Emotionotopy: Gradients encode emotion dimensions in right temporo-parietal territories

Giada Lettieri[1+], Giacomo Handjaras[1+], Emiliano Ricciardi[1], Andrea Leo[1], Paolo Papale[1], Monica Betta[1], Pietro Pietrini[1#], Luca Cecchetti[1#*]

[1]MoMiLab Research Unit, IMT School for Advanced Studies Lucca, Lucca, Italy
[+]Denotes equal first author contribution
[#]Denotes equal senior author contribution

[*]Corresponding author:
Luca Cecchetti
IMT School for Advanced Studies Lucca
Piazza San Francesco, 19, 55100 Lucca - Italy
Email: luca.cecchetti@imtlucca.it

Abstract

Humans use emotions to decipher complex cascades of internal events. However, which mechanisms link descriptions of affective states to brain activity is unclear, with evidence supporting either local or distributed processing. A biologically favorable alternative is provided by the notion of gradient, which postulates the isomorphism between functional representations of stimulus features and cortical distance. Here, we used fMRI activity evoked by an emotionally charged movie and continuous ratings of the perceived emotion intensity to reveal the topographic organization of affective states. We found that three orthogonal and spatially overlapping gradients encode the *polarity, complexity* and *intensity* of emotional experiences in right temporo-parietal territories. The spatial arrangement of these gradients allows the brain to map a variety of affective states within a single patch of cortex. As this organization resembles how sensory regions represent psychophysical properties (e.g., retinotopy), we propose *emotionotopy* as a principle of emotion coding.

Introduction

Emotions promptly translate inner experiences into specific patterns of interpretable behaviors. To understand our own affective states and to infer those of others represent crucial aspects both when humans directly relate to each other and when they simply observe social interactions. Through years, the relevance of such abilities motivated the quest for models that optimally associate behavioral responses to emotional experiences.

In this regard, seminal works pointed toward the existence of discrete basic emotions, characterized by distinctive and culturally stable facial expressions[1], patterns of autonomous nervous system activity[2,3] and bodily sensations[4]. Happiness, surprise, fear, sadness, anger and disgust represent the most frequently identified set of basic emotions[5], though alternative models propose that other emotions, such as pride or contempt, should be included for their social and biological relevance[6]. To prove the neurobiological validity of these models, neuroscientists investigated whether basic emotions are consistently associated with specific patterns of brain responses across subjects. Findings show that activity in amygdala, medial prefrontal, anterior cingulate, insular, middle/inferior frontal, and posterior superior temporal cortex, is associated to the perceived intensity of emotions and supports their recognition[7-10]. However, this perspective has been challenged by other studies, which failed to demonstrate significant associations between single emotions and activity within distinct cortical areas or networks[11-13].

An alternative theory proposes that behavioral, physiological and subjective characteristics of emotions would be more appropriately described along a number of continuous cardinal dimensions[14,15], generally one governing pleasure versus displeasure (i.e., valence) and another one the strength of the experience (i.e., arousal). While these two dimensions have been reliably and consistently described, alternative models propose that additional dimensions, such as dominance or unpredictability, are needed to adequately explain affective states[16,17]. Neuroimaging studies also demonstrated that stimuli varying in valence and arousal elicit specific and reliable brain

responses[18,19], which have been recently employed to decode emotional experiences[20]. Activity recorded in insula, amygdala, ventral striatum, anterior cingulate, ventromedial prefrontal and posterior territories of the superior temporal cortex, is associated to transitions between positive and negative valence and fluctuations in arousal[21,22].

Of note, other than in the ventromedial prefrontal regions, studies using either discrete emotion categories[9,10,12] or emotion dimensions[22-25] have shown responses in the posterior portion of the superior temporal cortex, extending to temporo-parietal territories. Furthermore, these temporo-parietal regions are fundamental for social cognition, as they support empathic processing[26,27] and the attribution of intentions, beliefs and emotions to others[28,29].

However, despite this large body of evidence, it remains to be determined whether emotional experiences are better described through discrete basic emotions or emotion dimensions. Moreover, regardless of the adopted model, it is still debated how emotion features are spatially encoded in the brain[8,11,13,30-32]. As a matter of fact, while findings support the role of distinct regions[7], others indicate the recruitment of distributed networks in relation to specific affective states[33].

An alternative and biologically favorable perspective may be provided by the notion of gradient. Gradients have been proven a fundamental organizing principle through which the brain efficiently represents and integrates stimuli coming from the external world. For instance, the location of a stimulus in the visual field is easily described through two orthogonal spatially overlapping gradients in primary visual cortex: rostrocaudal for eccentricity and dorsoventral for polar angle[34]. Thus, using functional magnetic resonance imaging (fMRI) and retinotopic mapping, one can easily predict the location of a stimulus in the visual field considering the spatial arrangement of recruited voxels with respect to these orthogonal gradients. Crucially, recent investigations revealed that gradients support the representation of higher-order information as well[35-37], with features as animacy or numerosity being topographically arranged onto the cortical mantle[35,38,39].

Following this view, we hypothesize that affective states are encoded in a gradient-like manner in the human brain. Specifically, different affective states would be mapped onto the cortical mantle through spatially overlapping gradients, which would code either the intensity of discrete emotions (e.g., weak to strong sadness) or, alternatively, the smooth transitions along cardinal dimensions (e.g., negative to positive valence). In either case, the pattern of brain activity could be interpreted according to emotion gradients to predict the subjective affective state.

Here, we tested this hypothesis using moment-by-moment ratings of the perceived intensity of emotions elicited by an emotionally charged movie. To unveil cortical regions involved in emotion processing, behavioral ratings were used as predictors of fMRI activity in an independent sample of subjects exposed to the same movie. The correspondence between functional characteristics and the relative spatial arrangement of distinct patches of cortex was then tested to reveal the existence of emotion gradients.

Results

Emotion Ratings

A group of Italian native speakers continuously rated the perceived intensity of six basic emotions[5] (i.e., *happiness*, *surprise*, *fear*, *sadness*, *anger* and *disgust*) while watching an edited version of the Forrest Gump movie (R. Zemeckis, Paramount Pictures, 1994). We first assessed how much each basic emotion contributed to the behavioral ratings and found that *happiness* and *sadness* explained 28% and 36% of the total variance, respectively. Altogether, *fear* (18%), *surprise* (8%), *anger* (7%), and *disgust* (3%) explained the remaining one-third of the total variance. We also evaluated the agreement in ratings of the six basic emotions (Figure 1A), and found that *happiness* (Spearman's $\rho = 0.476 \pm 0.102$, range 0.202 - 0.717), *fear* ($\rho = 0.522 \pm 0.134$, range 0.243 - 0.793), *sadness* ($\rho = 0.509 \pm 0.084$, range 0.253 - 0.670), and *anger* ($\rho = 0.390 \pm 0.072$, range 0.199 - 0.627) were consistent across all the subjects, whereas *surprise* ($\rho = 0.236 \pm 0.099$, range 0.010 - 0.436) and *disgust* ($\rho = 0.269 \pm 0.115$, range 0.010 - 0.549) were not. Nonetheless, ratings for these latter emotions were on average significantly different from a null distribution of randomly assigned emotion ratings (p-value < 0.05).

To reveal emotion dimensions, we averaged across subjects the ratings of the six basic emotions, measured their collinearity (Figure 1B) and performed Principal Component (PC) analysis (Figure 1C). The first component reflected a measure of *polarity* (PC_1: 45% explained variance) as positive and negative emotions demonstrated opposite loadings. The second component was interpreted as a measure of *complexity* (PC_2: 24% explained variance) of the perceived affective state, ranging from a positive pole where *happiness* and *sadness* together denoted inner conflict and ambivalence, to a negative pole mainly representing fearful events. The third component was a measure of *intensity* (PC_3: 16% explained variance), since all the six basic emotions showed positive loadings (Figure 1C). Altogether, the first three components explained approximately 85% of the total variance. We further assessed the stability of the PCs and found that

only these first three components (*polarity*: ρ = 0.610 ± 0.089, range 0.384 - 0.757; *complexity*: ρ = 0.453 ± 0.089, range 0.227 - 0.645; *intensity*: ρ = 0.431 ± 0.071, range 0.258 - 0.606), hereinafter *emotion dimensions*, were consistent across all the subjects (Figure 1D). The fourth PC described movie segments during which participants experienced *anger* and *disgust* at the same time (PC_4: 8% explained variance, ρ = 0.329 ± 0.128, range -0.003 - 0.529), whereas the fifth PC was mainly related to *surprise* (PC_5: 6% explained variance, ρ = 0.214 ± 0.090, range 0.028 - 0.397). Notably, these two PCs were not consistent across all the subjects, even though their scores were on average significantly different from a null distribution (p-value < 0.05). Scores of the sixth PC were not significantly consistent across subjects (PC_6: 1% explained variance, p-value > 0.05).

--- Figure 1 ---

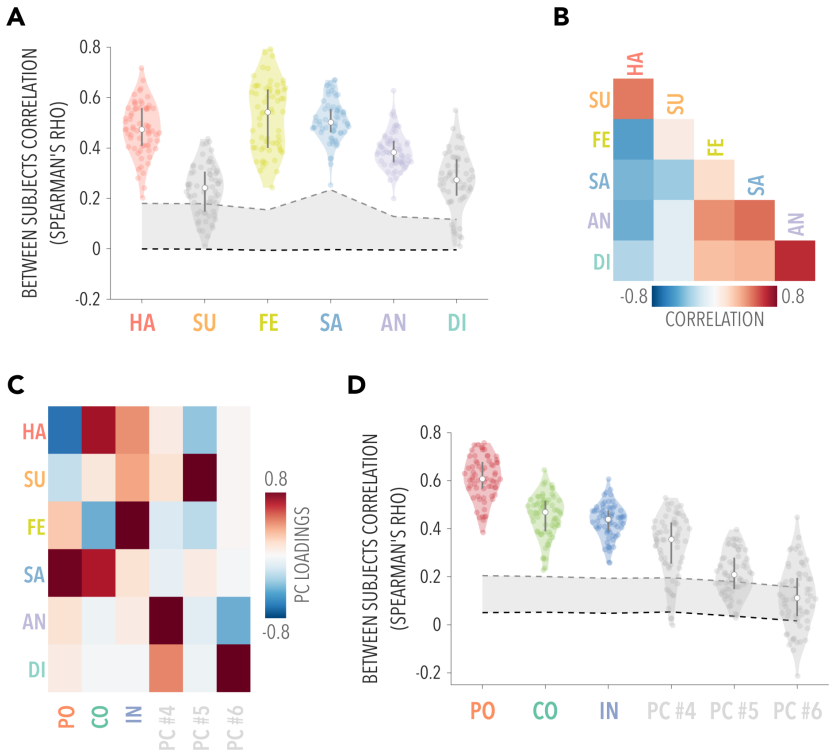

Figure 1 Emotion ratings - **A**. Violin plots show the agreement (Spearman's ρ coefficient) of the six basic emotions across subjects. Gray shaded area represents the null distribution of behavioral ratings and dashed lines the mean and 95th percentile of the null distribution. **B**. Correlation matrix showing Spearman's ρ values for pairings of basic emotions. **C**. Principal Component Analysis: loadings of the six principal components. Explained variance was 45% for *polarity*, 24% for *complexity* and 16% for *intensity*. **D**. Violin plots show the agreement (Spearman's ρ coefficient) of the six principal components across subjects. Gray shaded area represents the null distribution of behavioral ratings and dashed lines the mean and 95th percentile of the null distribution. HA = *Happiness*, SU = *Surprise*, FE = *Fear*, SA = *Sadness*, AN = *Anger*, DI = *Disgust*, PC = Principal component, PO = *Polarity*, CO = *Complexity*, IN = *Intensity*.

Richness of the Reported Emotional Experience

In our behavioral experiment, participants were allowed to report the perceived intensity of more than one emotion at a time. Thus, the final number of elicited emotional states might be greater than the original six emotion categories. To measure the richness of affective states reported by our participants, we performed dimensionality reduction and clustering analyses on group-averaged behavioral ratings. Results revealed the existence of 15 distinct affective states throughout the movie (Figure 2). Among these, some states were characterized by a single basic emotion, whereas others by a peculiar mixture of them. Combinations of distinct emotions likely expressed secondary affective states, as *ambivalence* (i.e., cluster *j* depicting movie scenes in which *happiness* and *sadness* are simultaneously experienced) or *resentment* (i.e., cluster *i* representing movie segments in which a mixture of *sadness*, *anger* and *disgust* is perceived). Of note, this evidence is supported also by single-subject reports, in which the 38% (SE: ±2.3%) of timepoints were associated to a single emotion, the 29% (SE: ±3.5%) to two basic emotions and the 6% (SE: ±1.4%) to the concurrent experience of three distinct emotions. Altogether, these results show that the Forrest Gump movie evoked complex and multifaceted experiences, which cannot be reduced to the original six categories.

--Figure 2--

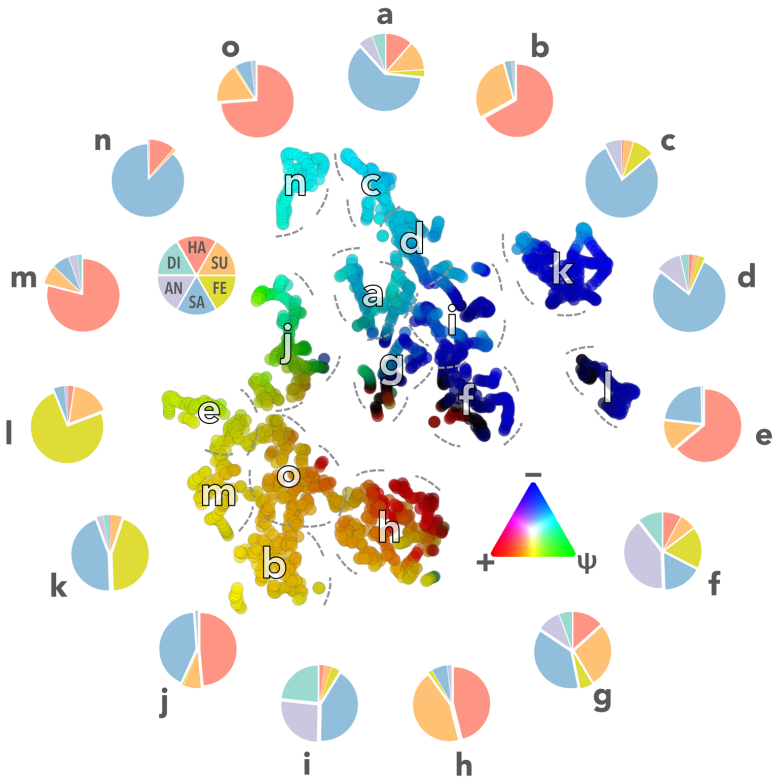

Figure 2 Richness of the emotional experience - Results of the dimensionality reduction (t-SNE) and clustering analyses (k-means) on the group-averaged behavioral ratings showing the existence of 15 distinct affective states throughout the movie. Each element represents a specific timepoint in the movie and the distance between elements depends on the statistical similarity of emotion ratings. Element color reflects the scores of the *polarity* and *complexity* dimensions: positive (+) and negative (-) events (i.e., polarity) are associated respectively to the red and blue channels, whereas *complexity* (Ψ) scores modulate the green channel. Pie charts show the relative contribution of the six basic emotions to each of the 15 identified clusters. HA = *Happiness*, SU = *Surprise*, FE = *Fear*, SA = *Sadness*, AN = *Anger*, DI = *Disgust*.

Brain Regions encoding Emotion Ratings

Emotion ratings obtained from the behavioral experiment were used as predictors of brain activity in independent subjects exposed to the same movie (*studyforrest* project; http://studyforrest.org[40]). The model significantly explained activity in right inferior frontal gyrus (IFG), rostral middle frontal gyrus (rMFG), medial superior frontal gyrus (mSFG), occipitotemporal sulcus (OTS), precentral sulcus (preCS), posterior part of the superior temporal sulcus/temporoparietal junction (pSTS/TPJ), middle occipital gyrus (MOG) and posterior middle temporal gyrus (pMTG). We also observed significant results in the left supramarginal gyrus (SMG) and pMTG ($q < 0.01$ FDR corrected and cluster size > 10; Figure 3A and Supplementary Table 1). Notably, the peak of association between emotion ratings and brain activity was located in

the right pSTS/TPJ, an important region for social cognition[12,22,26,28,29] (R^2 = 0.07 ± SE = 0.009; Center of Gravity – CoG: x = 61, y = -40, z = 19; noise ceiling lower bound 0.13, upper bound 0.23; Figure 3B and Supplementary Figure 1). The peak of association was also located in proximity (11 mm displacement) of the *reverse inference* peak for the term "TPJ" (CoG: x = 58, y = -50, z = 16) as reported in the NeuroSynth database (neurosynth.org; Figure 3B).

--- Figure 3 ---

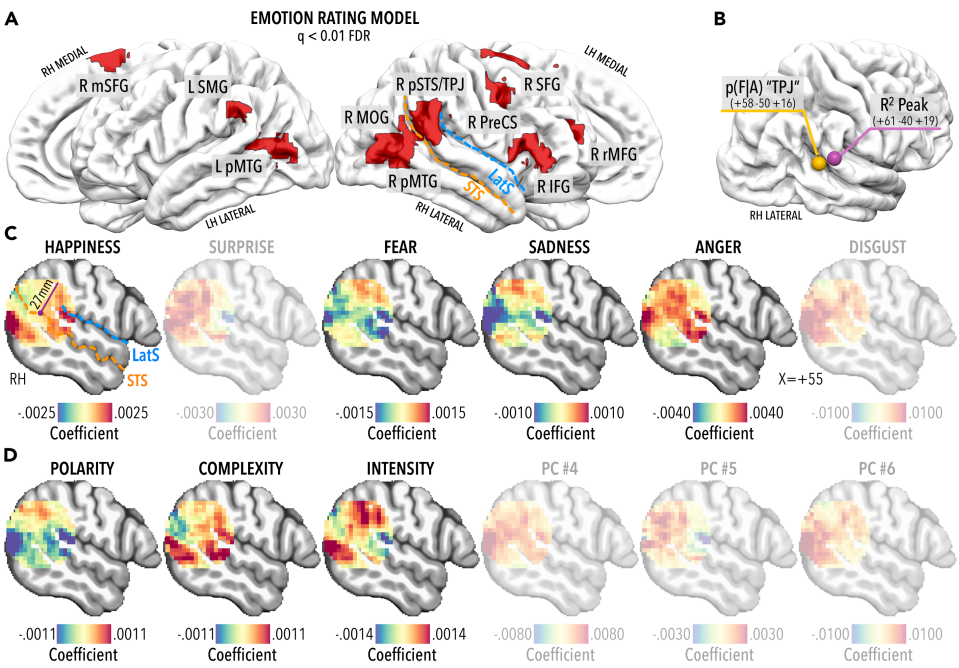

Figure 3 Encoding of emotion ratings - A. Brain regions encoding emotion ratings corrected for multiple comparisons through the False Discovery Rate method (q < 0.01). **B**. Peak of association between emotion ratings and brain activity (purple sphere) and *reverse inference peak* for the term "*TPJ*" as reported in the NeuroSynth database (yellow sphere). Coordinates represent the center of gravity in MNI152 space. **C**. β coefficients associated to basic emotions in a spherical region of interest (27mm radius) located at the *reverse inference peak* for the term "*TPJ*". Maps for emotions not consistent across all the subjects (i.e., *surprise* and *disgust*) are faded (see the *Agreement across subjects of the six basic emotions section*). **D**. β coefficients associated to *emotion dimensions* in a spherical region of interest (27mm radius) located at the *reverse inference peak* for the term "*TPJ*". Maps for components not consistent across all the subjects (i.e., PC_4, PC_5 and PC_6) are faded (see the *Agreement across subjects of the Emotion Dimensions section*). IFG = Inferior Frontal Gyrus, rMFG = rostral Middle Frontal Gyrus, mSFG = Medial Superior Frontal Gyrus, preCS = Precentral Sulcus, pSTS/TPJ = posterior part of the Superior Temporal Sulcus/Temporoparietal Junction, MOG = Middle Occipital Gyrus, pMTG = posterior Middle Temporal Gyrus, SMG = Supramarginal Gyrus, LatS = Lateral Sulcus, STS = Superior Temporal Sulcus.

Emotion Gradients in Right Temporo-Parietal Territories

We tested the existence of either basic emotion or *emotion dimension* gradients in a spherical region of interest located at the *reverse inference* peak for the term "TPJ". This analysis was

conducted on behavioral ratings consistent across all the subjects: *happiness*, *sadness*, *fear* and *anger* for basic emotions and *polarity*, *complexity* and *intensity* for *emotion dimensions*.

Using β coefficients obtained from the encoding analysis, we observed that, within right TPJ, voxels appeared to encode *happiness* in an anterior to posterior arrangement, *fear* and *sadness* in an inferior to superior manner, while *anger* showed a patchier organization (Figure 3C). With respect to *emotion dimensions*, voxels seemed to encode *polarity* and *intensity* in a more inferior to superior fashion, whereas *complexity* in a more posterior to anterior direction (Figure 3D).

To prove the existence and precisely characterize the orientation of these gradients, we tested the association between physical distance and functional characteristics of right TPJ voxels (Supplementary Figure 2). Results demonstrated that within a 15 mm radius sphere, the relative spatial arrangement and functional features of right TPJ were significantly and maximally correlated, either considering the basic emotion model ($\rho = 0.352$, p-value = 0.004, 95% Confidence Interval - CI: 0.346 to 0.357) or the *emotion dimension* one ($\rho = 0.399$, p-value < 0.001, 95% CI: 0.393 to 0.404; for alternative definitions of the right TPJ region see Supplementary Table 2).

Crucially, when focusing on each *emotion dimension*, results revealed the existence of three orthogonal and spatially overlapping gradients: *polarity* ($\rho = 0.241$, p-value = 0.041, 95% CI: 0.235 to 0.247), *complexity* ($\rho = 0.271$, p-value = 0.013, 95% CI: 0.265 to 0.277) and *intensity* ($\rho = 0.229$, p-value = 0.049, 95% CI: 0.223 to 0.235; Figure 4 and Supplementary Table 3). On the contrary, *happiness* ($\rho = 0.275$, p-value = 0.013, 95% CI: 0.269 to 0.281), but not other basic emotions (*fear*: $\rho = 0.197$, p-value = 0.091; *sadness*: $\rho = 0.182$, p-value = 0.160; *anger*: $\rho = 0.141$, p-value = 0.379; Supplementary Table 3), retained a gradient-like organization. Of note, the peculiar arrangement of group-level *emotion dimension* gradients (Figure 4) was also identified using single-subject fMRI data (Supplementary Figure 3 and Supplementary Table 4).

As any orthogonal rotation applied to the *emotion dimensions* would result into different gradients, we measured to what extent rotated solutions explained the topography of right TPJ. Therefore, we tested the correspondence between anatomical distance and the fitting of ~70,000 rotated versions of *polarity*, *complexity* and *intensity* (see Supplementary Materials for a comprehensive description). Results showed that the original unrotated *emotion dimensions* represented the optimal solution to explain the gradient-like organization of right temporo-parietal territories (Supplementary Figure 4).

Further, we performed a data-driven searchlight analysis to test whether right TPJ was the only region significantly encoding all the three *emotion dimension* gradients (please refer to Supplementary Materials for details). Results obtained from the meta-analytic definition of right TPJ were confirmed using this alternative approach ($q < 0.05$ FDR corrected and cluster size > 10; CoG: x = 58, y = -53, z = 21; Supplementary Figure 5), as no other region encoded the combination of *polarity*, *complexity* and *intensity* in a topographic manner.

Moreover, we conducted three separate searchlight analyses to characterize the spatial arrangement of single *emotion dimension* gradients (see Supplementary Material). *Polarity*, *complexity* and *intensity* maps revealed specific topographies: regions as the right preCS represented the three *emotion dimensions* in distinct - yet adjoining - subregions, whereas the right OTS encoded overlapping gradients of *complexity* and *intensit*y (Supplementary Figure 6).

When we explored whether the left hemisphere homologue of TPJ (CoG: x = -59, y = -56, z = 19) showed a similar gradient-like organization, we did not find significant associations between spatial and functional characteristics either for the basic emotion model ($\rho = 0.208$, p-value = 0.356) or the *emotion dimension* one ($\rho = 0.251$, p-value = 0.144; Supplementary Table 2). Specifically, neither any of the *emotion dimensions* (*polarity*: $\rho = 0.132$, p-value = 0.354; *complexity*: $\rho = 0.157$, p-value = 0.222; *intensity*: $\rho = 0.149$, p-value = 0.257) nor any of the basic emotions showed a gradient-like organization in left TPJ (*happiness*: $\rho = 0.158$, p-value = 0.216;

fear: ρ = 0.142, p-value = 0.293; *sadness*: ρ = 0.156, p-value = 0.213; *anger*: ρ = 0.073, p-value = 0.733; Supplementary Table 3).

Lastly, as spatial smoothness of functional data and cortical folding may affect the estimation of gradients, we performed additional analyses considering the unfiltered version of group-average brain activity and obtaining a measure of the anatomical distance respectful of cortical topology. Results showed that the topographic arrangement of *emotion dimensions* in right temporo-parietal territories was not affected by smoothing (Supplementary Figure 7) and respected the cortical folding (*polarity*: ρ = 0.248, p-value = 0.026, CI: 0.238-0.257; *complexity*: ρ = 0.314, p-value = 0.001, CI: 0.304-0.323; *intensity*: ρ = 0.249, p-value = 0.013, CI: 0.239-0.258). For details about this procedure and a comprehensive description of the results please refer to Supplementary Materials.

To summarize, *polarity*, *complexity* and *intensity* dimensions were highly consistent across individuals, explained the majority of the variance in behavioral ratings (85%) and were mapped in a gradient-like manner in right (but not left) TPJ. *Happiness* (28% of the total variance in behavioral ratings) was the only basic emotion to be consistent across subjects and to be represented in right TPJ. Importantly, though, *happiness* and *complexity* demonstrated high similarity both in behavioral ratings (ρ = 0.552) and in brain activity patterns (ρ = 0.878). Taken together, these pieces of evidence indicate the existence of *emotion dimension* gradients in right temporo-parietal territories, rather than the presence of discrete emotion topographies.

--- Figure 4 ---

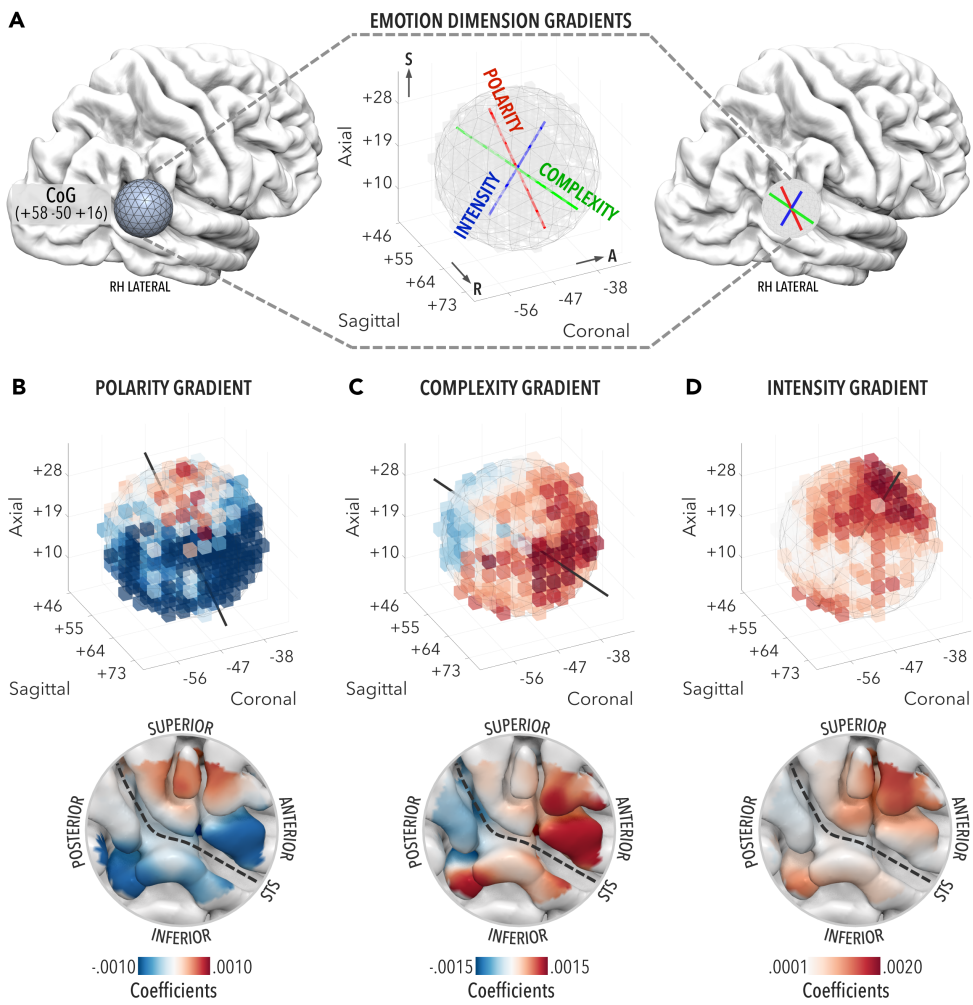

Figure 4 Emotion gradients in right TPJ - A. We revealed three orthogonal and spatially overlapping *emotion dimension* gradients (*polarity*, *complexity* and *intensity*) within a region of interest located at the *reverse inference peak* for the term "*TPJ*" (15mm radius sphere). Symmetry axis of the region of interest represents the main direction of the three gradients **B**. β coefficients of the *polarity* dimension are mapped through a gradient with an inferior to superior direction. **C**. β coefficients of the *complexity* dimension are mapped through a gradient with a posterior to anterior direction. **D**. β coefficients of the *intensity* dimension are mapped through a gradient with an inferior to superior direction. For single-subjects results please refer to Supplementary Figure 6. Lowermost row depicts the arrangement of the emotion dimension gradients in surface space. CoG = Center of Gravity, STS = Superior Temporal Sulcus.

Do Emotion Dimension Gradients Simply Encode Portrayed Emotions?

In movie watching, actions and dialogues are not usually directed toward the observer and, the reported subjective experience is very likely influenced by character emotions, intentions and beliefs. Therefore, we tested whether the gradient-like organization of right TPJ can be explained simply considering emotions of movie characters. To this aim, we took advantage of publicly available tagging data of Forrest Gump[41], in which participants indicated the portrayed emotion of each character and whether such an emotion was directed toward the character itself (*self-directed*;

e.g., Forrest feeling sad) or toward another one (*other-directed*; e.g., Forrest feeling happy for Jenny). As in Labs and colleagues[41], these reports constituted two *third-person emotion attribution* descriptions, which we used as models of the attribution of affective states to others (please refer to Supplementary Materials for details).

On average, *subjective ratings* shared the 11.4% ± 8.6% of the variance with the *self-directed emotion attribution model* and the 35.3% ± 16.1% with the *other-directed model*, indicating that the subjective emotional experience could be inferred from portrayed emotions only in part. Moreover, in line with previous studies highlighting the role of right TPJ in the attribution of mental states to others[28,29,42], the *other-directed emotion attribution* model significantly explained activity in this region (Supplementary Figure 8). However, none of the first six components obtained from the *other-directed emotion attribution model* (i.e., 87% of the explained variance) retained a topographic organization in right TPJ (Supplementary Table 5). In addition, we used canonical correlation analysis to transform the *other-directed model* into the space defined by subjective emotion ratings and tested whether starting from a third-person complex description of portrayed emotions, one can fully reconstruct the brain topography of *emotion dimensions*. Noteworthy, only the first aligned component was mapped in a topographic manner within right TPJ (*reconstructed polarity*: $\rho = 0.221$, p-value = 0.036; *reconstructed complexity*: $\rho = 0.150$, p-value = 0.384; *reconstructed intensity*: $\rho = 0.207$, p-value = 0.092). This finding suggests that the information coded in the *other-directed emotion attribution model* is not sufficient to fully reconstruct the topography of the subjective emotional experience.

Characterization of Emotion Dimension Gradients

To detail how right TPJ gradients encode perceived affective states, we have reconstructed fMRI activity for movie segments connoted by either positive or negative *polarity*, as well as higher or lower *complexity* and *intensity*. The orientation of the three *emotion dimension* gradients was

represented by the symmetry axis of our region of interest. Specifically, for *polarity* events connoted by positive emotions increased activity in ventrorostral territories, lying close to the superior temporal sulcus, whilst highly negative events augmented hemodynamic activity in dorsocaudal portions of right TPJ, extending to the posterior banks of Jensen sulcus (Figure 5A and 3D).

Events connoted by higher *complexity* (e.g., concurrent presence of *happiness* and *sadness*) were associated to signal increments in rostrolateral territories of right TPJ, whereas those rated as having lower *complexity* (e.g., fearful events) increased hemodynamic activity in its caudal and medial part, encompassing the ascending ramus of the superior temporal sulcus (Figure 5B and 3D). Higher levels of *intensity* were related to increased activity in rostrodorsal and ventrocaudal territories, reaching the ascending ramus of the lateral sulcus and posterior portions of the middle temporal gyrus, respectively. On the contrary, low-*intensity* events augmented hemodynamic activity in a central belt region of right TPJ, located along the superior temporal sulcus (Figure 5C and 3D). Noteworthy, the orthogonal arrangement of *polarity* and *complexity* and the fact that *intensity* was represented both superiorly and inferiorly to the superior temporal sulcus determined that the variety of emotional states elicited by the Forrest Gump movie (see Figure 2) could be mapped within a single patch of cortex.

--- Figure 5 ---

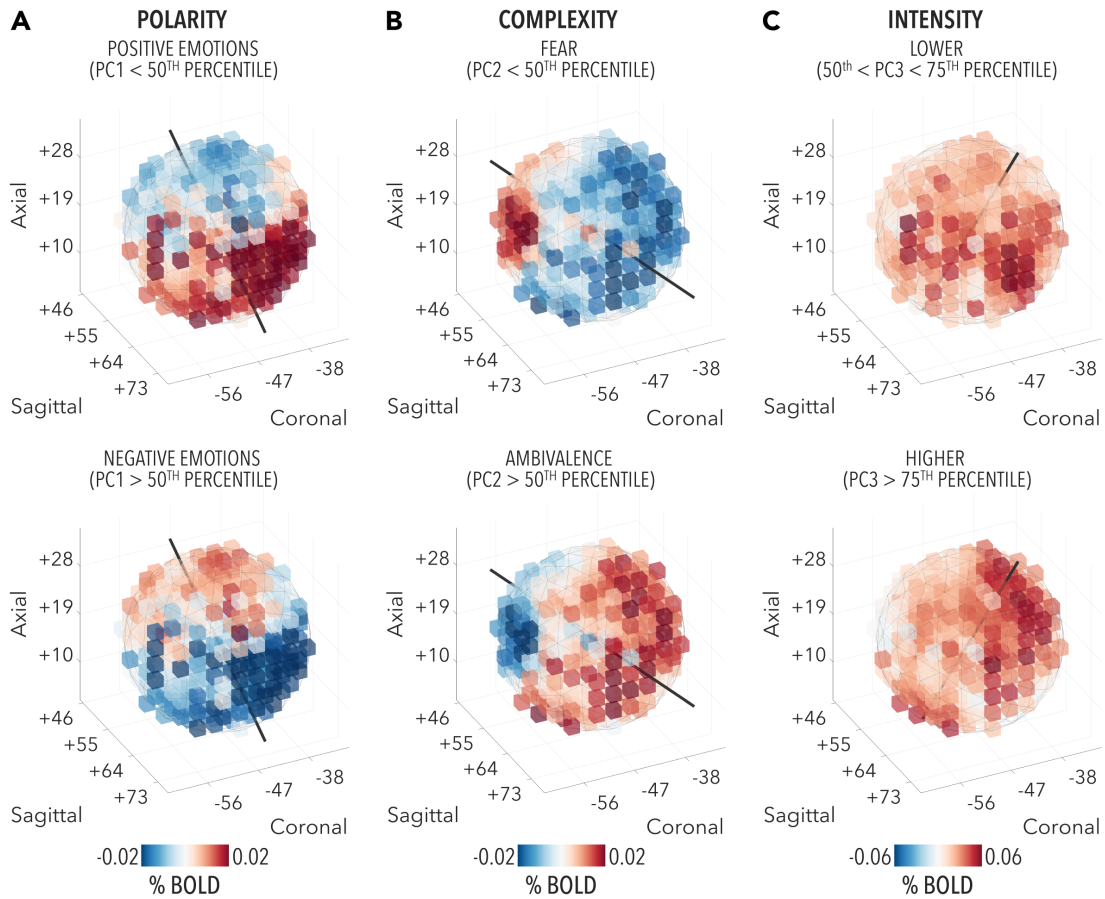

Figure 5 Characterization of Emotion Dimension Gradients in right TPJ - A. Average hemodynamic activity in right TPJ related to the scores below and above the 50th percentile for *polarity*. **B**. Average hemodynamic activity in right TPJ related to the scores below and above the 50th percentile for *complexity*. **C**. Since *intensity* is not bipolar as the other two components (i.e., scores ranged from ~0 to positive values only), for this dimension we mapped the average TPJ activity above the 75th percentile and within 50th and 75th percentile. PC = Principal Component.

Moreover, in sensory areas, topographies result from the maximal response of neurons to a graded stimulus feature. To parallel right TPJ *emotion dimension* gradients with those observed in primary sensory regions, we investigated whether distinct populations of voxels were selective for specific *polarity*, *complexity* and *intensity* scores. Thus, we employed the population receptive field method[43] to estimate the tuning curve of right TPJ voxels for each *emotion dimension*. The map of voxel selectivity was consistent with the topography obtained from the original gradient estimation for the three *emotion dimensions* (*polarity*: $\rho = 0.547$, p-value = 0.001; *complexity*: $\rho = 0.560$, p-value < 0.001 and *intensity*: $\rho = 0.596$, p-value < 0.001). Specifically, results demonstrated the existence of four populations of voxels tuned to specific *polarity* values, which encoded highly and mildly positive and negative events, respectively (Figure 6A). Also, two distinct populations of

voxels were tuned to maximally respond during cognitively mediated affective states (i.e., highly and mildly positive *complexity* values), and two other populations were selective for emotions characterized by higher and lower levels of automatic responses (i.e., highly and mildly negative *complexity* values; Figure 6B). Lastly, for the *intensity* dimension two specific populations of voxels were engaged depending on the strength of the emotional experience (Figure 6C). This further evidence favored the parallel between emotion and sensory gradients.

--Figure 6--

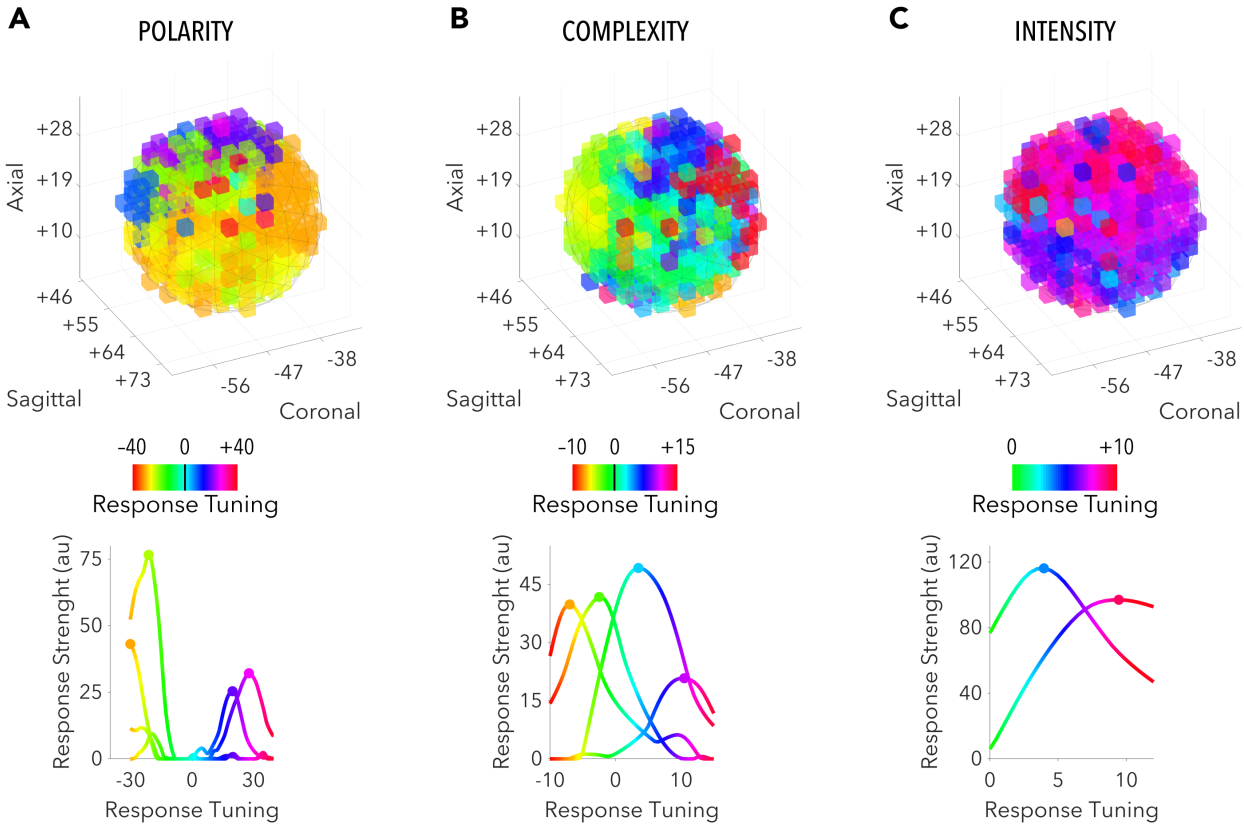

Figure 6 Population receptive field estimates in right TPJ - Response selectivity maps of right TPJ voxels for (**A**) *polarity*, (**B**) *complexity* and (**C**) *intensity*. Preferred responses of distinct populations of voxels were obtained using non-negative matrix factorization (Supplementary Figure 9). Components explaining at least 5% of the variance were plotted as a tuning curve (lowermost row) after averaging all the possible tuning width values for each *emotion dimension* score.

Discussion

Previous studies reported that activity of individual brain regions codes distinct emotion features[7], whereas others suggested that a distributed network of cortical areas conjointly interacts to represent affective states[33]. Interestingly, though, the possibility that gradients may encode the emotional experience as function of either basic emotions, or emotion dimensions, has never been explored. The topological isomorphism between feature space and cortical distances has been successfully adopted to relate psychophysical characteristics of stimuli to patterns of activity in sensory regions[34]. Nonetheless, this biologically advantageous mechanism has been recently proven to lie at the basis of the cortical representation of higher-level (e.g., semantic) features as well[35,38,39]. Building upon this evidence, we tested whether different affective states could be mapped onto the cortical mantle through spatially overlapping gradients.

We demonstrated that the topography of right TPJ, a crucial cortical hub for social cognition[12,22,26,28,29], is described by *emotion dimensions*, rather than by single basic emotions. Indeed, within this region, we discovered three orthogonal and spatially overlapping gradients encoding the *polarity, complexity* and *intensity* of the emotional experience. The peculiar arrangement of these gradients allows a gamut of emotional experiences to be represented in a single patch of cortex, including affective states perceived as pleasant, unpleasant or ambivalent, connoted by calmness or excitement and mediated by primitive reactions or mentalization processes. Therefore, TPJ organization resembles the one that can be observed in primary sensory areas, where stimulus properties are topographically arranged onto the cortical mantle, as in the case of eccentricity and polar angle in the primary visual cortex (V1), frequency in the primary auditory region (A1) and body parts in the primary somatosensory area (S1). In this regard, the evidence that *emotion dimensions* are encoded in a gradient-like manner supports the existence of a biologically plausible mechanism for the coding of affective states, which we named *emotionotopy*. Indeed, as in vision precise portions of V1 map distinct locations of the visual field, specific regions

of temporo-parietal territories code unique emotional experiences. This emerged also from the analysis of response tuning, showing how within each *emotional hemifield* of *polarity* and *complexity*, populations of voxels code specific levels of the emotional experience.

As for the case of polar angle and eccentricity in the primary visual cortex, right TPJ *emotion dimension* gradients are lower-dimensional descriptions of the underlying neural activity. The retinotopic representation of azimuth and elevation in V1 overlaps with local maps of ocular dominance and orientation tuning. Therefore, multiple neural codes exist at different spatial scales and the ability to capture either global or local representations relates to the resolution of the imaging technique. Our data provide evidence for a lower-dimensional, yet biologically favorable, neural code to represent emotions in temporo-parietal regions. Considering the parallel with the organization of sensory areas, we believe that the topography of right TPJ does not prevent the existence of other neural codes, especially considering the coexistence of global and local representations and the multifaceted nature of this region.

Furthermore, the fact that subjective reports explained the activity of other cortical modules is not necessarily in contrast with the topographic organization of TPJ. In fact, as in vision a rich and complex percept relies on both the primary visual cortex to extract fundamental features and other regions to process specific stimulus properties (i.e., V5 for motion), so in emotion processing TPJ may represent an hub embedded in a distributed network of regions carrying out distinct computations.

Richness of the Emotional Experience in Movie Watching

In our study, we employed a naturalistic continuous stimulation paradigm since it fosters emotional contagion and empathic reactions, leading to complex subjective emotional experiences, akin to real life[22]. Indeed, we found that within a 60 seconds time window, emotion transitions represented in the Forrest Gump movie are similar to those experienced in real life[44] and are predicted by a mental model of emotion co-occurrence[45] (please refer to Supplementary Materials

for details). This supports the ecological validity of our stimulus and emphasizes that movies can be successfully adopted to nurture emotional resonance[46-48] also in the fMRI setting.

In movie watching, actions and dialogues generally are not directed toward the observer. Thus, the subjective emotional experience results, on the one hand, from narrative choices aimed at fostering empathic responses and emotional contagion[49] and, on the other hand, from perspective-taking and mentalizing processes[50,51]. The fact that character intentions and beliefs shape the subjective experience in a bystander may also explain the high between-subjects agreement in reports of experienced emotions[46-48]. This is in line with the consistency of behavioral ratings of *happiness*, *fear*, *sadness* and *anger* present in our data. Noteworthy, *surprise* and *disgust* were not consistent across all participants and, even though this may appear as a contradiction with respect to the supposed[52] universalism of basic emotions, it should be noted that our stimulus was not specifically built to reflect the well-established definition of these six emotions. For instance, some of our subjects reported that movie scenes rated as disgusting were mainly associated to situations for which interpreting the context was necessary (e.g., the principal of the school using his power to obtain sexual favors), rather than to repulsive images. This cognitive interpretation of the basic emotion *disgust* was apparently not present in all the subjects, with some of them relying more on its well-established definition for their ratings. Also, the use of six distinct emotion categories offered the opportunity to compare the basic emotion model with the emotion dimension one starting from the same data. In fact, as we (and others[15,17]) have demonstrated, emotion dimensions can be easily derived from reports based on single emotions, whereas the opposite may not be feasible. For instance, *happiness* is associated to positive valence, yet the intensity of such an experience may be different depending on the context (e.g., win the lottery versus meet an old friend). Moreover, while the definition of basic emotions is common across individuals, ratings based on emotion dimensions require participants to be acquainted with the meaning of psychological constructs (e.g., *dominance*[16]).

Nonetheless, single basic emotions provide a coarse description of subjective experiences, since humans may perceive a complex blend of apparently conflicting emotions[53], and affective states could emerge from psychological processes not directly reducible to single emotions[54]. Our rating model, though, does account for this possibility, since subjects were allowed to report throughout the movie the occurrence of more than one emotion, as when simultaneously experiencing *happiness* and *sadness*. Besides the divergences in literature on the precise temporal overlap in perceiving conflicting emotions[53,55], when subjects are free to detail their personal experience, this peculiar emotional state seems to arise.

Therefore, even though our rating model was based on six emotion categories, the possibility to report the perceived intensity of more than one emotion at a time provided a rich online description of subjects' emotional experiences. Indeed, using this procedure we identified fifteen distinct affective states elicited by the Forrest Gump movie (Figure 2), a number compatible with the one reported by previous studies aimed at investigating a wide range of emotion categories (e.g., *craving*, *terror*)[56,29].

Polarity, Complexity and Intensity of the Emotional Experience

With respect to emotion dimensions, the components we identified were deliberately interpreted not following any known model. Nonetheless, the first dimension, *polarity*, mainly relates to positive against negative emotions as in valence[14], whereas the third one, *intensity*, is unipolar and mimics arousal[14]. We considered the second component as a measure of *complexity* of the emotional state. Indeed, this dimension contrasts events in the movie rated as fearful, an emotion with a fast and automatic response[57], against scenes characterized by ambivalence, where cognitive processes play a significant role in generating “mixed emotions”[58]. Even though this component does not pertain to classical emotion dimension theories, an interesting interpretation may relate *complexity* to the involvement of Theory of Mind[28] in emotion perception[59]. In addition, a recent study on the mental representation of emotions[45] described the *human mind* component as a

cardinal dimension of the affective space. In accordance with the definition provided by the authors, this dimension maps states *"[...] purely mental and human specific vs. bodily and shared with animals"*, which is in line with our interpretation of *complexity*.

Right Temporo-Parietal Gradients Do Not Simply Encode Portrayed Emotions

In our analyses, we used the collected behavioral ratings to explain brain activity in independent subjects. In line with previous studies,[10,12] results highlighted a set of regions located mainly in the right hemisphere (Figure 3 and Supplementary Table 1). Interestingly, the peak of association between emotion ratings and brain activity was located in right pSTS/TPJ. This cortical area has been consistently identified as having a central role in the attribution of mental states to others, as demonstrated by functional neuroimaging[28,29], noninvasive transcranial stimulation[60] and lesion studies[61]. In addition, this region spans across the posterior portion of the superior temporal sulcus, which is implicated in emotion perception[10,12,22,62,63]. In line with this, we demonstrated that activity in right TPJ is significantly explained by the process of emotion attribution to others and by the subjective emotional experience as well. Crucially, though, the information coded in the *other-directed emotion attribution model* is not sufficient to fully reconstruct the topography of the subjective emotional experience. In this regard, we reason that the empathic resonance and emotional contagion elicited by emotions of movie characters produced a remapping of events within a subjective framework. The final reported experience was then mapped in right temporo-parietal territories following the three cardinal axes represented by *emotion dimensions*. This view would reconcile previous studies demonstrating the involvement of right TPJ in the representation of subjective emotional experience[10,22,63], in empathic processes[26,27] and in the attribution of beliefs and emotions to others.[28,29,42]

In addition, in the current study, ratings of the emotional experience elicited by an American movie in Italian participants explained brain activity of German subjects. This suggests that the topographic representation of emotions exists regardless of linguistic or micro-cultural differences.

What may depend on the cultural background of each individual is instead the mapping of distinct emotional states within these gradients. For instance, ruminative thinking, sadness and apathy characterize *melancholy* in the Western culture and we speculate that such an emotion would be mapped in the brain as a negative state having high *complexity*. However, if different levels of *polarity*, *complexity* and *intensity* characterize *melancholy* in other cultures (e.g., Eastern), this emotion would be mapped differently with respect to the three *emotion dimension* topographies.

Of note, the optimal description of *emotion dimension* gradients is obtained considering a 15mm radius sphere centered at the meta-analytic peak for the term "TPJ". This region of interest is ~42% larger in volume as compared to the *reverse inference TPJ map* (i.e., the likelihood that the term "TPJ" is used in a study given the presence of reported activation), but also ~32% smaller than the *forward inference* definition (i.e., the likelihood that a region will activate if a study uses the "TPJ" term; please refer to Supplementary Materials for further details). Therefore, *emotion dimension* gradients are best represented in a patch of cortex that approximates the definition of right TPJ based on brain activation studies. Nevertheless, the fact that emotion topography generalizes across a range of spatial scales (i.e., up to 27mm radius sphere) motivated the use of "*temporo-parietal territories*" in describing the location and size of *emotion dimension* gradients.

Limitations

Our study presents the following limitations: first, the effect size we report for the relationship between emotion ratings and brain activity appears to be relatively small (i.e., 7% of explained variance in right TPJ). However, we would like to emphasize three aspects: (1) first, brain regions significantly encoding emotions are selected after rigorous correction for multiple comparisons; (2) second, the magnitude of the effect is in line with recent fMRI literature on the coding of emotions in the brain[29] and the evaluation of the noise ceiling suggests that our *emotion dimension* model explains between 30% (i.e., upper bound) and 54% (i.e., lower bound) of right TPJ activity; (3)

lastly, we used a parsimonious encoding model, in which only six predictors explained 3,595 samplings of brain activity.

Second, although using a larger set of emotion categories the same *polarity*, *complexity* and *intensity* dimensions still emerged (see Supplementary Materials for details), we would like to emphasize that the collected emotion ratings are specific for the present stimulus. Therefore, alternative movies depicting *horror*, or *sexual desire*[56] may produce different emotion dimensions. Indeed, independent studies are required to prove the generalizability of the topographic organization of *polarity, complexity* and *intensity dimensions* within right TPJ.

Conclusions

In summary, our results showed that moment-by-moment ratings of perceived emotions explain brain activity recorded in independent subjects. Most importantly, we demonstrated the existence of orthogonal and spatially overlapping right temporo-parietal gradients encoding *emotion dimensions*, a mechanism that we named *emotionotopy*.

Methods

In the present study, we took advantage of a high quality publicly available dataset, part of the *studyforrest* project[40] (http://studyforrest.org), to demonstrate the existence of a gradient-like organization in brain regions coding emotion ratings. Particularly, we used moment-by-moment scores of the perceived intensity of six basic emotions elicited by an emotionally charged movie (Forrest Gump; R. Zemeckis, Paramount Pictures, 1994), as predictors of fMRI activity in an independent sample. We then tested the correspondence between the fitting of the emotion rating model in TPJ voxels and their relative spatial arrangement to reveal the existence of orthogonal spatially overlapping gradients.

Behavioral Experiment

Participants. To obtain moment-by-moment emotion ratings during the Forrest Gump movie, we enrolled 12 healthy Italian native speakers (5F; mean age 26.6 years, range 24-34). None of them reported to have watched the movie in one year period prior to the experiment. Subjects signed an informed consent to participate in the study, had the right to withdraw at any time and received a small monetary compensation for their participation. The study was conducted in accordance with the Declaration of Helsinki and was approved by the local IRB (Protocol N°1485/2017).

Experimental setup. We started from the Italian dubbed version of the movie, edited following the exact same description reported in the *studyforrest* project (eight movie segments ranging from a duration of 11 to 18 minutes). The movie was presented in a setting free from distractions using a 24" monitor with a resolution of 1920x1080 pixels connected to a MacBook™ Pro running Psychtoolbox[64] v3.0.14. Participants wore headphones in a noiseless environment (Sennheiser™ HD201; 21-18,000 Hz; Maximum SPL 108dB) and were instructed to continuously rate the subjective perceived intensity (on a scale ranging from 0 to 100) of six basic emotions[5] throughout

the entire movie: *happiness*, *surprise*, *fear*, *sadness*, *anger* and *disgust*. Specific buttons mapped the increase and decrease in intensity of each emotion and subjects were instructed to represent their inner experience by freely adjusting or maintaining the level of intensity. Participants were allowed to report more than one emotion at the same time and ratings were continuously recorded with a 10Hz-sampling rate. Subjects were presented with the same eight movie segments employed in the fMRI study one after the other, for an overall duration of 120 minutes. Further, before starting the actual emotion rating experiment, all participants performed a 20 minutes training session to familiarize with the experimental procedure. Specifically, they had to reproduce various levels of intensity for random combinations of emotions that appeared on the screen every ten seconds.

Behavioral data pre-processing. For each subject, we recorded six timeseries representing the moment-by-moment perceived intensity of basic emotions. First, we downsampled timeseries to match the fMRI temporal resolution (2s) and, afterwards, we introduced a lag of 2s to account for the delay in hemodynamic activity. The resulting timeseries were then temporally smoothed using a moving average procedure (10s window). This method allowed us to further account for the uncertainty of the temporal relationship between the actual onset of emotions and the time required to report the emotional state.

Agreement across subjects of the six basic emotions. To verify the consistency in the occurrence of affective states while watching the Forrest Gump movie, we computed the Spearman's ρ correlation coefficient across subjects for each of the six ratings (Figure 1B). Statistical significance of the agreement was assessed by generating a null distribution of random ratings using the IAAFT procedure (Iterative Amplitude Adjusted Fourier Transform[65]; Chaotic System Toolbox), which provided surrogate data with the same spectral density and temporal autocorrelation of the averaged ratings across subjects (1,000 surrogates).

Basic emotion and emotion dimension models. Preprocessed and temporally smoothed single-subject emotion ratings were averaged to obtain six group-level timeseries representing the basic emotion model. After measuring the Spearman's ρ between pairings of basic emotions (Figure 1B), we performed principal component (PC) analysis and identified six orthogonal components, which constituted the emotion dimension model (Figure 1C).

Agreement across subjects of the emotion dimensions. To verify the consistency across subjects of the PCs, we computed the agreement of the six components by means of a leave-one-subject-out cross validation procedure (Figure 1D). Specifically, for each iteration, we performed PC analysis on the left-out subject behavioral ratings and on the averaged ratings of all the other participants. The six components obtained from each left-out subject were rotated (Procrustes analysis, reflection and orthogonal rotation only) to match those derived from all the other participants. This procedure generated for each iteration (i.e., for each of the left-out subjects) six components, which were then compared across individuals using Spearman's ρ similarly to what has been done for the six basic emotions. To assess the statistical significance, we created a null distribution of PCs from the generated surrogate data of the behavioral ratings, as described above (1,000 surrogates).

Richness of the reported emotional experience. Although subjects were asked to report their inner experience using six emotion categories, their ratings were not limited to binary choices. Indeed, at each timepoint raters could specify the perceived intensity of more than one emotion, leading to the definition of more complex affective states as compared to the basic ones. To further highlight this aspect, we performed dimensionality reduction and clustering analyses on emotion timeseries. Starting from emotion ratings averaged across participants, we selected timepoints characterized by the highest intensity (i.e., by summing the six basic emotions and setting the threshold to the 50th percentile) and applied Barnes-Hut t-distributed stochastic neighbor embedding[56,66] (t-SNE; perplexity = 30; theta = 0.05). The algorithm measures the distances between timepoints in the six-

dimensional space defined by the basic emotions as joint probabilities according to a Gaussian distribution. These distances are projected onto a two-dimensional embedding space using a Student's t probability distribution and by minimizing the Kullback–Leibler divergence. To further describe the variety of affective states elicited by the movie, we then applied k-means clustering analysis to the projection of timepoints in the t-SNE manifold and determined the number of clusters using the silhouette criterion[67].

fMRI Experiment

We selected data from the phase II of the *studyforrest* project, in which 15 German mother tongue subjects watched an edited version of the Forrest Gump movie during the fMRI acquisition. Participants underwent two 1-hour sessions of fMRI scanning (3T, TR 2s, TE 30ms, FA 90°, 3mm ISO, FoV 240mm, 3599 tps), with an overall duration of the experiment of 2h across eight runs. Subjects were instructed to inhibit any movement and simply enjoy the movie (for further details[40]). We included in our study all participants that underwent the fMRI acquisition and had the complete recordings of the physiological parameters (i.e., cardiac trace) throughout the scanning time (14 subjects; 6F; mean age 29.4 years, range 20-40 years). For the fMRI pre-processing pipeline please refer to Supplementary Materials.

Encoding Analysis

Voxel-wise encoding[68,69] was performed using a multiple linear regression approach to measure the association between brain activity and the emotion ratings, constituted by the six principal components. Of note, performing a least square linear regression using either the six principal components or the six basic emotion ratings yields the same overall fitting (i.e., full model R^2), even though the coefficient of each column could vary among the two predictor sets.

To reduce the computational effort, we limited the regression procedure to gray matter voxels only (~44k with an isotropic voxel resolution of 3mm). We assessed the statistical significance of the R^2 fitting of the model for each voxel using a permutation approach, by generating 10,000 null encoding models. Null models were obtained by measuring the association between brain activity and surrogate data having the same spectral density and temporal autocorrelation of the original six PCs. This procedure provided a null distribution of R^2 coefficients, against which the actual association was tested. The resulting p-values were corrected for multiple comparisons through the False Discovery Rate[70] method ($q < 0.01$; Figure 3A, Supplementary Figure 1 and Supplementary Table 1). R^2 standard error was calculated through a bootstrapping procedure (1,000 iterations). Moreover, we conducted a noise-ceiling analysis for right TPJ data, similarly to what has been done by Ejaz and colleagues[71] (please see Supplementary Materials).

Emotion Gradients in right TPJ

Estimation of right TPJ topography. We tested the existence of emotion gradients by measuring the topographic arrangement of the multiple regression coefficients[72] in regions lying close to the peak of fitting for the encoding procedure (i.e., right pSTS/TPJ). To avoid any circularity in the analysis[73], we first delineated a region of interest (ROI) in the right pSTS/TPJ territories using an unbiased procedure based on the NeuroSynth[74] database v0.6 (i.e., *reverse inference* probability associated to the term "TPJ"). Specifically, we started from the peak of the "TPJ" NeuroSynth reverse inference meta-analytic map to draw a series of cortical ROIs, with a radius ranging from 9 to 27 mm. Afterwards, to identify the radius showing the highest significant association, for each spherical ROI we tested the relationship between anatomical and functional distance[75] (Supplementary Table 2). This procedure was performed using either multiple regression coefficients obtained from the three *emotion dimensions* or from the four basic emotions stable across all subjects. As depicted in Supplementary Figure 2, we built for each radius two dissimilarity matrices: one using the Euclidean distance of voxel coordinates, and the other one

using the Euclidean distance of the fitting coefficients (i.e., β values) of either the three *emotion dimensions* or the four basic emotions. The rationale behind the existence of a gradient-like organization is that voxels with similar functional behavior (i.e., lower functional distance) would also be spatially arranged close to each other on the cortex[75] (i.e., lower physical distance). The functional and anatomical dissimilarity matrices were compared using the Spearman's ρ coefficient. To properly address the significance of the anatomo-functional association, we built an *ad hoc* procedure that maintained the same spatial autocorrelation structure of TPJ in the null distribution. Specifically, we generated 1,000 IAAFT-based null models for the *emotion dimension* and the basic emotion data, respectively. These null models represented the predictors in a multiple regression analysis and generated a set of null β regression coefficients. Starting from these coefficients we built a set of functional dissimilarity matrices that have been correlated to the anatomical distance and provided 1,000 null Spearman's ρ coefficients, against which the actual anatomo-functional relationship was tested. Confidence intervals (CI, 2.5 and 97.5 percentile) for the obtained correlation values were calculated employing a bootstrap procedure (1,000 iterations). We also tested the existence of gradients in other brain regions encoding emotion ratings using a data-driven searchlight analysis. Results and details of this procedure are reported in Supplementary Materials.

Impact of spatial smoothing on emotion topography. To estimate the significance of right TPJ gradients we used null models built on emotion ratings, leaving untouched the spatial and temporal structure of brain activity. However, as spatial smoothness may still affect the estimation of gradients, we tested right TPJ topography using the group-average unfiltered data. In brief, all the steps described in the *fMRI data pre-processing* section (see Supplementary Materials) were applied, with the only exception of spatial filtering. Following this procedure, the estimated smoothness of the right TPJ region was 4.5x4.2x3.6mm (3dFWHMx tool). Using these data and the same procedure described in the *Estimation of right TPJ topography* paragraph, we measured the

significance of emotion gradients. Results are detailed in Supplementary Table 6 and Supplementary Figure 7.

Impact of cortical folding on emotion topography. The Euclidean metric does not take into account cortical folding. Indeed, because of the morphological characteristics of TPJ, which include a substantial portion of STS sulcal walls, the estimation of emotion gradients would benefit from the use of a metric respectful of cortical topology.

For this reason, we ran the Freesurfer recon-all analysis pipeline[76] on the standard space template[77] used as reference for the nonlinear alignment of single-subject data. We then transformed the obtained files in AFNI-compatible format (@SUMA_Make_Spec_FS). This procedure provided a reconstruction of the cortical ribbon (i.e., the space between pial surface and gray-to-white matter boundary), which has been used to measure the anatomical distance. In this regard, we particularly employed the Dijkstra algorithm as it represents a computationally efficient method to estimate cortical distance based on folding[78,79]. The single-subject unsmoothed timeseries were then transformed into the standard space, averaged across individuals and projected onto the cortical surface (AFNI 3dVol2Surf, map function: average, 15 steps). Afterwards, we performed a multiple linear regression analysis using principal components derived from emotion ratings as predictors of the unsmoothed functional data. This analysis was carried out within a cortical patch that well approximated the size of the 3D-sphere used in the original volumetric pipeline and centered at the closest cortical point with respect to the Neurosynth "TPJ" peak. Thus, for each regressor of interest, we obtained unsmoothed β values projected onto the cortical mantle. We then tested the existence of a gradient-like organization for each predictor, using the Dijkstra algorithm and the same procedure described above (i.e., *Estimation of right TPJ topography* paragraph). Results are detailed in Supplementary Table 6 and Figure 4.

Right Temporo-Parietal Gradients and Portrayed Emotions

We tested whether the gradient-like organization of right TPJ reflects portrayed emotions. Thus, we took advantage of publicly available emotion tagging data of the same movie, provided by an independent group[41]. Differently from our behavioral task, raters were asked to indicate the portrayed emotion of each character (e.g., Forrest Gump, Jenny) in 205 movie segments (average duration ~35s) presented in random order and labeled over the course of approximately three weeks. As also suggested by the authors[41], this particular procedure minimizes carry-over effects and help observers to exclusively focus on indicators of portrayed emotions. Importantly, in their editing, the authors respected the narrative of movie scenes (e.g., Forrest in front of Jenny's grave is a single cut with a duration of ~131s), so that the raters could have a clear understanding of what was shown on the screen. In addition, the possibility to tag emotions independently in each movie segment and to watch each scene more than once, allowed subjects to choose among a larger number of emotion categories[80] (N = 22), as compared to our set of emotions. Moreover, each observer was instructed to report with a binary label whether the portrayed emotion was directed toward the character itself (*self-directed*; e.g., Forrest feeling sad) or toward another character (*other-directed*; e.g., Forrest feeling happy for Jenny). These two descriptions served as *third-person emotion attribution models* and underwent the exact same processing steps (i.e., 2s lagging and temporal smoothing), which have been applied to our *subjective emotion rating model*. As the two *third-person emotion attribution models* included the four basic emotions found to be consistent across observers in our experiment (i.e., *happiness*, *fear*, *sadness* and *anger*), we have been able to directly assess the correlation for these ratings using Spearman's ρ.

We then measured the extent to which the two *third-person emotion attribution models* explained brain activity in right TPJ following the method described in the *Encoding analysis* section. As these two descriptions are higher in dimensionality as compared to our *subjective emotion rating model*, we assessed the significance of fitting using three different procedures: (A) matching the dimensionality across models by selecting the first six principal components only; (B) matching the

emotion categories in ratings, by performing PCA on the four basic emotions shared across models (i.e., *happiness*, *fear*, *sadness* and *anger*); (C) using the full model regardless of the dimensionality (i.e., six components for our *subjective emotion rating model* and 22 for each of the two *emotion attribution models*). In addition, to allow a direct and unbiased comparison between R^2 values obtained from different models, we performed cross-validation using a half-run split method (for further details please refer to Supplementary Materials).

Lastly, we tested whether right TPJ gradients encode *emotion attribution models*. Specifically, we evaluated two different scenarios: (1) the existence of right TPJ gradients encoding the 22 components of each *emotion attribution model*; (2) the possibility to identify emotion gradients following the multidimensional alignment[81] (i.e., canonical correlation analysis) of the 22-dimensional emotion attribution space to the 6-dimensional space defined by subjective ratings. These alternative procedures relate to two different questions: (1) whether the process of emotion attribution is associated to emotion gradients in right TPJ and (2) whether starting from a third-person complex description of portrayed emotions, one can reconstruct the subjective report of our raters. Results for these two procedures are detailed in Supplementary Table 5.

Characterization of Emotion Gradients in right TPJ

Principal orientation of right TPJ gradients. Once the optimal ROI radius was identified, we tested the gradient-like organization of right TPJ for each individual *emotion dimension* and basic emotion (Supplementary Table 3), using the same procedure described above (i.e., *Estimation of right TPJ topography* section). We calculated the numerical gradient of each voxel using β values. This numerical gradient estimates the partial derivatives in each spatial dimension (x, y, z) and voxel, and can be interpreted as a vector field pointing in the physical direction of increasing β values. Afterwards, to characterize the main direction of each gradient, rather than calculating its divergence (i.e., Laplacian of the original data[82,83]), we computed the sum of all vectors in the field

(Figure 4 and Supplementary Figure 2). This procedure is particularly useful to reveal the principal direction of linear gradients and provides the opportunity to represent this direction as the orientation of the symmetry axis of the selected ROI. The above-mentioned procedure was also adopted to assess the reliability of the emotion gradients in each subject. Results and details of this procedure are reported in Supplementary Materials. Furthermore, since gradients built on β coefficients could reflect positive or negative changes in hemodynamic signal depending on the sign of the predictor, we represented the average TPJ activity during movie scenes characterized by specific affective states (Figure 5).

Population receptive field estimates in right TPJ. We investigated whether distinct populations of voxels are selective for specific affective states. To this aim, we employed the population receptive field method[43] (pRF) and estimated the tuning curve of right TPJ voxels for each predictor found to be topographically encoded within this region. We modeled the tuning curve of each voxel as a Gaussian distribution, in which μ represented the preferred score of the predictor and σ the width of the response. The optimal combination of tuning parameters was selected among ~5k plausible values of μ (5th-95th percentile of the scores of each predictor - 0.5 step) and σ (ranging from 1 to 12 - 0.25 step), sampled on a regular grid. Each emotion timeseries was then filtered using these ~5k Gaussian distributions and fitted in brain activity using a linear regression approach. This produced t-values (i.e., β/SE β) expressing the goodness of fit of μ and σ combinations, for each right TPJ voxel. The principal tuning of voxels was then obtained by selecting the combination characterized by the highest t-value across the ~5k samples (Figure 6).

To estimate the similarity between tunings (i.e., μ parameters) obtained from the pRF approach and our original results (i.e., β coefficients of the gradient estimation), we computed Spearman's ρ across right TPJ voxels. The significance of such an association was tested against a null distribution of β coefficients obtained through the IAAFT procedure (N = 1,000).

Lastly, we further characterized the prototypical responses of populations of voxels as function of affective states. To do so, we used the non-negative matrix factorization[84] and decomposed the multivariate pRF data (i.e., voxels t-values for each μ and σ) into an approximated matrix of lower rank (i.e., 10, retaining at least 90% of the total variance). This method allows parts-based representations, as the tuning of right TPJ voxels is computed as a linear summation of non-negative basis responses. The results of this procedure are summarized in Supplementary Figure 9. All the analyses were performed using MATLAB R2016b (MathWorks Inc., Natick, MA, USA).

Acknowledgements

We would like to thank all the people behind the *studyforrest* project, especially Michael Hanke, Annika Labs and colleagues for sharing the Forrest Gump portrayed emotions data and Mark A. Thornton and Diana I. Tamir for making available the real life emotion transitions data. We also thank Maria Luisa Catoni for her valuable suggestions. M.B. was in part supported by the PRIN (Research projects of national interest) 2015 – Italian Ministry of Education, University and Research (MIUR) Prot. 2015AR52F9 granted to P.P.

Author Contributions

G.L., G.H. and L.C., conceived the study, designed the behavioral experiment, developed the code, performed behavioral and fMRI data analysis. G.L., G.H., L.C. and P.P. interpreted the obtained results and drafted the manuscript. A.L., Pa.P. and M.B. contributed to the fMRI data analysis. L.C., E.R. and P.P. critically revised the manuscript. All the authors approved the final version of the manuscript.

Data Availability

The data that support the findings of this study are available from OpenScience foundation, https://osf.io/tzpdf. Raw fMRI data are available at http://studyforrest.org. Real life experience-sampling dataset is available at https://osf.io/zrdpa. Portrayed emotions dataset is available at https://f1000research.com/articles/4-92/v1.

Code Availability

The code and the preprocessed data are publicly available at OpenScience foundation (https://osf.io/tzpdf, doi:10.17605/OSF.IO/TZPDF).

Competing Interests

The authors declare no competing interests.

References

1. Panksepp, J. Toward a general psychobiological theory of emotions. *Behav Brain Sci* **5**(3), 407-422 (1982).

2. Kreibig, S. D. Autonomic nervous system activity in emotion: A review. *Biol Psychol* **84**(3), 394-421 (2010).

3. Stephens, C. L., Christie, I. C., & Friedman, B. H. Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis. *Biol Psychol* **84**(3), 463-473 (2010).

4. Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. K. Bodily maps of emotions. *Proc Natl Acad Sci USA* **111**(2), 646-651 (2014).

5. Ekman, P. An argument for basic emotions. *Cognition & emotion* **6**(3-4), 169-200 (1992).

6. Tracy, J. L., & Randles, D. Four models of basic emotions: a review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emot Rev* **3**(4), 397-405 (2011).

7. Vytal, K., & Hamann, S. Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *J Cogn Neurosci* **22**(12), 2864-2885 (2010).

8. Saarimäki, H., et al. Discrete neural signatures of basic emotions. *Cer Cor* **26**(6), 2563-2573 (2015).

9. Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. Supramodal representations of perceived emotions in the human brain. *J Neurosci* **30**(30), 10127-10134 (2010).

10. Kragel, P. A., & LaBar, K. S. Multivariate neural biomarkers of emotional states are categorically distinct. *Soc Cogn Affect Neurosci* **10**(11), 1437-1448 (2015).

11. Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. The brain basis of emotion: a meta-analytic review. *Behav Brain Sci* **35**(3), 121-143 (2012).

12. Kober, H., et al. Functional grouping and cortical–subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *Neuroimage* **42**(2), 998-1031 (2008).

13. Touroutoglou, A., Lindquist, K. A., Dickerson, B. C., & Barrett, L. F. Intrinsic connectivity in the human brain does not reveal networks for 'basic'emotions. *Soc Cogn Affect Neurosci* **10**(9), 1257-1265 (2015).

14. Russell, J. A. A circumplex model of affect. *J Pers Soc Psychol* **39**(6), 1161 (1980).

15. Smith, C. A., & Ellsworth, P. C. Patterns of cognitive appraisal in emotion. *J Pers Soc Psychol* **48**(4), 813 (1985).

16. Mehrabian, A., & Russell, J. A. *An approach to environmental psychology*. the MIT Press (1974).

17. Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. The world of emotions is not two-dimensional. *Psychol Sci* **18**(12), 1050-1057 (2007).

18. Anderson, A. K., et al. Dissociated neural representations of intensity and valence in human olfaction. *Nat Neurosci* **6**(2), 196 (2003).

19. Wager, T. D., Phan, K. L., Liberzon, I., & Taylor, S. F. Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *Neuroimage* **19**(3), 513-531 (2003).

20. Kassam, K. S., Markey, A. R., Cherkassky, V. L., Loewenstein, G., & Just, M. A. Identifying emotions on the basis of neural activation. *PloS One* **8**(6), e66032 (2013).

21. Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. *Cer Cor* **26**(5), 1910-1922 (2015).

22. Nummenmaa, L., et al. Emotions promote social interaction by synchronizing brain activity across individuals. *Proc Natl Acad Sci USA* **109**(24), 9599-9604 (2012).

23. Mourao-Miranda, J., et al. Contributions of stimulus valence and arousal to visual activation during emotional perception. *Neuroimage* **20**(4), 1955-1963 (2003).

24. Kensinger, E. A., & Schacter, D. L. Processing emotional pictures and words: effects of valence and arousal. *Cogn Affect Behav Neurosci* **6**(2), 110-126 (2006).

25. Mather, M., Mitchell, K. J., Raye, C. L., Novak, D. L., Greene, E. J., & Johnson, M. K. Emotional arousal can impair feature binding in working memory. *J Cogn Neurosci* **18**(4), 614-625 (2006).

26. Morelli, S. A., Rameson, L. T., & Lieberman, M. D. The neural components of empathy: predicting daily prosocial behavior. *Soc Cogn Affect Neurosci* **9**(1), 39-47 (2012).

27. Morelli, S. A., & Lieberman, M. D. The role of automaticity and attention in neural processes underlying empathy for happiness, sadness, and anxiety. *Front Hum Neurosci* **7**, 160 (2013).

28. Saxe, R., & Kanwisher, N. People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". *Neuroimage* **19**(4), 1835-1842 (2003).

29. Skerry, A. E., & Saxe, R. Neural representations of emotion are organized around abstract event features. *Curr Biol* **25**(15), 1945-1954 (2015).

30. Barrett, L. F., & Wager, T. D. The structure of emotion: Evidence from neuroimaging studies. *Curr Dir Psychol Sci* **15**(2), 79-83 (2006).

31. Saarimäki, H., et al. Distributed affective space represents multiple emotion categories across the human brain. *Soc Cogn Affect Neurosci* **13**(5), 471-482 (2018).

32. Clark-Polner, E., Johnson, T. D., & Barrett, L. F. Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cer Cor* **27**(3), 1944-1948 (2017).

33. Wager, T. D., et al. A Bayesian model of category-specific emotional brain responses. *PLoS Comput Biol* **11**(4), e1004066 (2015).

34. Sereno, M. I., et al. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268**(5212), 889-893 (1995).

35. Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**(7600), 453 (2016).

36. Margulies, D. S., et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc Natl Acad Sci USA* **113**(44), 12574-12579 (2016).

37. Huntenburg, J. M., Bazin, P. L., & Margulies, D. S. Large-scale gradients in human cortical organization. *Trends Cogn Sci* **1**, 21-31 (2018).

38. Sha, L., et al. The animacy continuum in the human ventral vision pathway. *J Cogn Neurosci* **27**(4), 665-678 (2015).

39. Harvey, B. M., Klein, B. P., Petridou, N., & Dumoulin, S. O. (2013). Topographic representation of numerosity in the human parietal cortex. *Science*, *341*(6150), 1123-1126.

40. Hanke, M., et al. A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. *Sci Data* **3**, 160092 (2016).

41. Labs, A., et al. Portrayed emotions in the movie" Forrest Gump". *F1000Research* **4** (2015).

42. Van Overwalle, F. Social cognition and the brain: a meta-analysis. *Hum Brain Mapp* **30**(3), 829-858 (2009).

43. Dumoulin, S. O., & Wandell, B. A. Population receptive field estimates in human visual cortex. *Neuroimage* **39**(2), 647-660 (2008).

44. Trampe, D., Quoidbach, J., & Taquet, M. Emotions in everyday life. *PloS one* **10**(12), e0145450 (2015).

45. Thornton, M. A., & Tamir, D. I. Mental models accurately predict emotion transitions. *Proc Natl Acad Sci USA* **114**(23), 5982-5987 (2017).

46. Philippot, P. Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition and emotion* **7**(2), 171-193 (1993).

47. Gross, J. J., & Levenson, R. W. Emotion elicitation using films. *Cognition & emotion* **9**(1), 87-108 (1995).

48. Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition & emotion* **24**(7), 1153-1172 (2010).

49. Smith, M. *Engaging characters: Fiction, emotion, and the cinema*. Oxford: Clarendon Press (1995).

50. Lombardo, M. V., et al. Shared neural circuits for mentalizing about the self and others. *J Cogn Neurosci* **22**(7), 1623-1635 (2010).

51. Raz, G., et al. Cry for her or cry with her: context-dependent dissociation of two modes of cinematic empathy reflected in network cohesion dynamics. *Soc Cogn Affect Neurosci* **9**(1), 30-38 (2013).

52. Mesquita, B., & Walker, R. Cultural differences in emotions: A context for interpreting emotional experiences. *Behav Res Ther* **41**(7), 777-793 (2003).

53. Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. Can people feel happy and sad at the same time?. *J Pers Soc Psychol* **81**(4), 684 (2001).

54. Lindquist, K. A., & Barrett, L. F. A functional architecture of the human brain: emerging insights from the science of emotion. *Trends Cogn Sci* **16**(11), 533-540 (2012).

55. Berrios, R., Totterdell, P., & Kellett, S. Eliciting mixed emotions: a meta-analysis comparing models, types, and measures. *Front Psychol* **6**, 428 (2015).

56. Cowen, A. S., & Keltner, D. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proc Natl Acad Sci USA* **114**(38), E7900-E7909 (2017).

57. Adolphs, R. The biology of fear. *Curr Biol* **23**(2), R79-R93 (2013).

58. Russell, J. A. Mixed emotions viewed from the psychological constructionist perspective. *Emot Rev* **9**(2), 111-117 (2017).

59. Mitchell, R. L., & Phillips, L. H. The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia* **70**, 1-10 (2015).

60. Donaldson, P. H., Rinehart, N. J., & Enticott, P. G. Noninvasive stimulation of the temporoparietal junction: A systematic review. *Neurosci Biobehav Rev* **55**, 547-572 (2015).

61. Campanella, F., Shallice, T., Ius, T., Fabbro, F., & Skrap, M. Impact of brain tumour location on emotion and personality: A voxel-based lesion–symptom mapping study on mentalization processes. *Brain* **137**(9), 2532-2545 (2014).

62. Kragel, P. A., & LaBar, K. S. Decoding the nature of emotion in the brain. *Trends Cogn Sci* **20**(6), 444-455 (2016).

63. Burnett, S., & Blakemore, S. J. Functional connectivity during a social emotion task in adolescents and in adults. *european J Neurosci* **29**(6), 1294-1301 (2009).

64. Kleiner, M., et al. What's new in Psychtoolbox-3. *Perception* **36**(14), 1 (2007).

65. Schreiber, T., & Schmitz, A. Improved Surrogate Data for Nonlinearity Tests. *Phys Rev Lett* 77(4), 635-638 (1996).

66. Maaten, L. V. D., & Hinton, G. Visualizing data using t-SNE. *J Mach Learn Res* **9**(Nov), 2579-2605 (2008).

67. Rousseeuw, P. J., & Kaufman, L. Finding groups in data. *Hoboken: Wiley Online Library* (1990).

68. Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. Encoding and decoding in fMRI. *Neuroimage* **56**(2), 400-410 (2011).

69. Handjaras, G., et al. Modality-independent encoding of individual concepts in the left parietal cortex. *Neuropsychologia* **105**, 39-49 (2017).

70. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* **57**(1), 289-300 (1995).

71. Ejaz, N., Hamada, M., & Diedrichsen, J. Hand use predicts the structure of representations in sensorimotor cortex. *Nat Neurosci* **18**(7), 1034 (2015).

72. Leo, A., et al. A synergy-based hand control is encoded in human motor cortical areas. *Elife* **5**, e13420 (2016).

73. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* **12**(5), 535 (2009).

74. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods* **8**(8), 665 (2011).

75. Yarrow, S., Razak, K. A., Seitz, A. R., & Seriès, P. Detecting and quantifying topography in neural maps. *PloS one* **9**(2), e87178 (2014).

76. Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* **61**(4), 1402-1418 (2012).

77. Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage* (47), S102 (2009).

78. Fischl, B., Sereno, M. I., & Dale, A. M. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**(2), 195-207 (1999).

79. Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J., & Coalson, T. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cer Cor* **22**(10), 2241-2262 (2011).

80. Ortony, A., Clore, G. L., & Collins, A. *The cognitive structure of emotions*. Cambridge university press (1990).

81. Bilenko, N. Y., & Gallant, J. L. Pyrcca: regularized kernel canonical correlation analysis in python and its applications to neuroimaging. *Frontiers in neuroinformatics* **10**, 49 (2016).

82. Haak, K. V., Marquand, A. F., & Beckmann, C. F. Connectopic mapping with resting-state fMRI. *Neuroimage* **170**, 83-94 (2017).

83. Glasser, M. F., et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**(7615), 171-178 (2016).

84. Lee, D. D., & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**(6755), 788 (1999).

Figure Legends

Figure 1. Emotion ratings

A. Violin plots show the agreement (Spearman's ρ coefficient) of the six basic emotions across subjects. Gray shaded area represents the null distribution of behavioral ratings and dashed lines the mean and 95th percentile of the null distribution. **B**. Correlation matrix showing Spearman's ρ values for pairings of basic emotions. **C**. Principal Component Analysis: loadings of the six principal components. Explained variance was 45% for *polarity*, 24% for *complexity* and 16% for *intensity*. **D**. Violin plots show the agreement (Spearman's ρ coefficient) of the six principal components across subjects. Gray shaded area represents the null distribution of behavioral ratings and dashed lines the mean and 95th percentile of the null distribution. HA = *Happiness*, SU = *Surprise*, FE = *Fear*, SA = *Sadness*, AN = *Anger*, DI = *Disgust*, PC = Principal component, PO = *Polarity*, CO = *Complexity*, IN = *Intensity*.

Figure 2. Richness of the emotional experience

Results of the dimensionality reduction (t-SNE) and clustering analyses (k-means) on the group-averaged behavioral ratings showing the existence of 15 distinct affective states throughout the movie. Each element represents a specific timepoint in the movie and the distance between elements depends on the statistical similarity of emotion ratings. Element color reflects the scores of the *polarity* and *complexity* dimensions: positive (+) and negative (-) events (i.e., *polarity*) are associated respectively to the red and blue channels, whereas *complexity* (Ψ) scores modulate the green channel. Pie charts show the relative contribution of the six basic emotions to each of the 15 identified clusters. HA = Happiness, SU = Surprise, FE = Fear, SA = Sadness, AN = Anger, DI = Disgust.

Figure 3. Encoding of emotion ratings

A. Brain regions encoding emotion ratings corrected for multiple comparisons through the False Discovery Rate method ($q < 0.01$). **B**. Peak of association between emotion ratings and brain activity (purple sphere) and reverse inference peak for the term "TPJ" as reported in the NeuroSynth database (yellow sphere). Coordinates represent the Center of Gravity in MNI152 space. **C**. β coefficients associated to basic emotions in a spherical region of interest (27mm radius) located at the reverse inference peak for the term "TPJ". Maps for emotions not consistent across all the subjects (i.e., surprise and disgust) are faded (see the *Agreement across subjects of the six basic emotions* section). **D**. β coefficients associated to emotion dimensions in a spherical region of interest (27mm radius) located at the reverse inference peak for the term "TPJ". Maps for components not consistent across all the subjects (i.e., PC_4, PC_5 and PC_6) are faded (see the *Agreement across subjects of the Emotion Dimensions* section). IFG = Inferior Frontal Gyrus, rMFG = rostral Middle Frontal Gyrus, mSFG = Medial Superior Frontal Gyrus, preCS = Precentral Sulcus, pSTS/TPJ = posterior part of the Superior Temporal Sulcus/Temporoparietal Junction, MOG = Middle Occipital Gyrus, pMTG = posterior Middle Temporal Gyrus, SMG = Supramarginal Gyrus, LatS = Lateral Sulcus, STS = Superior Temporal Sulcus.

Figure 4. Emotion gradients in right TPJ

A. We revealed three orthogonal and spatially overlapping emotion dimension gradients (*polarity*, *complexity* and *intensity*) within a region of interest located at the reverse inference peak for the term "TPJ" (15mm radius sphere). Symmetry axis of the region of interest represents the main direction of the three gradients **B**. β coefficients of the *polarity* dimension are mapped through a gradient with an inferior to superior direction. **C**. β coefficients of the *complexity* dimension are mapped through a gradient with a posterior to anterior direction. **D**. β coefficients of the *intensity* dimension are mapped through a gradient with an inferior to superior direction. For single-subjects

results please refer to Supplementary Figure 6. Lowermost row depicts the arrangement of the *emotion dimension* gradients in surface space. CoG = Center of Gravity, STS = Superior Temporal Sulcus.

Figure 5. Characterization of Emotion Dimension Gradients in right TPJ

A. Average hemodynamic activity in right TPJ related to the scores below and above the 50th percentile for *polarity*. **B**. Average hemodynamic activity in right TPJ related to the scores below and above the 50th percentile for *complexity*. **C**. Since *intensity* is not bipolar as the other two components (i.e., scores ranged from ~0 to positive values only), for this dimension we mapped the average TPJ activity above the 75th percentile and within 50th and 75th percentile. PC = Principal Component.

Figure 6. Population receptive field estimates in right TPJ

Response selectivity maps of right TPJ voxels for (**A**) *polarity*, (**B**) *complexity* and (**C**) *intensity*. Preferred responses of distinct populations of voxels were obtained using non-negative matrix factorization (Supplementary Figure 9). Components explaining at least 5% of the variance were plotted as a tuning curve (lowermost row) after averaging all the possible tuning width values for each *emotion dimension* score.