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 behavioral descriptions of affective states to brain activity is unclear, as evidence supports either local or distributed processing. A biologically favorable alternative is provided by the notion of gradients, which postulates the isomorphism between 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 topographical organization of affective states. Right TPJ activity is explained by orthogonal and spatially overlapping gradients encoding the *polarity*, *complexity* and *intensity* of emotional experiences. The peculiar arrangement of this three-dimensional functional space allows the brain to map a wide gamut of affective states. As this organization resembles the coding of psychophysical properties in sensory regions (i.e., retinotopy in V1), we propose *emotionotopy* as the underlying principle of emotion perception in TPJ.

Introduction

Emotions promptly translate inner experiences into specific patterns of interpretable behaviors. The ability to infer others' affective states represents a crucial aspect both when humans directly relate to each other and when they simply observe social interactions. Through years, the relevance of such an ability motivated psychologists to search for models that optimally associate behavioral responses to subjective emotional experiences. In this regard, seminal works pointed toward the existence of discrete basic emotions, characterized by distinctive and culturally stable facial expressions[1], specific body postures, 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 (for a summary[6]). To prove the neurobiological validity of these models, neuroscientists investigated whether basic emotions would elicit specific patterns of brain responses, consistently across subjects. Findings show that activity of the medial prefrontal, the anterior cingulate, the superior temporal, the insular, the middle and inferior frontal cortex, as well as of the amygdala, is associated to the perceived intensity of emotions and supports their recognition[7-9] (for a review[10,11]). 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[12,13].

An alternative theory proposes that behavioral, physiological and subjective characteristics of emotions would be more appropriately described along a number of continuous cardinal dimensions[14-17], 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 novelty, are needed to adequately explain affective states[18,19]. Neuroimaging studies also demonstrated that stimuli varying in valence and arousal elicit specific and reliable brain

responses[20,21], which have been recently employed to decode emotional experiences[22]. Activity recorded in the insula, the amygdala, the ventral striatum, the anterior cingulate, the superior temporal and the ventromedial prefrontal cortex, is associated to transitions between positive and negative valence and fluctuations in arousal[23,24].

Thus, despite this large body of evidence, it remains to be determined whether the subjective emotional experience is better explained through basic emotion or emotion dimension models. Moreover, regardless of the adopted model, it is still debated how emotion features are spatially encoded in the brain[8,13,25-28]. 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[29].

An alternative and biologically favorable perspective may be provided by the notion of gradients. 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[30]. 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 two orthogonal gradients. Crucially, recent investigations revealed that gradients support the representation of higher-order information as well[31-33], with features as animacy or numerosity being topographically arranged onto the cortical mantle[31,34].

Following this view, we hypothesize that a gradient-like organization may represent the anatomo-functional grounding principle of the human emotional system. Specifically, different affective states would be mapped onto the cortical mantle through spatially overlapping gradients, either representing specific single emotions (e.g., sadness) or, alternatively, emotion dimensions (e.g., valence). The pattern of brain activity would therefore 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, which we named *emotionotopy*.

Results

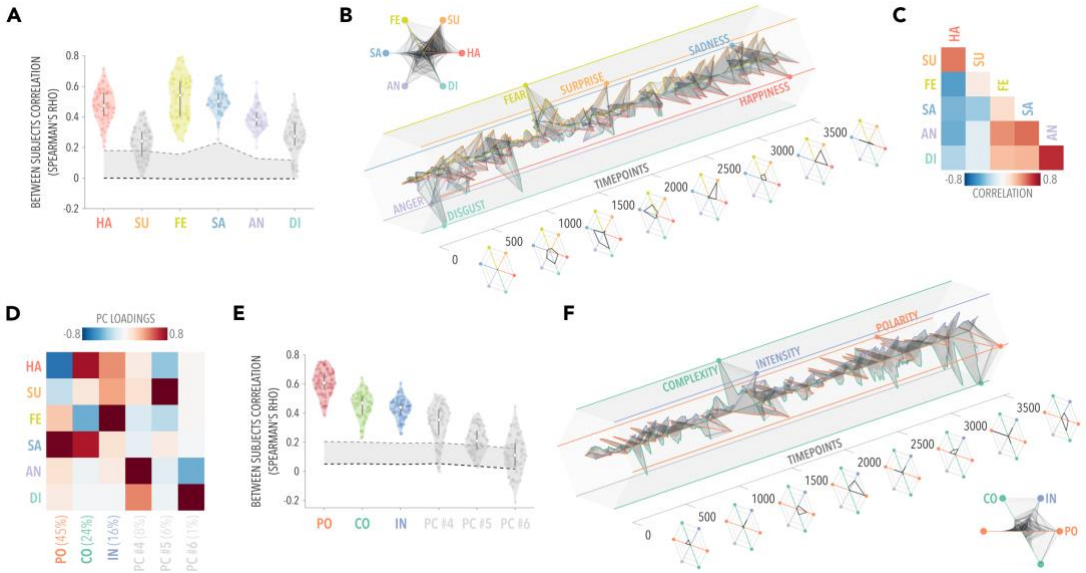
Behavioral Experiment: Emotion Ratings

A group of Italian native speakers continuously rated the perceived intensity of six basic emotions (i.e., happiness, surprise, fear, sadness, anger and disgust[5]) 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 and performed Principal Component (PC) analysis (Figure 1B-D). The first component reflected a measure of *polarity* (PC_1: 45% explained variance) as sadness and happiness yielded positive and negative loadings, respectively. 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 consistently showed only positive loadings (Figure 1D). 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 1E). The fourth PC described movie segments where participants experienced anger and disgust at the same time (ρ = 0.329 ± 0.128, range -0.003 - 0.529), whereas the fifth PC was mainly related to surprise (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 for the sixth PC were not significantly consistent across subjects (p-value > 0.05).

--- Figure 1 ---

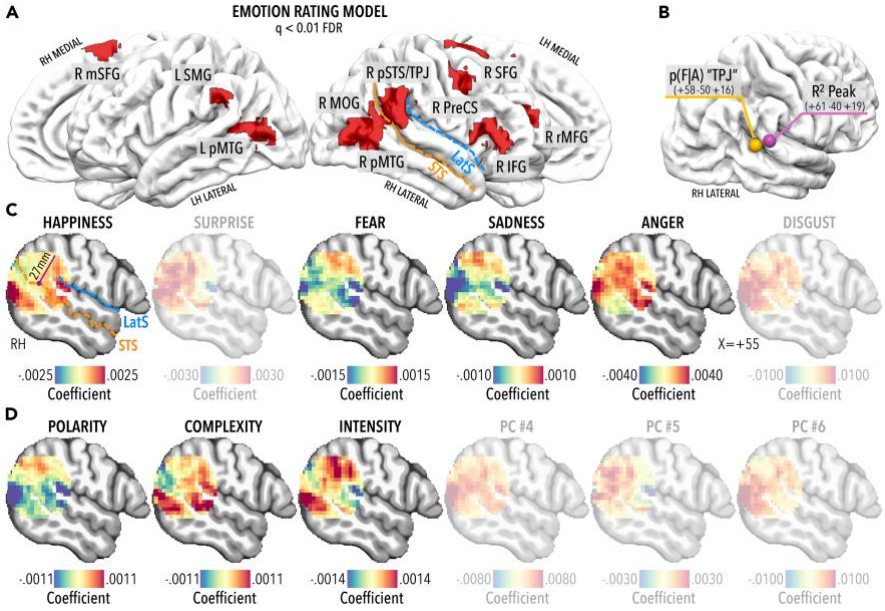


Brain Regions encoding Emotion Ratings

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). The model significantly explained activity in right inferior frontal gyrus (IFG), rostral middle frontal gyrus (rMFG), medial superior frontal gyrus (mSFG), transverse occipital sulcus (TOS), 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 2A and Supplementary Table 1). Notably, the peak of association between emotion ratings and brain activity was located in the right pSTS/TPJ (R^2 = 0.07 ± SE = 0.009; Center of Gravity – CoG: x = 61, y = -40, z = 19; Figure 2B and Supplementary Figure 1), an important hub for the processing of emotional events and social cognition[35]. 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 2B).

--- Figure 2 ---



Emotion Gradients in right TPJ

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 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 2C). 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 2D).

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 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).

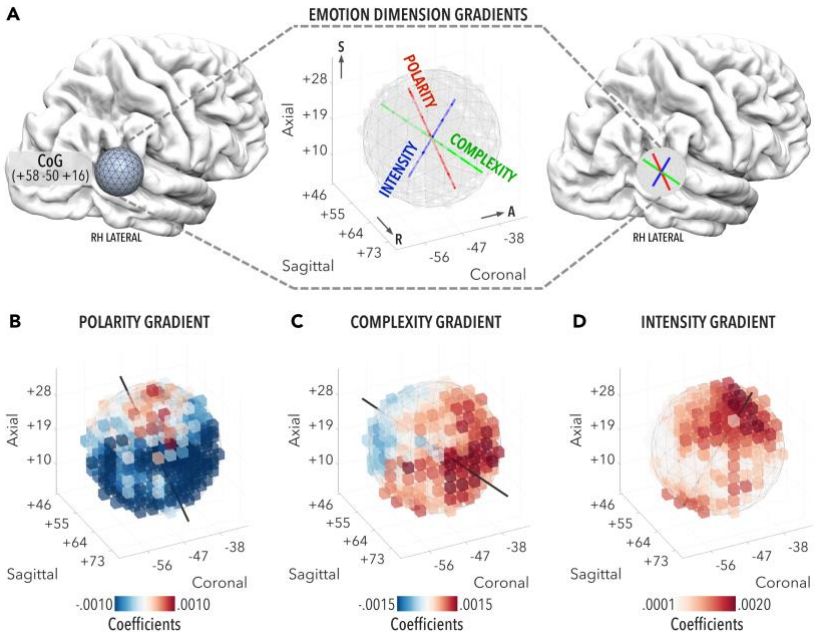
Crucially, when focusing on each *emotion dimension*, results revealed the existence of three orthogonal and spatially overlapping gradients: *polarity* ($\rho = 0.241$, p-value = 0.041, 95% CI: 0.235 to 0.247), *complexity* ($\rho = 0.271$, p-value = 0.013, 95% CI: 0.265 to 0.277) and *intensity* ($\rho = 0.229$, p-value = 0.049, 95% CI: 0.223 to 0.235; Figure 3 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 3) was also identified using single-subject fMRI data (Supplementary Figure 6). Results obtained from the meta-analytic definition of right TPJ were also confirmed by a data-driven searchlight analysis (Supplementary Figure 7). Specifically, we tested the existence of *emotion dimension* gradients in a spherical region of interest (15mm radius) moving across brain regions significantly encoding emotion ratings (see *Brain Regions encoding Emotion Ratings* paragraph). We found that *emotion dimensions* were topographically arranged within a patch of cortex centered in right pSTS/TPJ only ($q < 0.05$ FDR corrected and cluster size > 10; CoG: x = 58, y = -53, z = 21). Here, *polarity* ($\rho = 0.289$, p-value = 0.008), *complexity* ($\rho =$

0.318, p-value = 0.002) and *intensity* (ρ = 0.250, p-value = 0.024) retained a specific gradient-like organization, similar to the one highlighted using the hypothesis-driven approach.

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 (ρ =0.208, p-value = 0.356) or the *emotion dimension* one (ρ =0.251, p-value = 0.144; Supplementary Table 2). Specifically, neither any of the *emotion dimensions* (*polarity*: ρ = 0.132, p-value = 0.354; *complexity*: ρ = 0.157, p-value = 0.222; *intensity*: ρ = 0.149, p-value = 0.257) nor any of the basic emotions showed a gradient-like organization in left TPJ (*happiness*: ρ = 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).

To summarize, we found that *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. We also found that *happiness* (28% of the total variance in behavioral ratings) was the only basic emotion consistent across subjects and 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 support the existence of *emotion dimension* gradients in right temporo-parietal territories, rather than the coding of single basic emotions. To better detail how these 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*. Results for this procedure are detailed in the next section.

--- Figure 3 ---



Characterization of Emotion Dimension Gradients in right TPJ

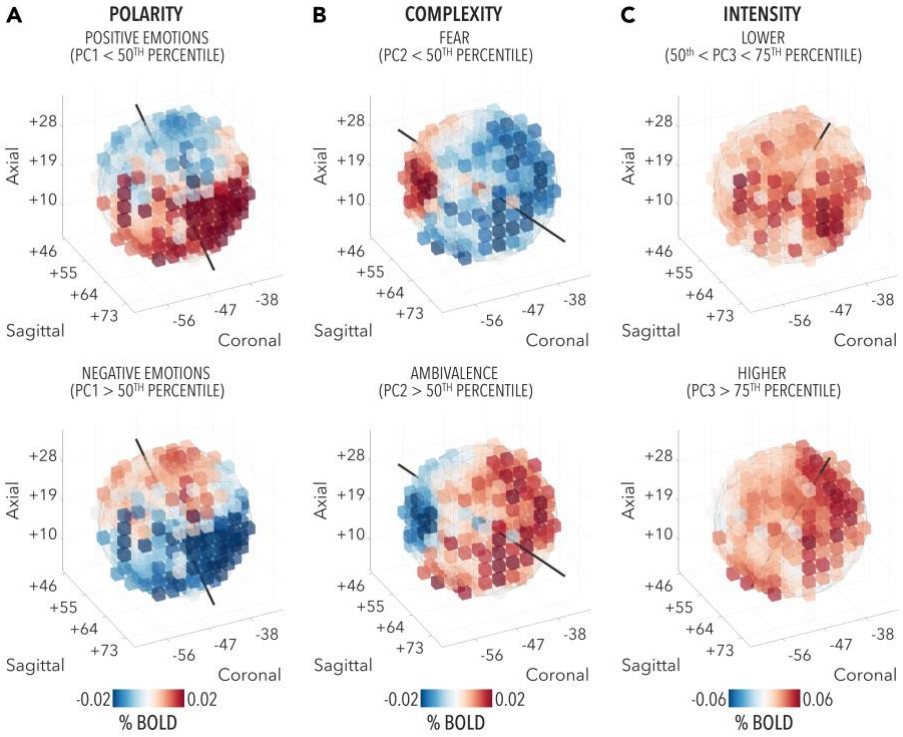
The orientation of the three *emotion dimension* gradients was represented as the symmetry axis of our region of interest. Specifically, for *polarity* highly positive events (i.e., events connoted by positive emotions) increased activity in ventrorostral territories, lying close to the superior temporal sulcus, whilst highly negative events augmented hemodynamic activity in dorsocaudal portions of right TPJ, extending to the posterior banks of Jensen sulcus (Figure 4A and 2D).

Events connoted by higher *complexity* (i.e., concurrent presence of happiness and sadness) were associated to signal increments in rostrolateral territories of right TPJ, whereas those rated as having lower *complexity* (i.e., fearful events) increased hemodynamic activity in its caudal and medial part, encompassing the ascending ramus of the superior temporal sulcus (Figure 4B and 2D). 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 4C and 2D).

Noteworthy, the orthogonal arrangement of *polarity* and *complexity* in right TPJ and the fact that *intensity* was represented both superiorly and inferiorly to the superior temporal sulcus determined

that all the possible combinations of emotional states elicited by the 'Forrest Gump' movie could be mapped within this region.

--- Figure 4 ---



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[29]. 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 adopted to successfully relate psychophysical characteristics of stimuli to patterns of activity in sensory regions[30]. Nonetheless, this biologically advantageous mechanism has been recently proven to lie at the basis of cortical representation of higher-level (e.g., semantic) features as well[31,34]. 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 activity of right TPJ, a crucial cortical hub for social cognition[36], is better described through *emotion dimensions*, rather than using single basic emotions. Crucially, within this region, we discovered three orthogonal and spatially overlapping gradients encoding the *polarity, complexity* and *intensity* of the emotional experience. Therefore, TPJ organization resembles the one that can be observed in primary sensory cortices, where stimulus properties are topographically arranged onto the cortical mantle, as for eccentricity and polar angle in the primary visual cortex (V1), pitch in the primary auditory region (A1) and body parts in the primary somatosensory area (S1). In this regard, the evidence that temporo-parietal territories encode emotion dimensions in a gradient-like manner represents an alternative and biologically plausible mechanism of coding the subjective affective states in the human brain. Thus, we propose that as retinotopy represents the organizing principle of sight in V1, cochleotopy of hearing in A1 and somatotopy of somesthesis in S1, *emotionotopy* constitutes the basis of emotion perception in right TPJ. Indeed, this peculiar organization permits the cortical representation of numerous affective

states: as in vision precise portions of V1 map distinct locations of the visual field, specific regions of temporo-parietal territories code unique emotional experiences.

In our study, we employed a naturalistic continuous stimulation paradigm since it fosters emotional contagion and empathic reactions, leading to complex personal emotional experiences, akin to real life[24]. Indeed, the simple observation of others can trigger the behavioral and physiological representation of affective states in a bystander. In this regard, movies offer a compelling method to nurture emotional resonance and are now widely adopted in fMRI investigations[37,38]. Also, when experiencing emotions, a complex and multifaceted cascade of internal events takes place, and different psychological processes, such as mentalization and emotion recognition, concur to shape the final percept[39]. In our study, we specifically asked an independent sample of subjects to give a continuous rating of the perceived intensity of six basic emotions considered culturally stable[5], and having a common definition across individuals[1]. Although we asked our participants to focus on their inner experience, it is likely that common intrinsic and automatic processes, as empathic resonance[40] and the attribution of beliefs and emotions[41] to movie characters, participated in modulating their subjective reports. This can also be inferred from the high agreement in the behavioral ratings of happiness, fear, sadness and anger (89% of the total variance). Noteworthy, surprise and disgust were not consistent across all participants and, even though this may appear as a contradiction with respect to the supposed[42] 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 where 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. It is important to note that, in our study, the six emotions were adopted as a rating model that provided sufficiently detailed online descriptions of subjects' emotional experiences and also allowed to compare the basic

emotion model to the emotion dimension one starting from the same data. In fact, as we (and others[15,18]) 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*[19]).

Nonetheless, single basic emotions provide a coarse description of subjective experiences, since humans may perceive a complex blend of apparently conflicting emotions[17], and affective states could emerge from psychological processes not directly reducible to single emotions[35,43]. Importantly, our rating model accounts for this possibility, since subjects could report throughout the movie the occurrence of more than one emotion at the same time, as when simultaneously experiencing happiness and sadness. Although some authors criticized that such diametrically opposite affective states may coexist at the exact same moment[44], we have been able to reveal and track this ambivalence. Besides the divergences in literature on the precise temporal overlap in perceiving conflicting emotions[45,46], when subjects are free to detail their personal experience, as in our case, this peculiar emotional state seems to arise.

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[47], against scenes characterized by ambivalence, where cognitive processes play a significant role in generating "mixed emotions"[48]. 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[36] (ToM) in emotion perception[39]. In this regard, the two extremes

of this bipolar continuum characterize automatic responses on one side and cognitive processing of context and others' mental and affective states on the other one.

Starting from the collected behavioral ratings, we performed an encoding procedure to explain brain activity. Results highlighted a set of regions distributed onto the cortical mantle and located mainly in the right hemisphere, as previously reported also by others[26,29] (Figure 2 and Supplementary Table 1). Interestingly, the peak of association between emotion ratings and brain activity was located in right TPJ. This cortical area has been consistently identified as having a central role in moral judgment and intention attribution among other processes, as demonstrated by functional neuroimaging[35,36], noninvasive transcranial stimulation[49] and lesion studies[50]. In addition, this region spans across the posterior portion of the superior temporal sulcus, which seems to be implicated in social and emotion perception, independently from the sensory modality[10].

In the current study, we sought to disentangle whether basic emotions or emotion dimensions are represented in a gradient-like manner in right TPJ. The rationale behind this enquiry is that in topographically organized regions, stimuli are mapped with a specific spatial arrangement and their similarity in feature space is represented by cortical distance. Our analysis demonstrated that right TPJ retains a gradient-like organization for *emotion dimensions*, but not for basic emotions. In line with this, previous studies revealed a spatial overlap of valence and arousal in right temporo-parietal territories[24], even though the fine-grained topographical organization we found within this region has never been reported. Specifically, the peculiar and orthogonal arrangement of *polarity*, *intensity* and *complexity* gradients in right TPJ allows a gamut of emotional experiences to be represented on the human cortex, including affective states perceived as pleasant, unpleasant or ambivalent, connoted by calmness or excitement and mediated by primitive reactions or mentalization processes. Also, this finding is specific for right TPJ, as when we explored whether other brain regions associated to the continuous rating of emotional experience showed a gradient-like organization, we did not find any significant result (Supplementary Figure 7). Nonetheless, the fact that behavioral reports of the perceived intensity of emotions explained activity in other cortical

modules is not necessarily in contrast with the central role of TPJ. In fact, as in vision a rich and complex percept requires both the primary visual cortex to extract fundamental features and other regions to process specific stimulus properties (i.e., V5 for motion), in emotion perception TPJ may represent a crucial hub embedded in a distributed network of regions carrying out distinct computations.

One potential criticism of this study is that 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 four aspects: (1) firstly, brain regions significantly encoding emotions are selected after rigorous correction for multiple comparisons; (2) secondly, the magnitude of the effect is in line with recent fMRI literature on the coding of emotions in the brain[35]; (3) thirdly, the encoding procedure for psychophysical features of the stimulus yielded similar effect size, with volume energy explaining approximately 6% of the total variance in primary auditory cortex and 3% in inferior colliculus; (4) lastly, it should be noted that we used a parsimonious encoding model, where only 6 predictors explained 3,595 samplings of brain activity.

Another possible criticism is that emotion ratings were based on affective states elicited by the 'Forrest Gump' movie, which however evoked complex and multifaceted experiences across subjects. Particularly, throughout the movie 26% of timepoints (2s temporal resolution) were on average rated as neutral, 38% were connoted by a single emotion, 29% by two emotions and 6% by the concurrent experience of three distinct emotions. Nonetheless, further studies employing other naturalistic emotionally charged stimuli are needed to confirm the gradient-like organization of emotion dimensions in right TPJ.

In summary, our results demonstrate that moment-by-moment ratings of perceived emotions explain brain activity recorded in independent subjects and, most importantly, the existence of *emotionotopy* in right TPJ, where orthogonal and spatially overlapping gradients of emotion dimensions may represent the anatomo-functional principle of emotion perception in humans.

Methods

In the present study, we took advantage of a high quality publicly available dataset, part of the *studyforrest* project[51] (http://studyforrest.org), to demonstrate the existence of a gradient-like organization in brain regions coding the subjective emotional experience. Particularly, we used moment-by-moment ratings 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.

Data acquisition and processing

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[51] (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[52] 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 throughout

the entire movie: happiness, surprise, fear, sadness, anger and disgust[5]. 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 and even to report more than one emotion at the same time. 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 10 seconds.

Behavioral Data pre-processing. For each subject, we recorded six timeseries representing the moment-by-moment perceived intensity of six basic emotions. Firstly, for each basic emotion, 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 Spearman's correlation (ρ) across subjects for each of the six ratings (Figure 1A). Statistical significance of the agreement was assessed by generating a null distribution of random ratings using the IAAFT procedure (Iterative Amplitude Adjusted Fourier Transform[53]; Chaotic System Toolbox - mathworks.com/matlabcentral/fileexchange/1597-chaotic-systems-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 (Figure 1B). After measuring the Spearman's ρ between pairings of basic emotions (Figure 1C), we performed principal component (PC) analysis and identified six orthogonal components, which constituted the *emotion dimension* model (Figure 1D and F).

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 1E). Specifically, for each iteration, we performed PC analysis on the left-out subject behavioral ratings and on the averaged behavioral 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). All the analyses were performed using MATLAB R2016b (MathWorks Inc., Natick, MA, USA).

fMRI Experiment

fMRI Data Acquisition. We selected data from the phase II of the *studyforrest* project[51], 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[51]). 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).

fMRI data pre-processing

We employed ANTs[54] and AFNI[55] v.17.2.00 to preprocess MRI data (Supplementary Figure 3). First, structural images were brain extracted (antsBrainExtraction.sh) and non-linearly transformed to match the MNI152 template (3dQwarp). The estimated deformation field was subsequently used to bring single-subject activation maps from the original to the standard space.

Functional data were corrected for intensity spikes (3dDespike) and adjusted for slice timing acquisition (3dTshift). We also compensated head movements by registering each volume to the most stable timepoint (3dvolreg). In this regard, a rigid body transformation was adopted and the 6 estimated motion parameters were included as confounds in further analyses. The transformation matrices were also used to compute an aggregated measure - framewise displacement[56] - that highlighted timepoints affected by excessive motion. Functional data were linearly (align_epi_anat.py) and non-linearly registered to the T1w images, also correcting for phase distortion, and warped to match the MNI152 template using the already computed deformation field (3dNwarpApply). Furthermore, timeseries were smoothed until they reached a full width at half maximum of 6mm (Gaussian kernel, 3dBlurToFWHM). Lastly, we ruled out the effects of signal drifts, head motion and heartbeat (3dDeconvolve) to obtain timeseries of brain activity cleaned from these nuisance regressors.

Following the same procedure adopted for the behavioral processing, single-subject preprocessed fMRI data were averaged to obtain group-level hemodynamic activity and for each voxel the same windowing procedure was employed to temporally smooth data (moving average: 10s window; Supplementary Figure 4). From the obtained aggregated and smoothed timeseries, the timecourse of low-level acoustic (i.e., volume energy - RMS of the signal) and visual (i.e., Gabor contrast energy for 0.5 and 8 cyc/deg spatial frequencies for each frame) features of the movie were regressed out to mitigate the possible collinearities between emotion ratings and psychophysical properties of the stimulus (e.g., fearful events might be associated to sudden volume increases). Specifically, the

RMS value was estimated on 2s non-overlapping windows[57] matching the TR of the fMRI scan. For the low-level visual features instead, we modeled the canonical response of area V1[58]. Each movie frame was filtered with a set of oriented Gabor filters encompassing the lowest and highest limits of V1 spatial frequency selectivity (0.5 and 8 cyc/deg), as found by cell recordings in non-human[59] and by fMRI in humans primates[60,61]. Filters response was averaged across four orientations (i.e., 0, 45, 90, 135 deg) and all pixels, to obtain a global descriptor for each frequency in each frame. Subsequently, visual features were temporally averaged across frames, delayed and smoothed in time to match the fMRI data.

Overall, low-level features modelling generated three regressors of no interest (i.e., low and high spatial frequencies of movie frames and RMS of the audio track) that were regressed out from brain activity using a multiple regression analysis. Details of the procedure and the results are depicted in Supplementary Figure 5.

The obtained regression residuals, consisting of 3,595 timepoints, were considered as the dependent variable in the encoding analysis having emotional ratings as predictors.

fMRI encoding analysis

Voxel-wise encoding[62] 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 yield 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

principal components. This procedure generated a null distribution of R^2 coefficients, against which the actual association was tested. The resulting p-value was corrected for multiple comparisons through the False Discovery Rate[63] method ($q < 0.01$; Figure 2A, Supplementary Figure 1 and Supplementary Table 1). R^2 standard error was calculated trough a bootstrapping procedure (1,000 iterations). All the analyses were performed using MATLAB R2016b (MathWorks Inc., Natick, MA, USA).

Emotion Gradients in right TPJ

We tested the existence of emotion gradients by measuring the topographical arrangement of the multiple regression coefficients[64] in regions lying close to the peak of fitting for the encoding procedure (i.e., right posterior superior temporal sulcus-pSTS/temporoparietal junction-TPJ). To avoid any circularity in the analysis[65], we first delineated a region of interest (ROI) in the right pSTS/TPJ territories using an unbiased procedure based on the NeuroSynth[66] 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[67] (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 β coefficients related to the fitting 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., low functional distance) would also be spatially arranged close to each other on the cortex (i.e., low 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 Figure 7.

Characterization of Emotion Dimension Gradients in right TPJ

Once the optimal ROI radius was identified, we tested the existence of a gradient-like organization for each individual emotion dimension and basic emotion, using the same procedure described above (Supplementary Table 3). 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[68,69]), we computed the sum of all vectors in the field (Figure 3 and Supplementary Figure 2). This procedure is particularly useful to reveal the principal direction of each linear gradient 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 dimension* gradients in each subject. Results and details of this procedure are reported in Supplementary Table 4 and Supplementary Figure 6. Furthermore, since gradients built on β coefficients could reflect positive or negative changes in hemodynamic signal depending on the emotion dimension score, we represented the average TPJ activity during movie scenes having *polarity* and *complexity* scores below and above the 50th percentile. Instead, 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 (Figure 4).

All the analyses were performed using MATLAB R2016b (MathWorks Inc., Natick, MA, USA). The code and the preprocessed data are publicly available.

Acknowledgements

We would like to thank all the people behind the *studyforrest* project.

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 and drafted the manuscript. A.L., Pa.P. and M.B. contributed to the fMRI data analysis and revised the manuscript. E.R. and P.P. revised the manuscript.

Competing Interests statement

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. Decoding the nature of emotion in the brain. *Trends Cogn Sci* **20**(6), 444-455 (2016).

11. Hamann, S. What can neuroimaging meta-analyses really tell us about the nature of emotion?. *Behav Brain Sci* **35**(3), 150-152 (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. Lang, P. J. The emotion probe: studies of motivation and attention. *Am Psychol* **50**(5), 372 (1995).

17. 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).

18. 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).

19. Bakker, I., van der Voordt, T., Vink, P., & de Boon, J. Pleasure, arousal, dominance: Mehrabian and Russell revisited. *Curr Psychol* **33**(3), 405-421 (2014).

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

21. 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).

22. 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).

23. 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).

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

25. 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).

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

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

28. 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).

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

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

31. 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).

32. 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).

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

34. 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.

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

36. 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).

37. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. Intersubject synchronization of cortical activity during natural vision. *Science* **303**(5664), 1634-1640 (2004).

38. Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. Development of the social brain from age three to twelve years. *Nat Commun* **9**(1), 1027 (2018).

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

40. De Vignemont, F., & Singer, T. The empathic brain: how, when and why?. *Trends Cogn Sci* **10**(10), 435-441 (2006).

41. Shamay-Tsoory, S. G., & Aharon-Peretz, J. Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia* **45**(13), 3054-3067 (2007).

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

43. 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).

44. Russell, J. A., & Carroll, J. M. On the bipolarity of positive and negative affect. *Psychol Bull* **125**(1), 3 (1999).

45. Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time?. *Journal of personality and social psychology*, *81*(4), 684.

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

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

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

49. 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).

50. 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).

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

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

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

54. Avants, B. B., Tustison, N., & Song, G. Advanced normalization tools (ANTS). *Insight* **2**, 1-35 (2009).

55. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* **29**(3), 162-173 (1996).

56. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**(3), 2142-2154 (2012).

57. Lahnakoski, J. M., et al. Stimulus-related independent component and voxel-wise analysis of human brain activity during free viewing of a feature film. *PloS one* **7**(4), e35215 (2012).

58. Hubel, D. H., & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* **160**(1), 106-154 (1962).

59. Foster, K. H., Gaska, J. P., Nagler, M., & Pollen, D. A. Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *J Physiol* **365**(1), 331-363 (1985).

60. Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. Identifying natural images from human brain activity. *Nature* **452**(7185), 352 (2008).

61. Papale, P., et al. Foreground-background segmentation revealed during natural image viewing. *eNeuro* **5**(3) (2018).

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

63. 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).

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

65. 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).

66. 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).

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

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

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

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**. Time-varying radial plot displays group-level timeseries of the perceived intensity of six basic emotions. This model was obtained from averaging the preprocessed and temporally smoothed single-subject emotion ratings. **C**. Correlation matrix showing Spearman's ρ values for pairings of basic emotions. **D**. Principal Component Analysis: loadings and explained variance of the six principal components. **E**. 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. **F**. Time-varying radial plot shows the timecourse of the first three principal components: *polarity*, *complexity* and *intensity*. HA = Happiness, SU = Surprise, FE = Fear, SA = Sadness, AN = Anger, DI = Disgust, PC = Principal component, PO = Polarity, CO = Complexity, IN = Intensity.

Figure 2. 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, #5 and #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, TOS = Transverse Occipital Sulcus, 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 3. 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. CoG = Center of Gravity.

Figure 4. Characterization of Emotion Dimension Gradients in right TPJ

A. Average hemodynamic activity in right TPJ temporally related to the scores below and above the 50^{th} percentile for *polarity*. **B**. Average hemodynamic activity in right TPJ temporally related to the scores below and above the 50^{th} percentile for *complexity*. **C**. Average hemodynamic timepoints above the 75^{th} percentile and within 50^{th} and 75^{th} percentile for *intensity*. PC = Principal Component.