the sensitivity of pattern analysis (Misaki et al., 2013). Average white matter and out-of-brain signal were calculated by applying masks derived from the tissue probability maps available in SPM12 (Ashburner and Friston, 2005); specifically, the masks were created with probability > 0.75 as threshold on the respective probability maps. Linear detrending was then carried out using average white matter and out-of-brain signals as nuisance regressors; afterwards, voxel- and session-wise z-scoring were performed before analysis. For MVPA analysis, a time shift of 6s was also added throughout the analysis to account for hemodynamic response delay.

2.2.4. FMRI data analysis

Multivariate pattern analysis was carried out in Python with PyMVPA (Hanke et al., 2009) and Nilearn (Abraham et al., 2014) packages.

2.2.4.1. Selecting voxels responsive to affective picture categories. We first located voxels responsive to affective picture categories, i.e., those which showed differential activations between positive and negative valence pictures, and between high and low arousal pictures. This was done by univariate analysis with general linear models defined for each subject using a box-car function to model the picture blocks, with one regressor for each of the six picture categories detailed above. Two regressors of non-interest, average white matter signal and average out-of-brain signal, were entered together with a constant. Bidirectional *t*-contrasts were created for each subject comparing: (1) positive versus negative valence (collapsing over high and low arousal); (2) positive versus negative valence (high arousal only); (3) positive versus negative valence (low arousal only); and (4) high versus low arousal (collapsing over positive and negative valence). The first level maps were then entered into a random effects second-level analysis.

2.2.4.2. Inter-subject hyperalignment. Since we aimed to uncover shared neural representations of affective states across individuals, before doing so we considered ways to optimize brain alignment across individuals. To this end, we used functional response tuning as an extra step in addition to anatomical alignment (Haxby et al., 2014). Based on this particular approach called 'hyperalignment' (Haxby et al., 2011), individual subjects' voxel spaces were transformed into a common model space by comparing inter-subject neural responses to identical stimuli. For each subject, spatio-temporal responses to common stimuli (i.e., during movie-trailer watching) were treated as a single *n*-dimensional vector with *t* time points (where *n* is the number of voxels to be analysed). Using Procrustean transformation, these vectors were rotated to obtain optimal alignment (i.e., smallest Euclidean distances) among subjects with a three-pass procedure (Haxby et al., 2011): (1) begin with a randomly drawn pair of participants and obtain initial alignment, apply alignment iteratively by pairing the latest aligned average with a new participant, and obtain the average of all transformed vectors as new reference; (2) align all participants with the new reference, and generate the average of all transformed vectors as final reference; and (3) align all participants with the final reference. The procedure created, for each subject, a transformation matrix which could be applied on the corresponding voxel space. The transformation from individual voxel space to a common space has been shown to improve between-subject classification accuracy (Haxby et al., 2011).

Specifically, each participant's scan volumes of the 18 movie trailers (35 min 47 s in total) were re-sliced in the same order, resulting in a 1168-vol vector. (For the participant who did not watch the movie trailer M16 due to technical error, mean responses of that 63 vol from other participants were used instead.) Transformation matrices were estimated separately for each contiguous cluster, after which hyperalignment was applied to both picture viewing and movie-trailer watching data such that all participants' voxel responses were transformed to a common space.

2.2.4.3. Training and using valence and arousal classifiers based on shared patterns. After hyperalignment, two linear support vector machine (SVM) classifiers (one for valence and one for arousal) were trained on the picture viewing data. A leave-one-subject-out approach was used here, such that training data came from all-but-one participants. Then, the trained classifiers were used to decode the valence (positive or negative) and arousal (high or low) states of the remaining participant during picture viewing and movie-trailer watching. We decoded both the picture viewing data and the movie-trailer watching data volume-by-volume.

For the picture-viewing data, classification accuracy was defined as the percentage of volumes correctly categorized according to the valence and arousal categories of the pictures. Statistical significance of classification accuracy was evaluated based on the binomial distribution (Pereira and Botvinick, 2011).

For the movie-trailer data, the classifiers yielded for each movie-trailer two time series of classification probabilities (a continuous measure ranging 0–1), one for valence and one for arousal. The valence and arousal time series were then used in two ways. First, for each participant and each movie-trailer, we obtained the average classification probability for valence and arousal. We then entered the average probabilities

[1] It is possible to express the affective space in bipolar valence and arousal (Colibazzi et al., 2010) or unipolar positivity and negativity (Knutson et al., 2014), although there is insufficient evidence to determine which one better reflects the brain system (Lindquist et al., 2016). For the purpose of this study, we adopted the bipolar valence and arousal system since it did not involve transformation of the normed categories of the affective pictures nor transformation of the self-report ratings of the movie-trailers.

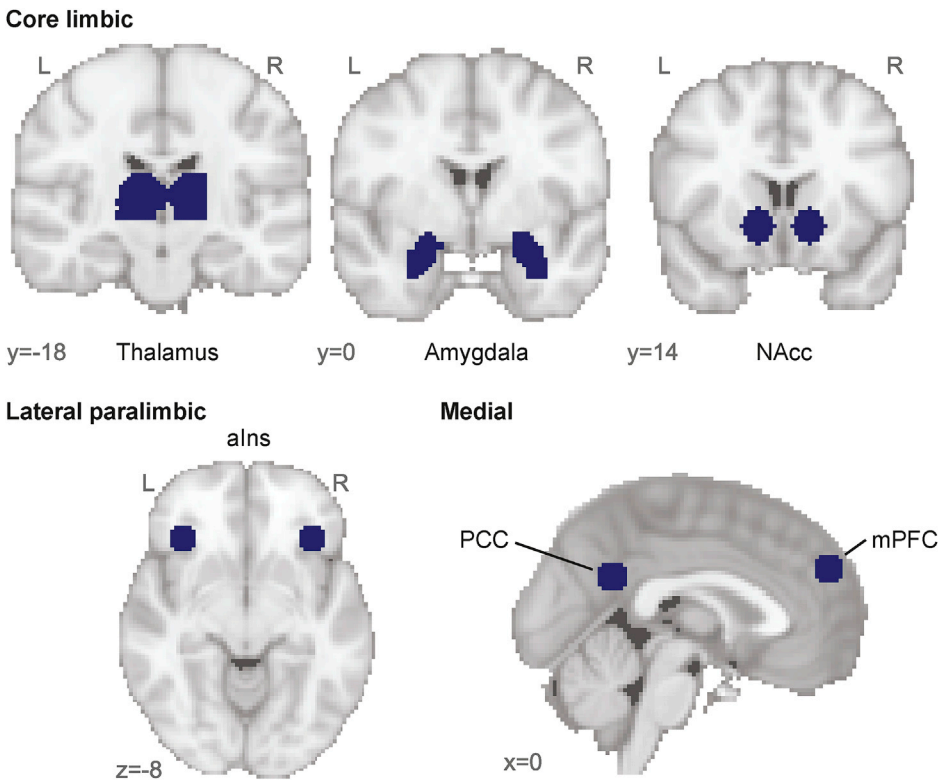

Fig. 2. Pre-defined regions of interest for mean activation analysis. NAcc = nucleus accumbens; aIns = anterior insula; PCC = posterior cingulate cortex; mPFC = medial prefrontal cortex.

to mixed effect regression models to examine the statistical relationship between self-report valence and arousal ratings and neural classification probabilities. Second, for each movie-trailer, we obtained the group average valence and arousal time series and compared them to the continuous self-report ratings obtained from a separate sample (details below).

2.2.4.4. Pre-defined regions of interest associated with affective processing. In addition to multivariate pattern analysis described above, we compared our findings with analysis based on mean activations at brain regions traditionally associated with affective processing (Knutson et al., 2014; Kober et al., 2008; Lindquist et al., 2012). We chose six regions of interest (ROIs) documented in extant literature for our analysis: amygdala, thalamus, nucleus accumbens (NAcc), anterior insula (aIns), posterior cingulate cortex (PCC), and medial prefrontal cortex (mPFC). The masks were defined from Automated Anatomical Labelling (amygdala and thalamus; Tzourio-Mazoyer et al., 2002), or created with 8 mm-radius spheres at published coordinates of NAcc [±11, 13, −6] and aIns [±35, 27, −9] (Knutson et al., 2014), and PCC [1, −54, 25] and mPFC [–2, 51, 29] (Kober et al., 2008). Using the category labels by Kober et al. (2008), we organized the six ROIs under three groups: core limbic, lateral paralimbic and medial (Fig. 2).

2.3. Online study

To further examine whether the valence and arousal time series from the neural data tracked the affective profiles of the videos, we recruited 51 U.S. participants through Amazon Mechanical Turk (29 women, mean age 38.0). They were paid at an hourly rate of $10. In this online study, they were asked to watch a random subset of the 18 movie trailers while providing continuous ratings of either valence (n = 24) or arousal (n = 27) for each of the videos. Continuous rating was achieved by using either a track pad or the scroll wheel of a mouse to control a vertical slider along the SAM visual analog scale (see Fig. 3 for a screenshot of the valence rating). Participants first practiced using the interface with an unrelated short video, then rated the movie-trailers in random order. The real-time movement of the slider was recorded and time-stamped with reference to the start of the video playback by the computer browser, and resampled at 1/TR (2.07s) Hz. On average, each movie trailer had 17 individual time-series in each affective dimension (valence and arousal; min = 14, max = 21). By averaging across participants, the valence and arousal self-report time series of the videos were obtained.

2.3.1. Examining correlations of affective time series between fMRI and online study

In the fMRI sample, we calculated the group average of valence and arousal classification probabilities at each time point of the videos, resulting in a single time series per movie-trailer. In the online sample, we calculated the group average at each corresponding time point of the movie-trailers. Concatenating the 18 movie-trailers, we obtained 1168-point time series of valence and arousal from each of the samples, and calculated the correlations between them. Due to the temporal autocorrelation structure inherent in the time series, we ascertained the statistical significance by permutation tests, in which the null correlation was calculated 10,000 times with the order of the 18 movie-trailers randomly shuffled at each time before concatenation (while keeping the time series *within* each movie-trailer intact). Empirical p-values were obtained by comparing the correlation with the null distribution.

2.4. Code accessibility and data availability

The data that support the findings of this study, including preprocessing and analysis scripts, are available in an open repository.[2]

[2] https://github.com/chanhangyee/affect_decoding.

Fig. 3. Example screenshot of the online rating task (screen was blurred due to copyright reasons).

Table 2
Clusters found in the thresholded maps (cluster size $k > 20$).

Cluster	Cluster size	t_{max}	X	y	z
Valence	126	5.58	−43	−69	−7
(negative > positive, high arousal)	59	6.18	6	−82	19
	47	4.74	44	−69	−10
Arousal	251	7.67	44	−68	0
(high > low)	219	7.97	−42	−75	1

3. Results

3.1. Self-report ratings of movie-trailers

The 18 movie-trailers evoked a range of valence and arousal responses, based on participants' self-report ratings (group-average valence [1–9, negative to positive] for movie-trailers: min = 4.60, median = 5.27, max = 6.35; group-average arousal [1–9, low to high]: min = 2.79, median = 4.77, max = 6.18; see supplementary material S3 for detailed descriptive statistics).

3.2. Brain areas responsive to affective picture categories

Significant results from *t*-contrasts, defined as $p < .001$ uncorrected and with minimum cluster extent of 20 voxels, revealed clusters with higher activation associated only with negative valence and high arousal. In addition, contrasts of valence yielded significant results only when high arousal pictures were used (Table 2). In both valence and arousal contrasts, we observed significant clusters at the lateral occipital cortex (LOC) (see Fig. 4A), while no significant voxels were found within the six pre-defined affective ROIs. (We further examined activations at those affective ROIs during picture viewing in a supplementary analysis S5.) We therefore targeted that area by creating a union of the two maps and applying light smoothing (3 mm FWHM) in order to obtain a LOC mask with two contiguous clusters (left: 340 voxels, right: 390 voxels) (Fig. 4B). Hyperalignment was conducted separately for each of the two clusters. The two corresponding transformation matrices were then applied on each participant's picture viewing and movie-trailer watching data.

3.3. Decoding affective categories of pictures using shared response classifiers

For each participant, linear SVM classifiers for valence and arousal were trained with picture viewing data pooled over the other 27 participants within the bilateral LOC. The classifiers were then applied on the remaining participant's picture viewing data to obtain valence and arousal classifications for each of the volumes under different picture categories. In the previous section, significant clusters for valence contrasts were only found in high arousal pictures; therefore, to optimize the valence classifiers, only neural responses to high arousal pictures (of positive and negative valence) were used (i.e., 1299 vol from the 28 subjects), whereas the arousal classifier was built with neural responses to all pictures (i.e., 2605 vol from the 28 subjects).

Using this leave-one-subject-out classification, the overall valence classification accuracy rate was 81.1% (chance: 50%), while the overall arousal classification accuracy rate was 67.2% (chance: 50%). Both figures were statistically significant ($p < 10^{-10}$) under the binomial test (Pereira and Botvinick, 2011). To ascertain whether the hyperalignment procedure boosted classification performance, we repeated the analysis without the procedure and found that classification accuracies were lower (valence: 70.4%, arousal: 62.3%) (Fig. 5A). The classification performance was comparable to the performance reported in a previous study with a similar design (Baucom et al., 2012). We further compared individual classification accuracies before and after hyperalignment. Wilcoxon signed-rank test showed that hyperalignment boosted performance (valence: Wilcoxon W = 10, $p < .0001$; arousal: W = 96, $p = .015$). After hyperalignment, classification accuracies for valence and arousal significantly exceeded the 50% chance level ($p < .05$) for the majority of participants (median accuracy for valence: 78.5%; arousal: 69.9%; Fig. 5B).

3.4. Decoding affective time series of movie-trailers using shared response classifiers

After training the valence and arousal classifiers based on affective pictures, they were applied on neural responses during movie-trailer watching in order to obtain valence and arousal time series, i.e., volume-by-volume classification probabilities for valence (negative-positive) and arousal (high-low).

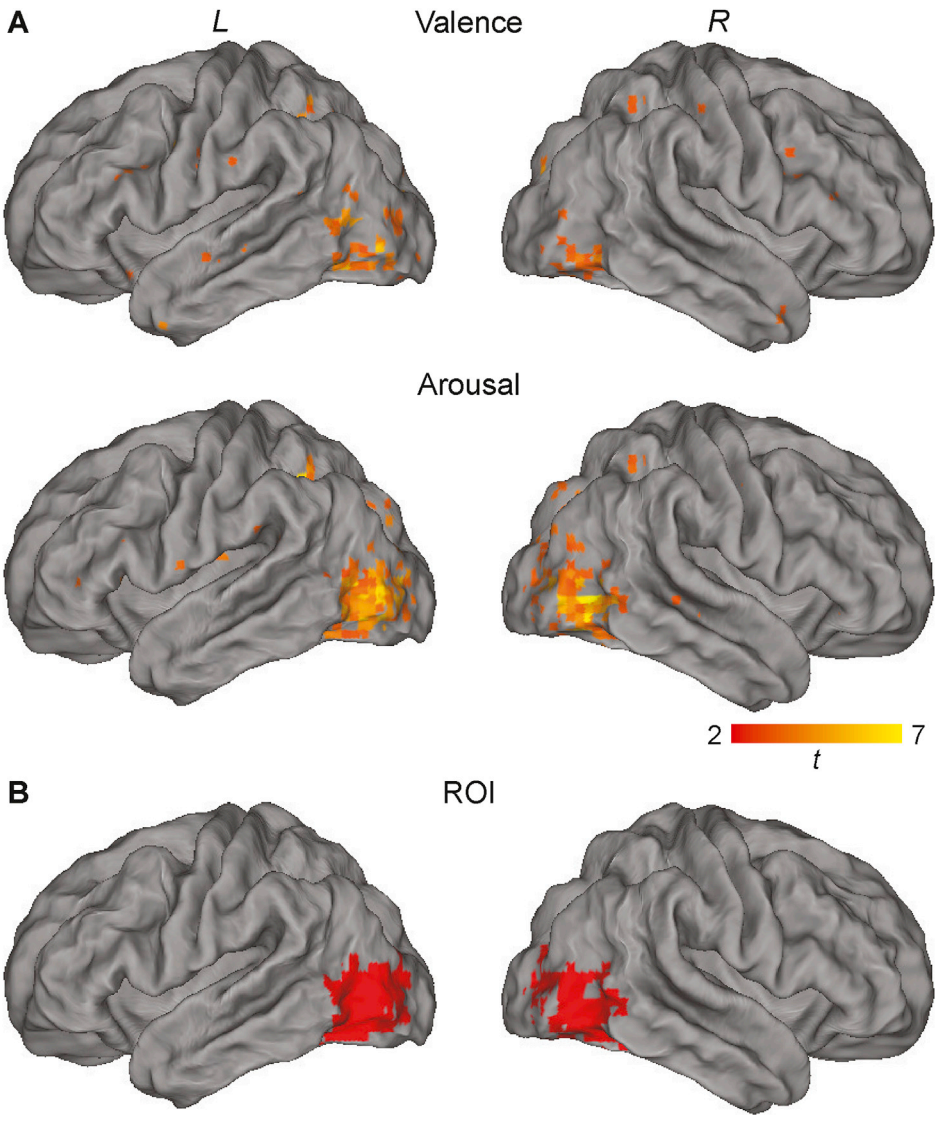

Fig. 4. (A) Statistical parametric maps thresholded at $p < .001$ uncorrected. (B) Bilateral regions of interest, defined by the union of voxels from the two statistical maps.

3.4.1. Validation with individual summary self-report ratings

Within a participant, the valence and arousal classification time series based on neural responses to each movie-trailer were averaged along the time axis. These average classification probabilities were then compared against the individual's own summary ratings of the corresponding movie-trailers (Fig. 6). The mean Fisher-transformed within-participant correlations between neural classification probability and self-report rating was 0.098 for valence (one-sample *t*-test against zero $t = 2.014$, $p = .054$), and 0.153 for arousal ($t = 3.671$, $p = .001$). We then examined the association between neural and self-report variables by using linear

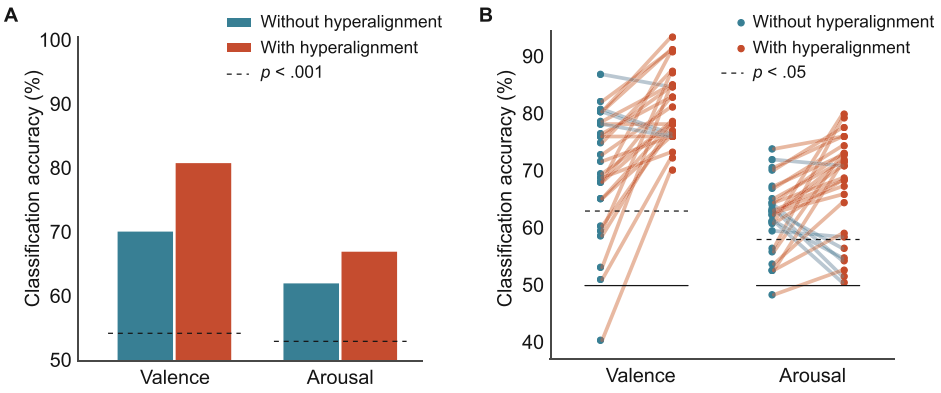

Fig. 5. (A) Overall and (B) individual leave-one-subject-out classification accuracies of affective picture categories with and without hyperalignment.

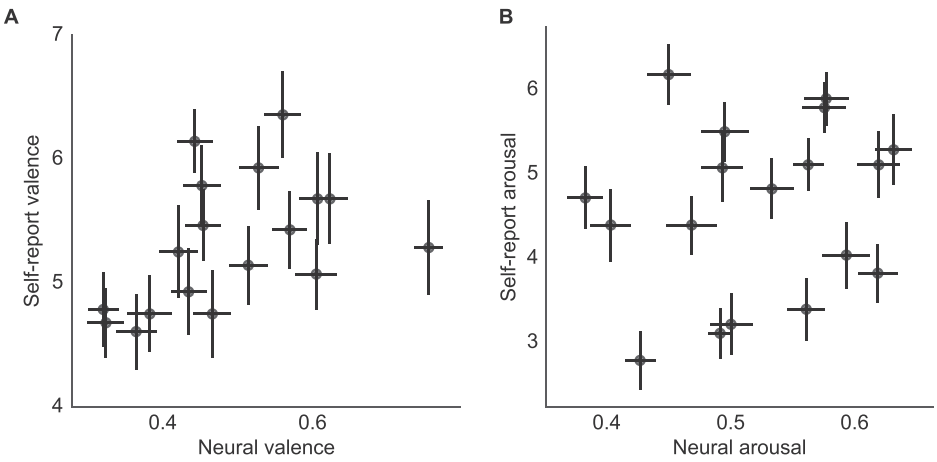

Fig. 6. Self-report ratings and neural classification probabilities of (A) valence and (B) arousal by movies (error bars represent standard error).

mixed-effects models with participant as random intercepts. Neural valence classification had a significant and positive effect on self-report valence rating, and neural arousal classification had a significant and positive effect on self-report arousal rating (Table 3A and B, models 1).

We further verified whether mean activations at pre-defined affective ROIs also predicted self-reported valence and arousal (Table 3A and B, models 2). Higher activation at medial ROIs was associated with more positive valence and higher arousal rating of the videos. Importantly, neural valence and arousal classifications remained significant predictors of self-report ratings after taking into account the ROI-based activations, and improved model fit as indicated by reduced Akaike's information criterion (AIC) (Table 3A and B, models 3). In order to verify that neural valence and arousal classifications based on LOC activity pattern were not driven by the mean activation level, we conducted a separate analysis where mean activation level at LOC was also entered as regressor. We found that LOC activation was not a significant predictor for self-reported arousal but was marginal for self-reported valence, and in either model the neural valence and arousal classification regressors remain significant after taking into account LOC activation (see supplementary analysis S6A). Mindful that activations at different ROIs could be correlated and thus posed an issue of collinearity, we repeated the analysis by entering regressors one at a time in a supplementary analysis (S6B) and found that the relationships between medial ROI activation and valence and arousal ratings remain significant. We also repeated the analysis without the hyperalignment procedure and found that with untransformed data, neural valence and arousal classifications were not significant predictors (supplementary analysis S6C). We further found that the time series of the valence and arousal classifications themselves contain movie-trailer-specific information that was shared across participants (supplementary analysis S7).

Table 3
Mixed effect regression models (from left to right, Model 1: neural classification only; Model 2: mean activations at pre-defined affective ROIs; Model 3: both) on self-report summary valence and arousal ratings, respectively, with participant as random intercepts. NAcc = nucleus accumbens; Lat. = lateral; aIns = anterior insula; mPFC = medial prefrontal cortex; PCC = posterior cingulate cortex.

A. Self-reported valence	Coef	SE	*p*	Coef	SE	*p*	Coef	SE	*p*
Neural classification									
Valence	0.733	0.439	.094				0.991	0.447	.027
Arousal	0.151	0.661	.819				−0.218	0.664	.742
Core limbic									
Amygdala				0.745	0.701	.288	0.941	0.710	.185
Thalamus				1.317	0.816	.107	1.341	0.825	.104
NAcc				−0.723	0.680	.287	−0.787	0.683	.250
Lat. paralimbic									
aIns				−1.442	0.843	.087	−1.192	0.855	.163
Medial									
mPFC				0.060	0.429	.888	−0.077	0.435	.860
PCC				1.814	0.563	.001	2.047	0.577	<.001
AIC		1989.5			1978.2			1977.2	
B. Self-reported arousal	**Coef**	**SE**	***p***	**Coef**	**SE**	***p***	**Coef**	**SE**	***p***
Neural classification									
Valence	−0.656	0.495	.185				−0.659	0.513	.199
Arousal	1.887	0.747	.011				1.945	0.762	.011
Core limbic									
Amygdala				−0.711	0.809	.379	−0.744	0.818	.363
Thalamus				0.396	0.941	.674	0.114	0.949	.904
NAcc				0.286	0.784	.715	0.294	0.785	.708
Lat. paralimbic									
aIns				0.706	0.971	.467	0.363	0.983	.712
Medial									
mPFC				−1.155	0.495	.020	−1.060	0.500	.034
PCC				−0.022	0.654	.973	−0.331	0.669	.621
AIC		2126.4			2135.1			2131.2	

3.4.2. Validation with aggregated continuous self-report ratings

For each movie-trailer, we obtained the average valence and arousal classification time series, and compared these with the continuous ratings by a separate online sample (see supplementary material S4 for the average continuous ratings for each movie-trailer). Time series correlation between neural classification of the fMRI sample and self-report ratings by the online sample was .134 (one-tailed empirical p value = .078) for valence and 0.222 (p = .003) for arousal, respectively (Fig. 7). These results suggest that the decoded time series of valence and arousal did indeed track the temporal profile of the movie-trailers, although the evidence was less strong for valence. (See supplementary analysis S8 for an analysis on activation time series extracted from affective ROIs.) As an illustration of how the classified affect time series might track narrative content, Fig. 8 shows the time series of one movie-trailer.

4. Discussion

In this study, we first identified that lateral occipital cortex contained information about valence and arousal evoked by IAPS pictures. We then showed that neural representations of affect within these brain areas were shared across participants, as evidenced by results from leave-one-subject-out classifications of the affective pictures. We also verified that functional alignment improved model performance. Using the classifiers trained on neural responses to these pictures, we then obtained moment-by-moment valence and arousal time series during movie-trailer watching. We showed that neural valence and arousal classifications were associated with self-report summary ratings of the individual. In addition, neural arousal classification time series tracked the continuous ratings of arousal of a separate sample, while weaker but consistent evidence was also found for valence classification.

The current findings provide supportive evidence that affective neural representations extracted from episodic affective states are similar to those of time-variable, and more naturalistic, emotional experiences. This is important, because it suggests that neural models constructed in a more controlled environment are also applicable in more realistic settings, where emotional states are changing rapidly. While behavioral studies have documented patterns of dynamic changes in emotions (Kuppens and Verduyn, 2017; Nielsen et al., 2008; Pe and Kuppens, 2012), the way in which these self-reported changes manifest at a neural level remained unknown. This study represents an effort into understanding dynamic emotional changes with neuroimaging (Spiers and Maguire, 2007).

It should be noted that, while the neural arousal classifications were strongly associated with self-reportratings of the affective experience during watching of the movie-trailers, this association was less strong for valence, even though the classification accuracy for valence within the affective pictures was relatively high (~80% after functional alignment). Thus, it seems that either our classifier for valence was less able to pick up subtle changes in valence during video watching, or that participants' self-reported evaluations for valence are somehow less accurate or more noisy. The current study does not allow for distinguishing between these two interpretations.

We identified the LOC as responsive to affective pictures, while the same GLM analysis did not reveal brain areas traditionally associated with affective processing, such as the amygdala, thalamus, aIns and NAcc. This result aligns with a previous study using a similar set of stimuli (IAPS pictures) and paradigm (block design) (Sabatinelli et al., 2006), which identified LOC activation, but not other affective regions, during viewing of affective pictures. More broadly speaking, LOC (among other affective brain regions) was identified in extant studies using affective visual stimuli (Aldhafeeri et al., 2012; Baucom et al., 2012; Bush et al., 2018b; Gerber et al., 2008; Kassam et al., 2013; Klasen et al., 2011; Nielen et al., 2009; Sabatinelli et al., 2005) but not in other modalities (Chikazoe et al., 2014; Colibazzi et al., 2010; Posner et al., 2009). Additional analysis of pre-defined affective ROIs (core limbic, lateral paralimbic and medial areas) showed that while negative valence-high

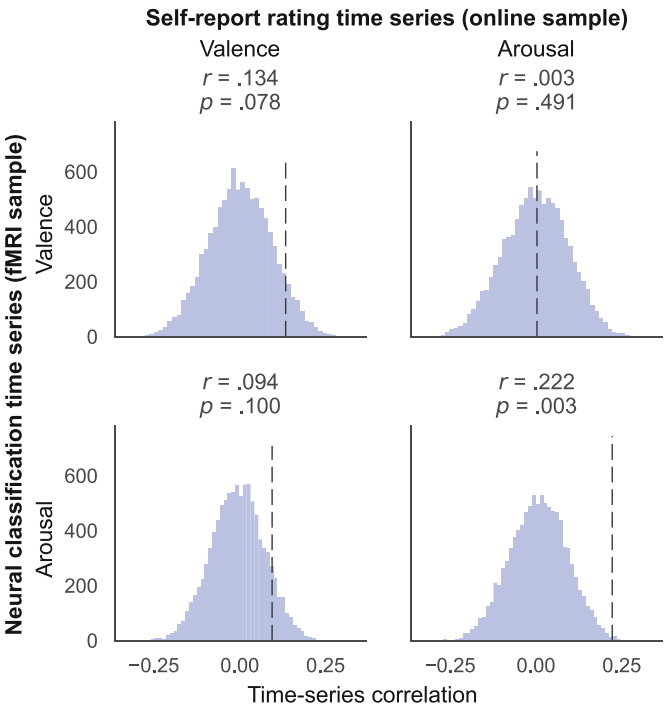

Fig. 7. Correlations between time-series of self-report ratings from online sample and neural scores from fMRI sample, with null distribution plot of the correlations from 10,000 permutations. The empirical p-values reported here are one-tailed.

arousal pictures drove up activation intensity in many of these areas, voxels in those areas did not survive statistical thresholding in a whole-brain analysis.

We speculate there may be several reasons for not finding these areas traditionally associated with affect. First, our preprocessing procedure deviated from previous studies. In order to retain spatial pattern information (Misaki et al., 2013), we chose a relatively light smoothing kernel (3 mm FWHM) as opposed to more commonly used kernels (6–8 mm). This decreased the sensitivity of locating significant voxels in second-level analysis, although for the purpose of this study the aim of the procedure was not localization but feature selection. We also restricted our feature selection to contiguous clusters ($k > 20$) since we planned for cluster-wise hyperalignment in later stages. (In additional analyses, we tested anatomically defined ROIs at amygdala and thalamus for neural pattern extractions but failed to obtain significant results.) Second, we adopted a block design for picture viewing, i.e., presenting successive pictures of the same affective categories within a block, and treat the block as having a homogenous affective state. This was in contrast with other studies where pictures were presented individually, separated by inter-trial intervals (Bush et al., 2018b, 2018a). It is likely that during a picture block (12 same-category pictures in 24s) participants experienced affective adaptation (Wilson and Gilbert, 2008) and thus attenuated their affective response towards the end of the block. Lastly, there is a distinction between emotion perception and experience (Wager et al., 2008). In the current study, participants were asked to observe affective stimuli (perception) instead of generate the feelings themselves (experience). A meta-analysis has shown that experience triggered a stronger reaction in the frontal cortex and subcortical structures, while perception accentuated activation in the sensory cortex (Wager et al., 2008). In sum, given the evidence that there exist both modality-general and modality-specific neural coding of affective states (Miskovic and Anderson, 2018), we speculate that the affective neural patterns we extracted in this study were specific to visual stimuli and may not be transferable to other modalities.

We also found that inter-subject functional alignment increased

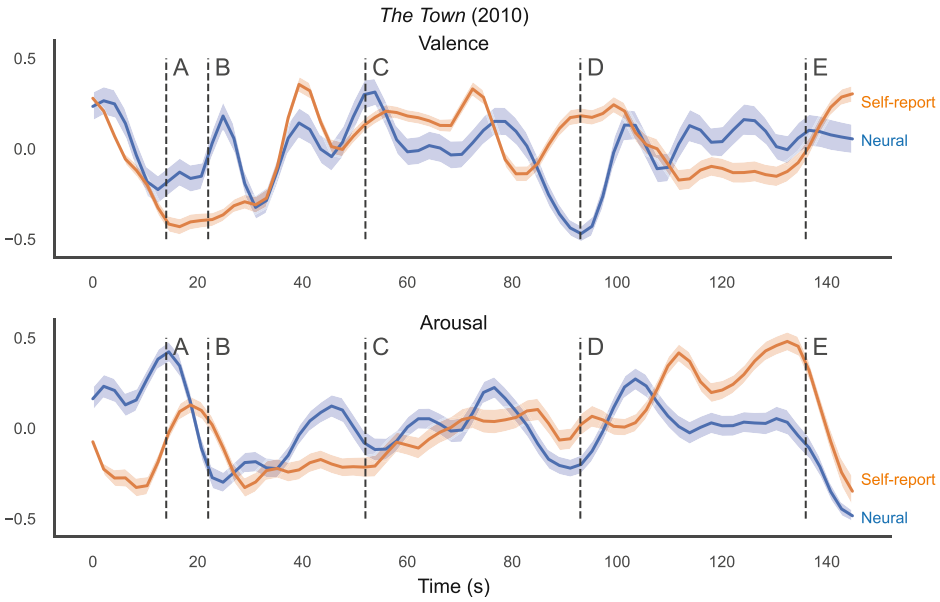

Fig. 8. This figure shows group-averaged neural (from fMRI sample) and self-report (from online sample) valence and arousal for the trailer of *The Town* (2010). For illustrative purposes, the series are rescaled to range from −0.5 to 0.5 and gaussian-smoothed with 1 TR sigma along the time axis. Shaded regions represent standard error. Notable moments in the video: A – Bank robbery and hostage taking; B – Studio vanity card; C – First romantic meeting between male and female leads (Ben Affleck and Rebecca Hall); D – Reveal that the male lead turns out to be kidnapper of the female lead; E – Title card and credits.

performance in classification accuracy. This confirms previous results on its use for improving cross-participant classification (Haxby et al., 2011). (In a supplementary analysis S6C, we repeated the procedure without hyperalignment and found that decoded valence and arousal during movie-trailer watching did not track self-report ratings.) However, most of the past applications were limited to visual and auditory processing circuits to uncover a common representational space for sensory percepts (Guntupalli, 2013; Haxby et al., 2011; Nishimoto and Nishida, 2016). More recent studies are expanding towards a whole-brain approach in a search of common space for higher-order processes (Guntupalli et al., 2016). While we applied hyperalignment closer to the original approach by Haxby et al. (2011), our findings suggest that this technique is also beneficial to decoding other mental processes such as emotions.

Some limitations of this study should be mentioned together with pointers for future research direction. First, the fact that only neural responses at LOC were used to predict affective content might suggest that the neural representations might be tied to specific low-level visual features instead of affective content. Although recent studies have shown that affective information could be uncovered based on occipital cortex activity (Bush et al., 2018a; Kragel et al., 2019), given the fact that we only used LOC and no other brain networks, further research should focus on establishing whether modality-general representations are capable of tracking dynamic change of specific emotions as well. Second, we did not obtain continuous self-report ratings of valence and arousal within the fMRI sample (Hutcherson et al., 2005; Nummenmaa et al., 2012; Raz et al., 2016b; Young et al., 2017), thus we could not compare neural and self-report time series within the same participant in the fMRI sample. Future research should verify if the current approach could also be used to study the individual variability in affective responses. Lastly, we adopted a dimensional model of core affect (Russell, 1980) in our study and the block design for presenting affective picture (Sabatinelli et al., 2006) collapsed the valence and arousal dimensions into binary categories. This prevented us from building a model that generates valence and arousal predictions on a continuous scale. Future research should therefore look into ways to: (a) develop models with continuous valence and arousal predictions, and (b) use categorical emotions for classification, which may reveal finer-grained affective changes across time that could not be captured by valence and arousal alone, such as co-existence of multiple emotions (Larsen and McGraw, 2011).

In summary, this study explored the feasibility of using neural representations from brief, stable affective episodes (picture viewing) to decode extended, dynamic affective sequences in a naturalistic experience (movie-trailers watching). Our findings demonstrate the potential of using pretrained neural representations to decode affective responses to naturalistic stimuli of an independent sample. Moreover, by extending affective neuroscience research from discrete states to dynamic sequences, this study highlights the possibility of augmenting behavioral studies of temporal changes of affect, which rely mostly on self-report measures, with neuroimaging data.

Declaration of competing interest

The authors declare no competing financial interests.

CRediT authorship contribution statement

Hang-Yee Chan: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing. **Ale Smidts:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision. **Vincent C. Schoots:** Conceptualization, Methodology, Investigation, Writing - original draft. **Alan G. Sanfey:** Conceptualization, Methodology, Supervision. **Maarten A.S. Boksem:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision.

Acknowledgement

The authors gratefully acknowledge financial support from the Erasmus Research Institute of Management (ERIM) and the Dutch national e-infrastructure with the support of SURF Cooperative.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116618.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Front. Neuroinf. 8 https://doi.org/10.3389/fninf.2014.00014.

Adolphs, R., Nummenmaa, L., Todorov, A., Haxby, J.V., 2016. Data-driven approaches in the investigation of social perception. Philos. Trans. R. Soc. B Biol. Sci. https://doi.org/10.1098/rstb.2015.0367.

Aldhafeeri, F.M., Mackenzie, I., Kay, T., Alghamdi, J., Sluming, V., 2012. Regional brain responses to pleasant and unpleasant IAPS pictures: different networks. Neurosci. Lett. https://doi.org/10.1016/j.neulet.2012.01.064.
Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018.
Baucom, L.B., Wedell, D.H., Wang, J., Blitzer, D.N., Shinkareva, S.V., 2012. Decoding the neural representation of affective states. Neuroimage 59, 718–727. https://doi.org/10.1016/j.neuroimage.2011.07.037.
Becker, G.M., Degroot, M.H., Marschak, J., 1964. Measuring utility by a single-response sequential method. Behav. Sci. 9, 226–232. https://doi.org/10.1002/bs.3830090304.
Brans, K., Verduyn, P., 2014. Intensity and duration of negative emotions: comparing the role of appraisals and regulation strategies. PloS One. https://doi.org/10.1371/journal.pone.0092410.
Bush, K.A., Gardner, J., Privratsky, A., Chung, M.-H., James, G.A., Kilts, C.D., 2018a. Brain states that encode perceived emotion are reproducible but their classification accuracy is stimulus-dependent. Front. Hum. Neurosci. 12, 262. https://doi.org/10.3389/fnhum.2018.00262.
Bush, K.A., Privratsky, A., Gardner, J., Zielinski, M.J., Kilts, C.D., 2018b. Common functional brain states encode both perceived emotion and the psychophysiological response to affective stimuli. Sci. Rep. 8, 15444. https://doi.org/10.1038/s41598-018-33621-6.
Chan, H.-Y., Smidts, A., Schoots, V.C., Dietvorst, R.C., Boksem, M.A.S., 2019. Neural similarity at temporal lobe and cerebellum predicts out-of-sample preference and recall for video stimuli. Neuroimage 197, 391–401. https://doi.org/10.1016/j.neuroimage.2019.04.076.
Chapin, H., Jantzen, K., Scott Kelso, J.A., Steinberg, F., Large, E., 2010. Dynamic emotional and neural responses to music depend on performance expression and listener experience. PloS One 5, e13812. https://doi.org/10.1371/journal.pone.0013812.
Chikazoe, J., Lee, D.H., Kriegeskorte, N., Anderson, A.K., 2014. Population coding of affect across stimuli, modalities and individuals. Nat. Neurosci. 17, 1114–1122. https://doi.org/10.1038/nn.3749.
Colibazzi, T., Posner, J., Wang, Z., Gorman, D., Gerber, A., Yu, S., Zhu, H., Kangarlu, A., Duan, Y., Russell, J.A., Peterson, B.S., 2010. Neural systems subserving valence and arousal during the experience of induced emotions. Emotion 10, 377–389. https://doi.org/10.1037/a0018484.
Conroy, B.R., Singer, B.D., Guntupalli, J.S., Ramadge, P.J., Haxby, J.V., 2013. Inter-subject alignment of human cortical anatomy using functional connectivity. Neuroimage. https://doi.org/10.1016/j.neuroimage.2013.05.009.
Frijda, N.H., 2009. Emotions, individual differences and time course: Reflections. Cognit. Emot. https://doi.org/10.1080/02699930903093276.
Gerber, A.J., Posner, J., Gorman, D., Colibazzi, T., Yu, S., Wang, Z., Kangarlu, A., Zhu, H., Russell, J., Peterson, B.S., 2008. An affective circumplex model of neural systems subserving valence, arousal, and cognitive overlay during the appraisal of emotional faces. Neuropsychologia 46, 2129–2139. https://doi.org/10.1016/j.neuropsychologia.2008.02.032.
Gholipour, A., Kehtarnavaz, N., Briggs, R., Devous, M., Gopinath, K., 2007. Brain functional localization: a survey of image registration techniques. IEEE Trans. Med. Imag. https://doi.org/10.1109/TMI.2007.892508.
Goldin, P.R., Hutcherson, C.A.C., Ochsner, K.N., Glover, G.H., Gabrieli, J.D.E., Gross, J.J., 2005. The neural bases of amusement and sadness: a comparison of block contrast and subject-specific emotion intensity regression approaches. Neuroimage. https://doi.org/10.1016/j.neuroimage.2005.03.018.
Guntupalli, J.S., 2013. Whole brain hyperalignment: inter-subject hyperalignment of local representational spaces. Dartmouth College. https://doi.org/10.1349/ddlp.964.
Guntupalli, J.S., Hanke, M., Halchenko, Y.O., Connolly, A.C., Ramadge, P.J., Haxby, J.V., 2016. A model of representational spaces in human cortex. Cerebr. Cortex 26, 2919–2934. https://doi.org/10.1093/cercor/bhw068.
Hanke, M., Halchenko, Y.O., Sederberg, P.B., Olivetti, E., Fründ, I., Rieger, J.W., Herrmann, C.S., Haxby, J.V., Hanson, S.J., Pollmann, S., 2009. PyMVPA: a unifying approach to the analysis of neuroscientific data. Front. Neuroinf. 3, 3. https://doi.org/10.3389/neuro.11.003.2009.
Haxby, J.V., 2012. Multivariate pattern analysis of fMRI: the early beginnings. Neuroimage. https://doi.org/10.1016/j.neuroimage.2012.03.016.
Haxby, J.V., Connolly, A.C., Guntupalli, J.S., 2014. Decoding neural representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. 37, 435–456. https://doi.org/10.1146/annurev-neuro-062012-170325.
Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 80. https://doi.org/10.1126/science.1063736.
Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., Ramadge, P.J., 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72, 404–416. https://doi.org/10.1016/j.neuron.2011.08.026.
Hodes, R.L., Cook, E.W., Lang, P.J., 1985. Individual differences in autonomic response: conditioned association or conditioned fear? Psychophysiology 22, 545–560. https://doi.org/10.1111/j.1469-8986.1985.tb01649.x.
Hutcherson, C.A., Goldin, P.R., Ochsner, K.N., Gabrieli, J.D., Barrett, L.F., Gross, J.J., 2005. Attention and emotion: does rating emotion alter neural responses to amusing and sad films? Neuroimage 27, 656–668. https://doi.org/10.1016/j.neuroimage.2005.04.028.
Kassam, K.S., Markey, A.R., Cherkassky, V.L., Loewenstein, G., Just, M.A., 2013. Identifying emotions on the basis of neural activation. PloS One 8, e66032. https://doi.org/10.1371/journal.pone.0066032.
Kim, J., Schultz, J., Rohe, T., Wallraven, C., Lee, S.-W., Bulthoff, H.H., 2015. Abstract representations of associated emotions in the human brain. J. Neurosci. 35, 5655–5663. https://doi.org/10.1523/JNEUROSCI.4059-14.2015.
Klasen, M., Kenworthy, C.A., Mathiak, K.A., Kircher, T.T.J., Mathiak, K., 2011. Supramodal representation of emotions. J. Neurosci. 31, 13635–13643. https://doi.org/10.1523/JNEUROSCI.2833-11.2011.
Knutson, B., Katovich, K., Suri, G., 2014. Inferring affect from fMRI data. Trends Cognit. Sci. 18, 422–428. https://doi.org/10.1016/j.tics.2014.04.006.
Kober, H., Barrett, L.F., Joseph, J., Bliss-Moreau, E., Lindquist, K.A., Wager, T.D., 2008. Functional grouping and cortical–subcortical interactions in emotion: a meta-analysis of neuroimaging studies. Neuroimage 42, 998–1031. https://doi.org/10.1016/j.neuroimage.2008.03.059.
Kragel, P.A., LaBar, K.S., 2015. Multivariate neural biomarkers of emotional states are categorically distinct. Soc. Cognit. Affect Neurosci. 10, 1437–1448. https://doi.org/10.1093/scan/nsv032.
Kragel, P.A., Reddan, M.C., LaBar, K.S., Wager, T.D., 2019. Emotion schemas are embedded in the human visual system. Sci. Adv. 5, eaaw4358. https://doi.org/10.1126/sciadv.aaw4358.
Kuppens, P., Verduyn, P., 2017. Emotion dynamics. Curr. Opin. Psychol. 17, 22–26. https://doi.org/10.1016/j.copsyc.2017.06.004.
Lang, P.J., Bradley, M.M., Cuthbert, B.N., 2008. International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual, Technical Report A-8. Center for Research in Psychophysiology, University of Florida, Gainesville, FL.
Larsen, J.T., McGraw, A.P., 2011. Further evidence for mixed emotions. J. Pers. Soc. Psychol. 100, 1095–1110. https://doi.org/10.1037/a0021846.
Lindquist, K.A., Satpute, A.B., Wager, T.D., Weber, J., Barrett, L.F., 2016. The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. Cerebr. Cortex 26, 1910–1922. https://doi.org/10.1093/cercor/bhv001.
Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., Barrett, L.F., 2012. The brain basis of emotion: a meta-analytic review. Behav. Brain Sci. 35, 121–143. https://doi.org/10.1017/S0140525X11000446.
Mesquita, B., 2010. Emoting: a contextualized process. In: Mesquita, B., Barrett, L.F., Smith, E.R. (Eds.), The Mind in Context. Guilford Press, pp. 83–104.
Misaki, M., Luh, W.M., Bandettini, P.A., 2013. The effect of spatial smoothing on fMRI decoding of columnar-level organization with linear support vector machine. J. Neurosci. Methods. https://doi.org/10.1016/j.jneumeth.2012.11.004.
Miskovic, V., Anderson, A.K., 2018. Modality general and modality specific coding of hedonic valence. Curr. Opin. Behav. Sci. 19, 91–97. https://doi.org/10.1016/j.cobeha.2017.12.012.
Nielen, M.M.A., Heslenfeld, D.J., Heinen, K., Van Strien, J.W., Witter, M.P., Jonker, C., Veltman, D.J., 2009. Distinct brain systems underlie the processing of valence and arousal of affective pictures. Brain Cognit. 71, 387–396. https://doi.org/10.1016/j.bandc.2009.05.007.
Nielsen, L., Knutson, B., Carstensen, L.L., 2008. Affect dynamics, affective forecasting, and aging. Emotion 8, 318–330. https://doi.org/10.1037/1528-3542.8.3.318.
Nishimoto, S., Nishida, S., 2016. Lining up brains via a common representational space. Trends Cognit. Sci. 20, 565–567. https://doi.org/10.1016/j.tics.2016.06.001.
Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cognit. Sci. 10, 424–430. https://doi.org/10.1016/j.tics.2006.07.005.
Nummenmaa, L., Glerean, E., Viinikainen, M., Jaaskelainen, I.P., Hari, R., Sams, M., Jääskeläinen, I.P., Hari, R., Sams, M., 2012. Emotions promote social interaction by synchronizing brain activity across individuals. Proc. Natl. Acad. Sci. Unit. States Am. 109, 9599–9604. https://doi.org/10.1073/pnas.1206095109.
Nummenmaa, L., Saarimäki, H., 2019. Emotions as discrete patterns of systemic activity. Neurosci. Lett. 693, 3–8. https://doi.org/10.1016/j.neulet.2017.07.012.
Panksepp, J., 1982. Toward a general psychobiological theory of emotions. Behav. Brain Sci. 5, 407–422. https://doi.org/10.1017/S0140525X00012759.
Pe, M.L., Kuppens, P., 2012. The dynamic interplay between emotions in daily life: augmentation, blunting, and the role of appraisal overlap. Emotion 12, 1320–1328. https://doi.org/10.1037/a0028262.
Peelen, M.V., Atkinson, A.P., Vuilleumier, P., 2010. Supramodal representations of perceived emotions in the human brain. J. Neurosci. 30, 10127–10134. https://doi.org/10.1523/JNEUROSCI.2161-10.2010.
Pereira, F., Botvinick, M., 2011. Information mapping with pattern classifiers: a comparative study. Neuroimage 56, 476–496. https://doi.org/10.1016/j.neuroimage.2010.05.026.
Poldrack, R.A., 2011. Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. Neuron 72, 692–697. https://doi.org/10.1016/j.neuron.2011.11.001.
Poser, B.A., Versluis, M.J., Hoogduin, J.M., Norris, D.G., 2006. BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. Magn. Reson. Med. 55, 1227–1235. https://doi.org/10.1002/mrm.20900.
Posner, J., Russell, J.A., Gerber, A., Gorman, D., Colibazzi, T., Yu, S., Wang, Z., Kangarlu, A., Zhu, H., Peterson, B.S., 2009. The neurophysiological bases of emotion: an fMRI study of the affective circumplex using emotion-denoting words. Hum. Brain Mapp. 30, 883–895. https://doi.org/10.1002/hbm.20553.
Raz, G., Shpigelman, L., Jacob, Y., Gonen, T., Benjamini, Y., Hendler, T., 2016a. Psychophysiological whole-brain network clustering based on connectivity dynamics analysis in naturalistic conditions. Hum. Brain Mapp. 37, 4654–4672. https://doi.org/10.1002/hbm.23335.
Raz, G., Touroutoglou, A., Wilson-Mendenhall, C., Gilam, G., Lin, T., Gonen, T., Jacob, Y., Atzil, S., Admon, R., Bleich-Cohen, M., Maron-Katz, A., Hendler, T., Barrett, L.F., 2016b. Functional connectivity dynamics during film viewing reveal common

networks for different emotional experiences. Cognit. Affect Behav. Neurosci. 16, 709–723. https://doi.org/10.3758/s13415-016-0425-4.
Raz, G., Winetraub, Y., Jacob, Y., Kinreich, S., Maron-Katz, A., Shaham, G., Podlipsky, I., Gilam, G., Soreq, E., Hendler, T., 2012. Portraying emotions at their unfolding: a multilayered approach for probing dynamics of neural networks. Neuroimage. https://doi.org/10.1016/j.neuroimage.2011.12.084.
Russell, J.A., 1980. A circumplex model of affect. J. Pers. Soc. Psychol. 39, 1161–1178. https://doi.org/10.1037/h0077714.
Saarimäki, H., Ejtehadian, L.F., Glerean, E., Jääskeläinen, I.P., Vuilleumier, P., Sams, M., Nummenmaa, L., 2018. Distributed affective space represents multiple emotion categories across the human brain. Soc. Cognit. Affect Neurosci. 13, 471–482. https://doi.org/10.1093/scan/nsy018.
Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I.P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., Nummenmaa, L., 2016. Discrete neural signatures of basic emotions. Cerebr. Cortex 26, 2563–2573. https://doi.org/10.1093/cercor/bhv086.
Sabatinelli, D., Bradley, M.M., Fitzsimmons, J.R., Lang, P.J., 2005. Parallel amygdala and inferotemporal activation reflect emotional intensity and fear relevance. Neuroimage 24, 1265–1270. https://doi.org/10.1016/j.neuroimage.2004.12.015.
Sabatinelli, D., Lang, P.J., Keil, A., Bradley, M.M., 2006. Emotional perception: correlation of functional MRI and event-related potentials. Cerebr. Cortex 17, 1085–1091. https://doi.org/10.1093/cercor/bhl017.
Sabuncu, M.R., Singer, B.D., Conroy, B., Bryan, R.E., Ramadge, P.J., Haxby, J.V., 2010. Function-based intersubject alignment of human cortical anatomy. Cerebr. Cortex. https://doi.org/10.1093/cercor/bhp085.
Scherer, K.R., 2009. The dynamic architecture of emotion: evidence for the component process model. Cognit. Emot. https://doi.org/10.1080/02699930902928969.
Skerry, A.E., Saxe, R., 2014. A common neural code for perceived and inferred emotion. J. Neurosci. 34, 15997–16008. https://doi.org/10.1523/JNEUROSCI.1676-14.2014.
Spiers, H.J., Maguire, E.A., 2007. Decoding human brain activity during real-world experiences. Trends Cognit. Sci. 11, 356–365. https://doi.org/10.1016/j.tics.2007.06.002.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289. https://doi.org/10.1006/nimg.2001.0978.
Wager, T.D., Barrett, L.F., Bliss-moreau, E., Lindquist, K.A., Duncan, S., Kober, H., Joseph, J., Davidson, M.L., Mize, J., 2008. The neuroimaging of emotion. In: Lewis, M., Haviland-Jones, J.M., Barrett, L.F. (Eds.), Handbook of Emotions. Guilford Press, New York, pp. 249–271. https://doi.org/10.2307/2076468.
Wilson, T.D., Gilbert, D.T., 2008. Explaining away: a model of affective adaptation. Perspect. Psychol. Sci. 3, 370–386. https://doi.org/10.1111/j.1745-6924.2008.00085.x.
Young, C.B., Raz, G., Everaerd, D., Beckmann, C.F., Tendolkar, I., Hendler, T., Fernández, G., Hermans, E.J., 2017. Dynamic shifts in large-scale brain network balance as a function of arousal. J. Neurosci. https://doi.org/10.1523/jneurosci.1759-16.2017.