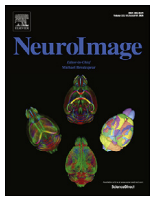
NeuroImage 222 (2020) 117258

Contents lists available at ScienceDirect

NeuroImage

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



Distinct dimensions of emotion in the human brain and their representation on the cortical surface

Naoko Koide-Majima[a], Tomoya Nakai[b,c], Shinji Nishimoto[b,c,d,*]

[a] Brother Industries Ltd., Aichi, Japan
[b] Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology, Osaka, Japan
[c] Graduate School of Frontier Biosciences, Osaka University, Osaka, Japan
[d] Graduate School of Medicine, Osaka University, Osaka, Japan

ARTICLE INFO

Keywords:
Emotion
fMRI
Semantic space

ABSTRACT

We experience a rich variety of emotions in daily life, and a fundamental goal of affective neuroscience is to determine how these emotions are represented in the brain. Recent psychological studies have used naturalistic stimuli (e.g., movies) to reveal high dimensional representational structures of diverse daily-life emotions. However, relatively little is known about how such diverse emotions are represented in the brain because most of the affective neuroscience studies have used only a small number of controlled stimuli. To reveal that, we measured functional MRI to obtain blood-oxygen-level-dependent (BOLD) responses from human subjects while they watched emotion-inducing audiovisual movies over a period of 3 hours. For each of the one-second movie scenes, we annotated the movies with respect to 80 emotions selected based on a wide range of previous emotion literature.

By quantifying canonical correlations between the emotion ratings and the BOLD responses, the results suggest that around 25 distinct dimensions (ranging from 18 to 36 and being subject-dependent) of the emotion ratings contribute to emotion representations in the brain. For demonstrating how the 80 emotion categories were represented in the cortical surface, we visualized a continuous semantic space of the emotion representation and mapped it on the cortical surface. We found that the emotion categories were changed from unimodal to transmodal regions on the cortical surface. This study presents a cortical representation of a rich variety of emotion categories, which covers many of the emotional experiences of daily living.

1. Introduction

We experience a rich variety of emotions in daily life, ranging from vivid feelings of anger or joy to more subtle sensitivities such as awe or empathetic pain. One of the central topics in affective neuroscience is to clarify how these various emotions are represented in the human brain. Most of the previous neuroimaging studies have addressed emotion representation in the brain (Chikazoe et al., 2014; Kragel and LaBar, 2014, 2016; Saarimaki et al., 2018; Saarimäki et al., 2016; Tettamanti et al., 2012; Wager et al., 2015; Winston, 2005) based on only a few to 15 categories (e.g., “fear,” “sadness,” and “happiness”) (Ekman, 2016; Johnson-Laird and Oatley, 1989; Lench et al., 2011; Mauss et al., 2005; Rottenberg et al., 2007), or rating scale for a few emotion dimensions (e.g., “arousal” and “valence”) (Barrett, 2006; Barrett and Bliss-moreau, 2009; Russell, 2009; Russell and Barrett, 1999). However, the emotional states referred to in previous studies might describe too confined a range to represent the full range of emotion we experience. It is unclear whether observations based only on these selected emotional states can be generalized to explain brain activity under more naturalistic conditions and how diverse emotions may be represented and mapped in the human brain.

Recent behavioral studies have revealed high dimensional organization of emotion representation using substantial behavioral data (Cowen et al., 2019, 2018; Cowen and Keltner, 2019, 2017). In these studies, researchers collected reports of emotional states elicited by a large number of emotion-inducing stimuli such as movie clips. The researchers revealed that the stimuli reliably elicited more than 20 distinct varieties of emotional experiences. When the reported emotional experiences were represented in a continuous space (a semantic space of emotion) where the distance between the experiences reflected the dissimilarity of the reports, the distribution of the reported emotional experiences displayed a smooth gradient between the categories of emo-

* Corresponding author at: Center for Information and Neural Networks, National Institute of Information and Communications Technology, Yamadaoka 1-4, Suita, Osaka 565-0871, Japan.
E-mail address: nishimoto@nict.go.jp (S. Nishimoto).

https://doi.org/10.1016/j.neuroimage.2020.117258
Received 13 December 2019; Received in revised form 2 August 2020; Accepted 5 August 2020
Available online 13 August 2020

tion. Hence, by such a data-driven approach using numerous emotion labels and stimuli such as movie clips, the researchers uncovered a more detailed representation of emotional states than that found by focusing on the specific emotional states based on the few to 15 emotional categories or few emotion dimensions. These behavioral studies suggested that an equivalently rich emotional stimuli may be used to probe comprehensive emotional representations in the brain, although no studies have yet been conducted to this effect.

Recent neuroimaging studies have also used continuous representational spaces to explain how perceptual categories such as semantics are organized in the brain (Huth et al., 2016, 2012). These studies used rich and naturalistic materials such as hours of movies or narrative stories as stimuli to reveal semantic representations of thousands of semantic categories by modeling the evoked brain activity using voxel-wise encoding models (Naselaris et al., 2011). Constructing semantic representational spaces enables us to examine the relative relationships between semantic categories and map the organization of these categories in a detailed manner across the cortex. By examining the estimated representations, smooth transition of category organization (Huth et al., 2012) and region domination of specific semantic domains (Huth et al., 2016) have been revealed. Such semantic space for emotion representation (= emotional semantic space) would provide a better understanding of the brain representation of emotion contrary to previous neuroimaging studies, which have focused on the localization of brain areas responsible for a restricted number of emotional states (Kragel and LaBar, 2014, 2016; Saarimaki et al., 2018; Saarimäki et al., 2016; Wager et al., 2015).

To provide a comprehensive understanding of the brain representation of emotion, we measured blood-oxygen-level-dependent (BOLD) responses in human subjects while they watched 3-h audiovisual movies consisting of 720 clips that were selected to induce various types of emotions. Each movie scene was rated by annotators in relation to each of the 80 emotion categories selected based on a wide range of previous reports (Plutchik, 1991; Russell and Barrett, 1999; Weidman et al., 2017; Strapparava and Valitutti, 2004; Cowen and Keltner, 2017; see Materials and Methods 2.5 Eighty Emotion Categories). Because we aimed to examine comprehensive representations of the emotions experienced in daily-life, we selected the 80 emotion categories to include as broad a range as possible. The ratings for the 80 emotion categories constitute greater variety than used in the previous fMRI studies (Chikazoe et al., 2014; Kragel and LaBar, 2014, 2016; Saarimaki et al., 2018; Saarimäki et al., 2016; Tettamanti et al., 2012; Wager et al., 2015; Winston, 2005). Using the emotion ratings and the evoked BOLD responses, we first examined how many dimensions of the emotional semantic space provided significant emotion to explain the BOLD responses. Next, to construct a semantic space for the emotion categories, we estimated the voxel-wise BOLD responses to the 80 emotion categories using a regularized linear regression model and applied a dimensionally-reduction technique to the emotion-category responses. Lastly, by performing a factor rotation analysis on the dimensions of a semantic space, we provided an interpretation of the dimensions and showed spatial transitions of each voxel's factor loadings in the cortex.

2. Materials and methods

2.1. Subjects

Eight healthy individuals (S1–S8; age 23-32; four females) with normal or corrected-to-normal vision participated in our experiments. We recruited subjects who consented to watch stimuli including extreme content, such as violent, disgusting or erotic representations. We confirmed their tolerance to the extreme content by showing them extra movie clips whose content was similar to those used in the fMRI experiment. All the subjects claimed not to feel nauseated after watching the extra movies, and none of the subjects had risk factors associated with fMRI scanning (metal implants, claustrophobia, pregnancy, experience of epileptic seizures, experience of head surgery). All the subjects accepted this and provided written informed consent. The ethics and safety committees of the National Institute of Information and Communications Technology approved the experimental protocol.

2.2. Experimental design

In our experiments, fMRI BOLD responses were recorded while subjects watched audiovisual stimuli. The stimuli consisted of 135 movie clips from a video-sharing site *vimeo* (https://vimeo.com/jp), which were selected to induce a rich variety of emotions. Examples of the movie genres were as follows: horror, violent drama, comedy, romance, fantasy, daily life scenes, and action movies. Movie clips were cut down to 10–20 s in length (mean of 15 s), and recreated as a sequence of stimuli by combining the selected clips in a random order.

The visual stimuli were presented at the center of a projector screen with 23.3 × 13.2 degrees of visual angle at 30 Hz. The audio stimuli were presented through MR-compatible earphones with an appropriate volume level for each subject. The subjects were instructed to watch the clips naturally as if watching TV show in daily life. For each subject, fMRI data were collected in 3 separate sessions over 3 or 4 days. Each session consisted of six movie-watching runs (each run lasting 610 s). A total of 18 runs were divided into 12 model training runs and 6 model testing runs. The model training runs were used to train encoding models and consisted of 480 different movie clips shown once each (total 7200 s). The model testing runs were used to assess model prediction accuracy and consisted of three different types of 900-s movie sequence shown four times each (total 3600 s). None of the movies in the training runs was shown in the test runs, thus the test run datasets can be used to assess the generalizability of the trained models to novel stimuli.

2.3. fMRI data acquisition

fMRI data were acquired using a 3T Siemens Trio TIM scanner (Siemens, Germany) with a standard Siemens 32-channel volume coil and a multiband gradient echo-planar imaging sequence (Moeller et al., 2010) (TR = 2000 ms, TE = 30 ms, flip angle = 62°; voxel size = 2 × 2 × 2 mm^3, matrix size = 96 × 96, 72 axial slices, FOV = 192 × 192 mm^2, multiband factor = 3). Anatomical data were collected on the same 3T scanner using T1-weighted MPRAGE (TR = 2530 ms, TE = 3.26 ms, flip angle = 9°, voxel size = 1 × 1 × 1 mm^3, matrix size = 256 × 256, 256 axial slices, FOV = 256 × 256 mm^2).

2.4. fMRI data preprocessing

The Statistical Parameter Mapping toolbox (SPM8, http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) was used to preprocess EPI data. We performed motion correction by aligning all of the EPI data to the first image from the first scan for each subject (Çukur et al., 2013; Nishida and Nishimoto, 2018; Nishimoto et al., 2011). For each voxel, responses were normalized by subtracting the mean response across all time points. Then, long-term trends were removed by subtracting the results of the median filter convolution (120-s time window). All analyses were performed on the individual brain space for each subject. We did not perform spatial smoothing on the individual imaging data. To define anatomical regions, for each subject, the cerebral cortex was segmented into 156 regions of the Destrieux atlas (Destrieux et al., 2010) by using FreeSurfer (Dale et al., 1999). The segmentation results in T1 space were registered to the EPI space using FreeSurfer functions, and each voxel was given one anatomical label.

2.5. Eighty emotion categories

In the present study, we rated movies in terms of 80 emotion categories, which were selected from various reports as follows: those posited as basic emotion theory (Plutchik, 1991) and core affective states (Russell and Barrett, 1999), those used in a recent meta-analytical

paper of emotion studies (Weidman et al., 2017), WordNet-Affect (Strapparava and Valitutti, 2004), and 27 emotion categories reported in a recent behavioral study (Cowen and Keltner, 2017). In addition, to select emotion categories that are relevant to emotional experiences induced by movies, we used affective labels used on the movie review site (https://movies.yahoo.co.jp/review/) as emotion categories. The 80 emotion categories are listed below: (1) love, (2) amusement, (3) craving, (4) joy, (5) nostalgia, (6) boredom, (7) calmness, (8) relief, (9) romance, (10) sadness, (11) admiration, (12) aesthetic appreciation, (13) awe, (14) confusion, (15) entrancement, (16) interest, (17) satisfaction, (18) excitement, (19) sexual desire, (20) surprise, (21) nervousness, (22) tension, (23) anger, (24) anxiety, (25) awkwardness, (26) disgust, (27) empathic pain, (28) fear, (29) horror (bloodcurdling), (30) laughing, (31) happiness, (32) friendliness, (33) ridiculousness, (34) affection, (35) liking, (36) shedding tears, (37) emotional hurt, (38) sympathy, (39) lethargy, (40) empathy, (41) compassion, (42) curiousness, (43) unrest, (44) exuberance, (45) appreciation of beauty, (46) fever, (47) scare (feel a cill), (48) daze, (49) positive-expectation, (50) throb, (51) sexiness, (52) indecency, (53) embarrassment, (54) oddness, (55) contempt, (56) alertness, (57) eeriness, (58) positive-emotion, (59) vigor, (60) longing, (61) tenderness, (62) pensiveness, (63) melancholy, (64) relaxedness, (65) acceptance, (66) unease, (67) negative-emotion, (68) hostility, (69) levity, (70) protectiveness, (71) elation, (72) coolness, (73) cuteness, (74) attachment, (75) encouragement, (76) annoyance, (77) positive-fear, (78) aggressiveness, (79) distress, and (80) stress

2.6. Emotion ratings and preprocessing

We collected ratings regarding each of the 80 emotion categories induced upon exposure to the movie stimuli used in our fMRI experiments. Note that we selected these 80 categories in an inclusive manner, where we aim to collect samples to probe as small potential differences as possible, instead to examine the category set that has known to be distinctive in the existing literature. By using these many categories, our modeling approach aims to examine how many of these categories indeed form distinctive dimensions from the aspect of brain representations.

To obtain the emotion ratings, we recruited 166 annotators who did not participate in the fMRI experiments. We instructed the annotators to rate how well an emotion category (e.g., “laughing”) matched the feeling they experienced in response to the movie scene, by assigning a value ranging from 0 (not matched at all) to 100 (perfectly matched). We also instructed the annotators to make ratings based not on the movie character’s feeling, but on their (the annotators’) own feeling. In the rating procedure, each annotator rated the matching degree every moment by dragging a mouse while watching the movie stimuli, and we stored the ratings at 1-s resolution. This procedure ensured that the ratings were based on the annotators’ emotional responses to the given movie scenes in a similar way to that of the experimental subjects during the fMRI experiments. For each of the 80 emotion categories, we collected four independent ratings by assigning four different annotators. To obtain all 80 emotion categories × 4 independent ratings, we recruited the total of 166 annotators, each of the 166 annotators gave ratings for one or two emotion categories. We assigned four different annotators to each emotion category, to obtain four independent ratings. To ensure the consistency of ratings, each annotator rated for each emotion category while watching 135 minutes of the entire 3-hour movie (note that the test movie stimuli included four repeats of 15-min movie scenes, and we collected annotations for unique movies. Thus the total duration of annotations was 135 minutes (120 minutes training movies and 15 minutes test movies) out of the 3-hour movie stimuli presented in the scanner). When one annotator rated for two emotion categories, the annotator first rated for one emotion category (e.g. “disgust”) while watching the whole movie stimuli. The same annotator then rated for another emotion category (e.g. “satisfaction”) while watching the whole movie stimuli again. To obtain reliable ratings, we first conducted an aptitude test for each candidate annotator. Specifically, we used a 246-s test movie and examined the consistency of temporal fluctuation of ratings between a template by one of the authors (NK) and each annotator in terms of each of two emotions “fear” and “disgust.” From the results, we recruited only candidates whose ratings showed significantly high correlations with the template ratings for both emotions ($p < 0.05$ for 246 time samples). The results of this exercise allowed us to recruit 166 annotators (correlation coefficients for fear: mean = 0.48, SD = 0.15; for disgust: mean = 0.61, SD = 0.15). The preprocessing procedure for the ratings is as follows. The ratings for each annotator and each emotion were de-trended by subtracting the results of median filtered signals (300 s time window). For each emotion category, the preprocessed ratings were averaged across those for the four annotators at 2-s resolution (i.e. BOLD sampling rate) and z-score normalized.

After collecting the ratings from the recruited annotators, we further examined the consistency of the ratings across annotators as a measure of the reliability of the ratings. For each emotion category, we quantified the consistency of the ratings as Pearson’s correlation coefficient between the ratings of each annotator and those averaged across the remaining three annotators. We obtained four correlation coefficients from four annotators, and averaged them. The mean of the correlation coefficients across four annotators and 80 emotion categories was 0.41 (Fig. S1), suggesting a certain level of consistency of the ratings across annotators.

2.7. Model fitting

To estimate the BOLD-response patterns to 80 emotion categories, we constructed a voxel-wise linear regression model to explain the BOLD responses (Huth et al., 2012; Nishimoto et al., 2011). We used L2-regularization in the regression model because it has been used in encoding models from previous neuroimaging studies that demonstrated a continuous space of entities using the model weights (Huth et al., 2016, 2012; Nakai and Nishimoto, 2020), similar to the present study. The stimulus vector (input vector) included the annotator-averaged emotion ratings (80 dimensions corresponding to 80 emotion categories) and sensory factors (visual and auditory features of 1000 dimensions each). We extracted these sensory factors from frames and the soundtrack of the movie stimuli, and used them to remove spurious correlation caused by visual and auditory responses in the BOLD signals (see “2.8 Removing spurious correlation between emotion ratings and sensory information” in Materials and Methods). We concatenated the emotion ratings with the sensory factors and used the resulting 2,080-dimensional stimulus vector as regressors. Further, the stimulus vector was concatenated with three temporal delays of 2, 4 and 6 s (total 6,240 dimensions) to capture the hemodynamic response. The model weights were optimized by least squares with L2-regularization. The regularization coefficient (γ) was optimized in 10-fold cross validation using 10 unique training-validation (9:1) subsets by randomized sampling from the training data (3,600 samples) of emotion ratings and BOLD responses. In each cross-validation step, we then constructed the regression model using a training subset and computed the prediction accuracy using a validation subset for each γ of 2^i, where $i = \{0, 1, \ldots, 25\}$. The prediction accuracy was quantified as an across-voxel average of the correlation coefficients between the actual and predicted training BOLD responses. We employed the best γ showing the highest accuracy across 10 repetitions within the cross validation. After the cross validation, we constructed the regression model with the best γ using the training data (3600 samples) and computed the prediction accuracy using the test data (1800 samples). In our main analyses, we used voxels with high prediction accuracy of the emotion-related BOLD responses after regressing out the sensory factors. In analyses using canonical correlation analysis (CCA) (see “Emotion dimensions based on the BOLD-response patterns” in Materials and Methods), we employed voxels with high prediction accuracies (uncorrected $p < 0.0001$) averaged across the 10 folds (3984–13,068 cortical voxels per subject). To construct the semantic space of emotion, we employed

voxels with high prediction accuracy (uncorrected $p < 0.0001$) for the test data (18,684–34,066 cortical voxels per subject).

2.8. Removing spurious correlation between emotion ratings and sensory information

To estimate unalloyed responses to emotion categories, spurious correlation with a sensory factor was explained away from model prediction for the BOLD responses to 80 emotion categories (Huth et al., 2012). For this purpose, we employed low-level visual and auditory features as the sensory factors and used them to fit the linear regression model, but these were not used in the model prediction. We extracted the visual and auditory features from frames and the soundtrack of the movie stimuli used in our experiments.

As low-level visual features, we employed the output of 2139 motion energy filters (Nishimoto et al., 2011). Each filter consists of quadrature pairs of spatiotemporal Gabor filters. Input frames were obtained at 15 Hz and resized from 720 × 1280 × 3 to 96 × 171 × 3, followed by cropping in the center to a size of 96 × 96. Then, the image was converted from RGB color to (CIE) L*A*B* color space, and the color information was discarded. The motion energy signals were yielded from the filter output and then log-transformed and averaged across 2 s (TR). Consequently, we obtained 2139 visual features, which represent preferences to spatial frequencies, temporal frequencies, and orientations. To minimize the computational burden, we reduced the original dimensions to 1,000 using singular value decomposition. These 1000 components preserved 83.8% of the variance explained in the original features.

As the low-level auditory features, we employed the output of the modulation-transfer function model (Chi et al., 2005). The spectrogram was generated using 128 bandpass filters (Ellis, 2009) with a window size of 25 ms and hop size of 10 ms. Then, the spectrogram was convolved with quadrature pairs of modulation-selective filters for 10 spectral modulation scales and 10 temporal modulation rates. The modulation energy was calculated using the same methods as reported by Nishimoto et al. (2011). Modulation energy was log-transformed, averaged across 2 s (TR), and further averaged within each of the 20 nonoverlapping frequency ranges logarithmically spaced in the frequency axis. From the results, we obtained 2,000 auditory features, which represent preferences to frequencies of audio signal and the temporal variation of the preference frequencies. As in the visual feature extraction, we reduced the original dimensions to 1000 to minimize the computational burden. The 1000 components preserved 93.4% of the variance explained in the original features.

As supplementary results, we showed the prediction accuracies of the BOLD responses using each of the following three features: emotion, visual, and auditory (Fig. S2). The accuracy was quantified using Pearson's correlation coefficients between the actual and predicted responses for the reserved test dataset of emotion ratings and BOLD responses that was not used to train the encoding models. We confirmed high prediction accuracy for the early visual and early auditory cortex from the visual and auditory features, respectively. This indicates that these two features could plausibly explain BOLD responses in the early visual and auditory cortices.

2.9. Emotion dimensions based on the BOLD-response patterns

To examine whether a large number of emotion labels, 80, is beneficial to explaining the BOLD responses, we estimated the number of correlated dimensions between BOLD responses and emotion ratings. To obtain the correlated dimensions, CCA was performed based on correlation in temporal fluctuations between ratings (delayed for 4 s) of each emotion category and the BOLD responses of each subject. In the CCA, the two types of factor loadings (A, B) were estimated to have the maximum correlation between the linear combinations of emotion ratings and the BOLD responses. Using the training data of emotion ratings and BOLD responses, we estimated the factor loadings (A^*, B^*) as follows:

$$A^*,\ B^* = \underset{A,\ B}{\operatorname{argmax}}\ corr(S \cdot A,\ \hat{R} \cdot B),$$

where $corr(i, j)$ denotes correlation coefficients between i and j. S (3600 × 80) is the emotion ratings. $\hat{R}$ (3600 × 1800) is the dimension-reduced BOLD responses of voxels showing good prediction accuracy (correlation coefficients, $p < 0.0001$) in 10-fold cross-validation in the linear regression model (see "Model fitting" in Materials and Methods). To validate the estimated A^* and B^*, we used test data of $\hat{R}$ (1800 × 1800) and S (1800 × 80)). Then, we quantified the significance of each dimension of A as a Pearson's correlation coefficient between each dimension of $S \cdot A$ and that of $\hat{R} \cdot B$, and counted the number of significant dimensions. The significance was defined by the statistical significance ($p < 0.01$; with Bonferroni correction for 80 emotion categories). To show how the number of significant dimensions depends on the number of emotion categories, we also estimated the number of significant dimensions for ratings when we used a restricted number (from 1 to 80) of emotion categories. For each restricted number of emotion categories, we obtained 50 sets of ratings by random sampling to estimate the correlated dimensions with the BOLD responses for each subject.

2.10. A semantic space for brain representation of 80 emotion categories

To construct a semantic space, we applied the dimension reduction technique to the BOLD-response patterns corresponding to the 80 emotion categories. Those BOLD-response patterns were estimated from emotion-category weights (18,684–34,066 cortical voxels × 240 regressors (80 emotion categories × 3 temporal delays)) of the regression model using voxels with high prediction accuracies ($p < 0.0001$, uncorrected; the number of significant voxels ranged from 18,684 to 34,066 cortical voxels for each subject). The emotion-category weights were averaged across three temporal delays and the averaged results (18,684–34,066 cortical voxels × 80 emotions) were vertically concatenated across the eight subjects. We termed them group weights and used as the BOLD-response patterns corresponding to the 80 emotion categories. We then used PCA to reduce the vertical (voxel) dimensions of the group weights to 25, which was determined as the number of emotion dimensions that contribute to explain the original BOLD responses (see results of the CCA). We defined the PCA space as the semantic space of emotion. Then, we conducted a varimax factor rotation on the principal components, to obtain an interpretation of each emotion dimension. Factors with high negative loadings were rotated to have the opposite direction by multiplying by −1. The rotated components were used as 25 emotion dimensions in analyses for transition of emotion representation in the cortex.

To confirm across-subject consistency of a semantic space, we performed a leave-one-subject-out method. For this, we constructed a 25-dimensional semantic space for a single subject (individual space) and also that for the weights concatenated across the remaining seven subjects (sub-group space). As a control semantic space, we also constructed a 25-dimensional semantic space by performing PCA (category-dimension reduction) on the emotion ratings of the training data. To support the across-subject consistency of individual spaces, we showed that the similarity to the sub-group space was higher than that to the control space for each subject. To quantify the similarity between semantic spaces, we calculated a Pearson's correlation coefficient in emotion distributions between each semantic space pair. The emotion distribution was quantified as pair-wise distances (correlation distance) of emotion categories in each semantic space.

2.11. Spatial transition of emotion dimensions in the cortex

To visualize how emotion-category weights were distributed through the cortical surface, we used the first three components of the semantic space to obtain RGB color values. Emotion-category weights of each

voxel were first projected to the semantic space consisting of the three components. Then, the projected coordinates were normalized to a range of 0–1 by z-scoring and linear scaling. The output result was used as an RGB color projected to voxel coordinates in the flattened cortical map. We can observe the weight transition of emotion categories as the RGB color transition on the cortical map. Furthermore, we assigned an RGB color to each emotion category by using three normalized components for each emotion category. This color was used for visualization of the semantic space in Fig. 2a. Each of the emotion-category colors implies an association with regions which have a similar color in the flattened cortical map.

In the cortical map for emotion-category weights, we observed a smooth transition in the following four areas: the postcentral area, the superior temporal area, the inferior parietal area, and the precuneus. To quantify the weight transition, for each voxel, we first computed the weight of the 25 emotion dimensions by multiplying the 80 dimensional weights [1 × 80] and the 25 emotion dimensions [80 × 25]. Then, we manually defined eight lines on a smooth transition in the four areas at each of the left and the right hemispheres (see Fig. 4a), with reference to the flattened cortical map using the in-house Matlab (MathWorks Inc.) GUI toolbox. To confirm that each line was located at similar anatomical location, we obtained the anatomical label of the Destrieux atlas (Destrieux et al., 2010) from each line for each subject (Fig. S3). The label was defined based on anatomical location using Freesurfer (Dale et al., 1999).

3. Results

3.1. Twenty-five dimensions of emotion ratings were significantly correlated with the BOLD responses

In this study, we aimed to elucidate the representations of a rich variety of emotions in the human brain using 80 categorical emotion labels. With this aim, we first used canonical correlation analysis (CCA) to examine whether this large number of emotion labels is beneficial to explaining the evoked BOLD responses and determine the number of distinct dimensions of emotion that contribute to the representations (Cowen et al., 2018; Cowen and Keltner, 2019, 2017). Specifically, we calculated the canonical correlation between the emotion ratings and the BOLD responses of the training dataset from each of the eight subjects, where factor loadings for each of the ratings and the responses were estimated to have the maximum correlation between them (see Materials and Methods). Then, we tested whether factor loadings demonstrated a significant correlation ($p < 0.01$, with Bonferroni correction) for each dimension of the ratings using a test dataset. To investigate how the number of significant dimensions we could detect was dependent on the number of emotional categories involved in the analysis, we sampled ratings for the restricted number (from 1 to 80) of emotion categories and quantified the number of significant dimensions (see Materials and Method). The results showed that the number of significant dimensions gradually increased in ascending order of the number of emotion categories (Fig. 1b). The number of significant factor loadings (= significant dimensions) using the full 80 emotion ratings ranged from 18 to 36 across subjects (S1: 18; S2: 31; S3: 26; S4: 26; S4: 24; S6: 36; S7: 24; S8: 24) and the median was 25. These findings suggest that a rich variety of up to at least 80 emotion categories is informative, as shown by the monotonic increase in the number of significant dimensions (Fig. 1b, right) to explain the BOLD responses, and that around 25 dimensions of the emotion ratings contribute to explaining the BOLD response patterns elicited with statistical significance using this experimental paradigm. In addition, we also provided results for the number of dimensions calculated at a different statistical significance ($p < 0.05$, Bonferroni correction for 80 emotion categories). In the case, we found the median across subjects was 30 for 80 emotion categories (Fig. S4).

3.2. A semantic space shows brain representation of 80 emotion categories

To construct a semantic space of emotion, we used an encoding modeling framework (Huth et al., 2012; Naselaris et al., 2011). First, an L2-regularized linear regression was performed on BOLD responses using the 80 emotion ratings for each movie scene as regressors (see Materials and Methods). The estimated weights from individual subjects were concatenated and was reduced the dimensionality from the original 80 to 25 using principal component analysis (PCA). The first 25 components explained 65% of the variance (Fig. S5a). The resultant number, 25, was accepted as a sufficient number of emotion-category dimensions to explain the BOLD responses, as suggested by the previous CCA (Fig. 1b). The semantic space of emotion was defined as the space consisting of the 25 dimensions. To maintain the quality of the semantic space, we used only voxels with high prediction accuracy in the regression model (see Materials and Methods).

In the semantic space of emotion, the distance between each pair of emotions represents the dissimilarity in the BOLD-response patterns between them. Fig. 2a shows the semantic space projected into two-dimensional space, while preserving the emotion-pair distances in the 25-dimensional space as much as possible. The color of each emotion category represents the first, second, and third PC coefficient as its red, green, and blue channel luminance, respectively. Emotion categories assigned dissimilar colors and distant locations were considered to have distinct brain representations. To examine how the representations of the 80 emotion categories are related, we performed hierarchical clustering on the coordinates of emotion categories in the semantic space (Fig. S6). Several clusters were found to include semantically similar emotion categories, namely Nervousness (including “nervousness” and “unease”), Curiousness (“curiousness,” “ridiculousness,” and “confusion”), Awkwardness (“awkwardness” and “hostility”), Joy (“joy,” “happiness” and “love”), Lethargy (“lethargy,” “nostalgia,” and “boredom”), Awe (“awe” and “aesthetic appreciation”), Sexual desire (“sexual desire” and “indecency”), Disgust (“disgust,” “empathic pain.” and “scare”), Interest (“interest,” “empathy,” and “emotional hurt”), and Fever (“fever,” “exuberance,” and “vigor”). For higher nodes of the hierarchy, the clusters with semantically similar meanings were combined into more abstract clusters as follows: emotions related to getting attention (Nervous-scared, Curiousness, and Awkwardness-related), peaceful emotions (Happiness, Amusement, and Awe-related) and stimulating emotions (Sexual desire and Disgust), and emotions related to being absorbed in (Interest and Fever). Distinct distributions for largely positive (Fig. 2a *upper part*) and negative (Fig. 2a *lower part*) emotions were also determined (see Fig. S7).

To interpret each principal component of the 25-dimensional semantic space, we performed varimax factor rotation (Fig. 2b). Each factor was shown to be weighted more highly for associated emotion categories such as “fear” and “horror.” Fig. 2c presents factor loadings for the four representative factors, where the top-left factor (ID1) contains high weights for “Disgust”-related emotions such as “disgust,” “empathic pain,” and “fear.” The top-right factor (ID2) contains high weights for “Awkwardness”-related emotions such as “awkwardness” and “laughing.” The bottom-left factor (ID4) contains high weights for “Sexual-desire”-related emotions such as “sexual-desire,” “sexiness,” and “levity.” The bottom-right factor (ID19) contains high weights for “Emotional-hurt”-related emotion categories such as “compassion,” “emotional hurt,” and “sadness.” Interpretations of the other factors are shown in Fig. S5b. We found that each dimension was associated with a specific emotion category and more than half of the factors corresponded to emotion dimensions addressed in the study by Cowen and Keltner (2017) as follows: “Awkwardness,” “Sexual desire,” “Entrancement,” “Amusement,” “Adoration (Friendliness),” “Boredom,” “Anger,” “Admiration,” “Joy,” “Emotional hurt (Sadness),” “Interest,” “Craving,” and “Nostalgia.” The other dimensions could also be interpreted as being similar to those in the report by Cowen and Keltner (2017), although the groupings were different. In our analysis, negative emotions such

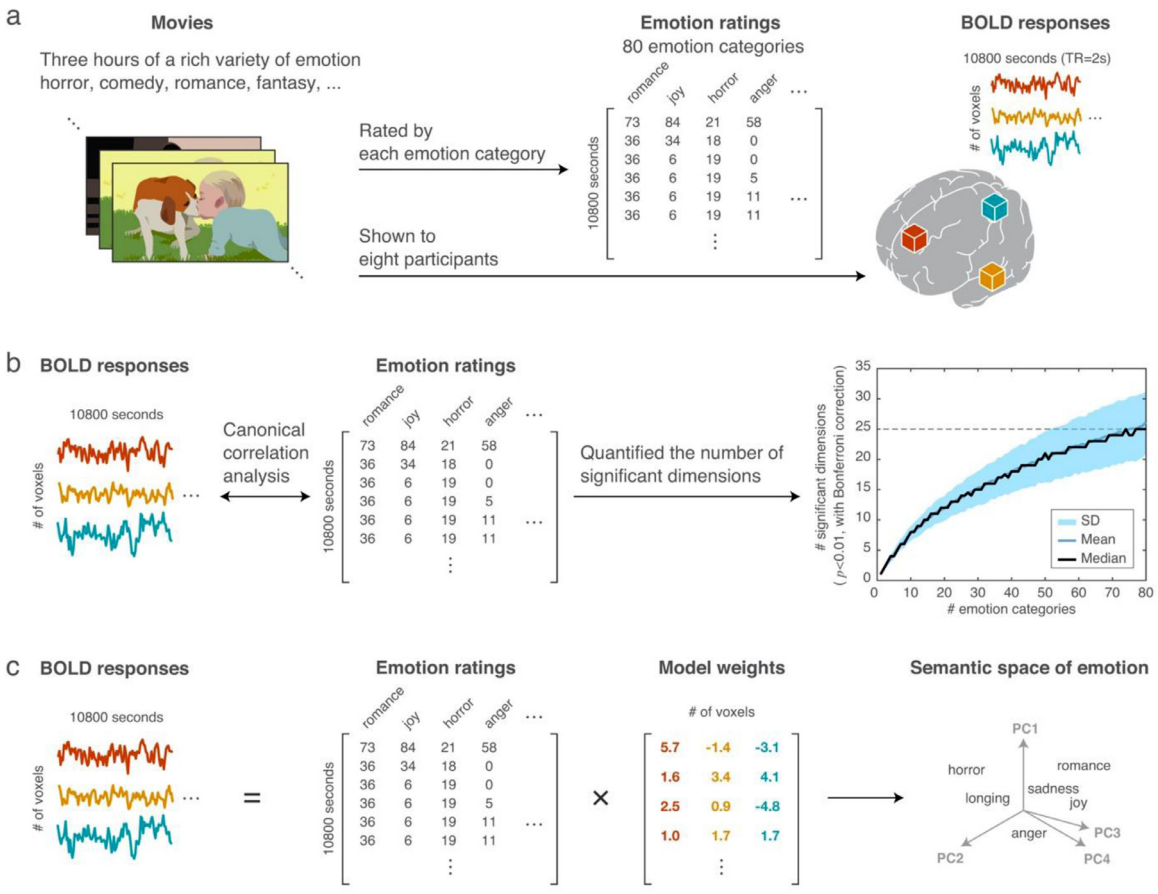


Fig. 1. Schematic of the experiment and procedure for constructing a semantic space of emotion. **a** BOLD responses for eight subjects were measured while they watched emotion-inducing movies for 3 h. Each movie scene was rated according to 80 emotion categories. **b** Canonical correlation analysis was performed to examine how many distinctive dimensions of the 80 emotion categories were informative to explain the associated BOLD responses. **c** The voxel-wise response was modeled as a linear weighted sum of the emotion ratings using an L2-regularized regression procedure. A semantic space was constructed by performing a dimension reduction on the estimated model weights.

as "disgust," "empathic pain," and "fear" were combined into one dimension, "Disgust." Furthermore, "tension," "nervousness," "scare," and "positive-fear" were combined into one dimension, "Nervous scared," whereas distinct dimensions were reported for the associated emotions ("Anxiety," "Horror," and "Fear") in the study by Cowen and Keltner (2017). In terms of positive emotions, we found some "Excitement"-related emotions ("excitement," "entrancement," "exuberance," "encouragement," and "fever") that provided low contributions to explaining the BOLD responses (Fig. S8). Differences in the contributions (the model weights) may be due to the relationships between the contributions and reliability of the emotion ratings across annotators. To examine these relationships, we first averaged the model weights across subjects and voxels and quantified the rating reliability as the consistency of the ratings across annotators (see "Emotion ratings and preprocessing"). We then quantified the relationship as Pearson's correlation coefficients between the averaged model weights and the rating reliabilities. The resultant correlation coefficient was 0.181 ($p \geq 0.05$), suggesting that such relationships were not the major reason for differences in the contributions across emotion categories.

The semantic space thus far was constructed from the averaged model across all subjects. To investigate whether the semantic space was consistent across individual subjects, we quantified subject-wise semantic spaces and examined their across-subject similarity (Fig. 2d). To this aim, we calculated the correlations between the representational distance matrix (Kriegeskorte et al., 2008) of the semantic space from one subject and that from the remaining seven subjects. We found that correlations between the individual and the group semantic spaces were high for all subjects (0.88–0.94, the mean is 0.92). Then, to examine whether the across-individual consistency of the semantic space was not explained by the brain representation of sensory information, we constructed individual and group semantic spaces from the voxel weights, excluding those in the early visual and auditory cortices (Fig. S9a). We defined the visual and the auditory cortices as regions with high prediction accuracies in explaining the BOLD responses from the visual and auditory features (see "Removing spurious correlation between emotion ratings and sensory information" in Materials and Methods). Also in this case, we found that the correlations between the individual and the group semantic spaces were high for all subjects (0.82–0.90, the mean is 0.87, Fig. S9b). Then, to verify that the inter-subject consistency of the semantic space was not from the stimulus features (the emotion ratings), we also computed the correlations between the representational distance matrix from each subject and that from the emotion ratings. We found that correlations between the individual and the group semantic spaces were higher than those between the individual and the emotion-rating semantic spaces (Fig. 2d magenta), suggesting that the inter-subject consistency was not from the stimulus features. In sum, our results suggest that the individual semantic space was consistent across the subjects and that the group semantic space could be used as a representative space for all subjects.

3.3. Spatial transition of 25 emotion dimensions in the cortex

To examine how the emotion dimensions were represented on the cortical surface, in the first instance, we assigned an RGB color according

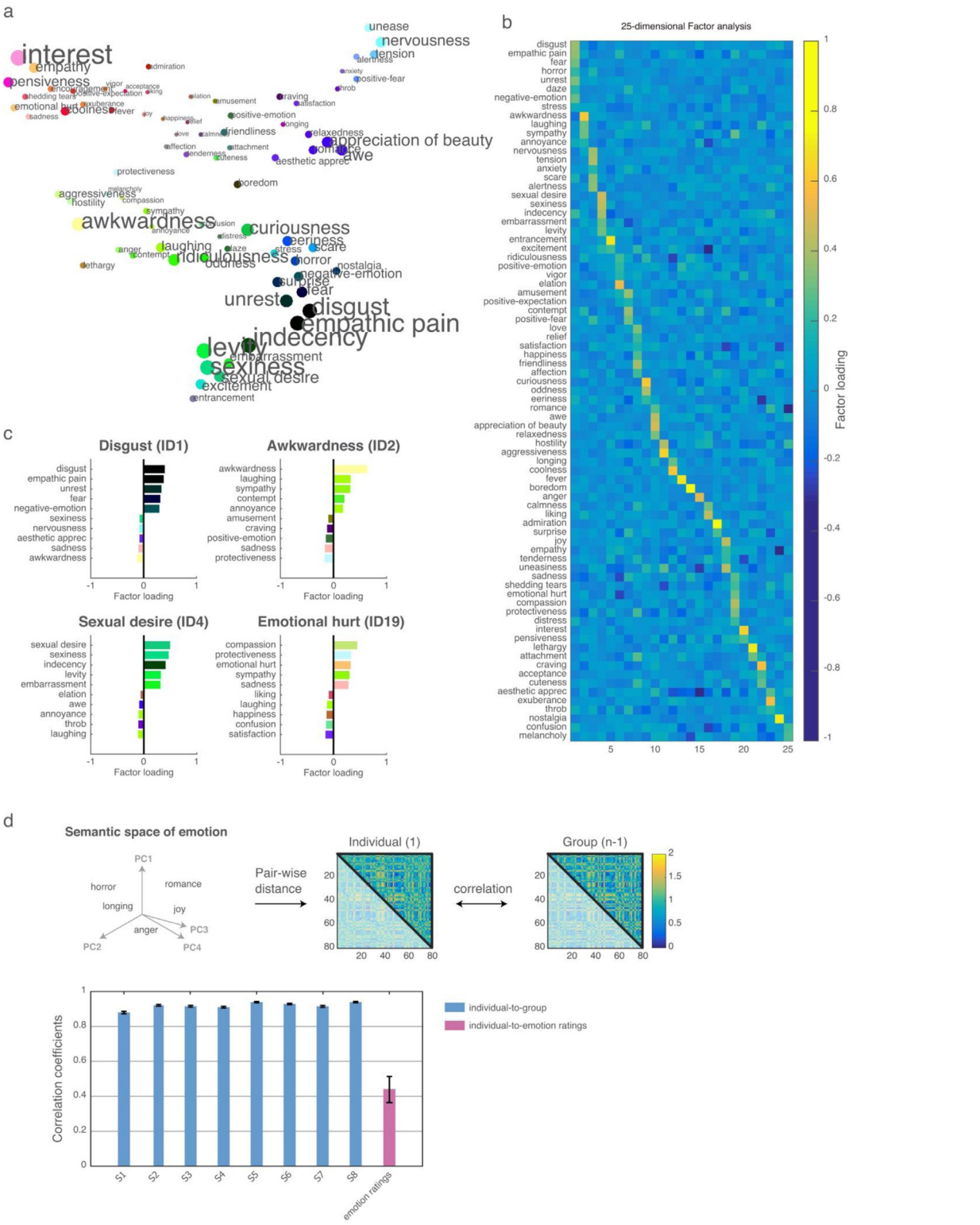


Fig. 2. A semantic space of emotion and varimax factor loading. **a** Eighty emotion categories were plotted according to distances between the estimated weights of the L2-regularized regression model, using *t*-distributed stochastic neighbor embedding (*t*-SNE Maaten and Hinton, 2008). A single RGB color was assigned to each emotion category according to the three main components of the estimated weights (see Materials and Methods). Both marker size and font size were modulated according to the average weight across voxels (larger size denotes higher average weight). **b** Factor loadings of 25 components (= 25 emotion dimensions) of the estimated weights after varimax rotation. **c** Factor loadings of four representative components: "Disgust," "Awkwardness," "Sexual desire," and "Emotional hurt." The bars indicate the five highest and lowest factors. **d** Inter-subject consistency of the semantic space. The individual consistency was quantified as the correlation between the representational distance matrix (emotion-pair-wise distances) of the individual semantic space and that of the averaged matrix for the remaining subjects. As a control to examine if the observed consistency was explained by the stimulus matrix (emotion rating), the individual-to-emotion ratings consistency was quantified as the correlation between the distance matrix of the individual semantic space and those of the emotion-rating semantic space. The bottom figure shows the individual-to-group correlations (Blue bar for each subject) and the individual-to-emotion ratings correlations (Magenta bar for all subjects). Errorbars denote 95%-confidence intervals obtained by the bootstrap method.

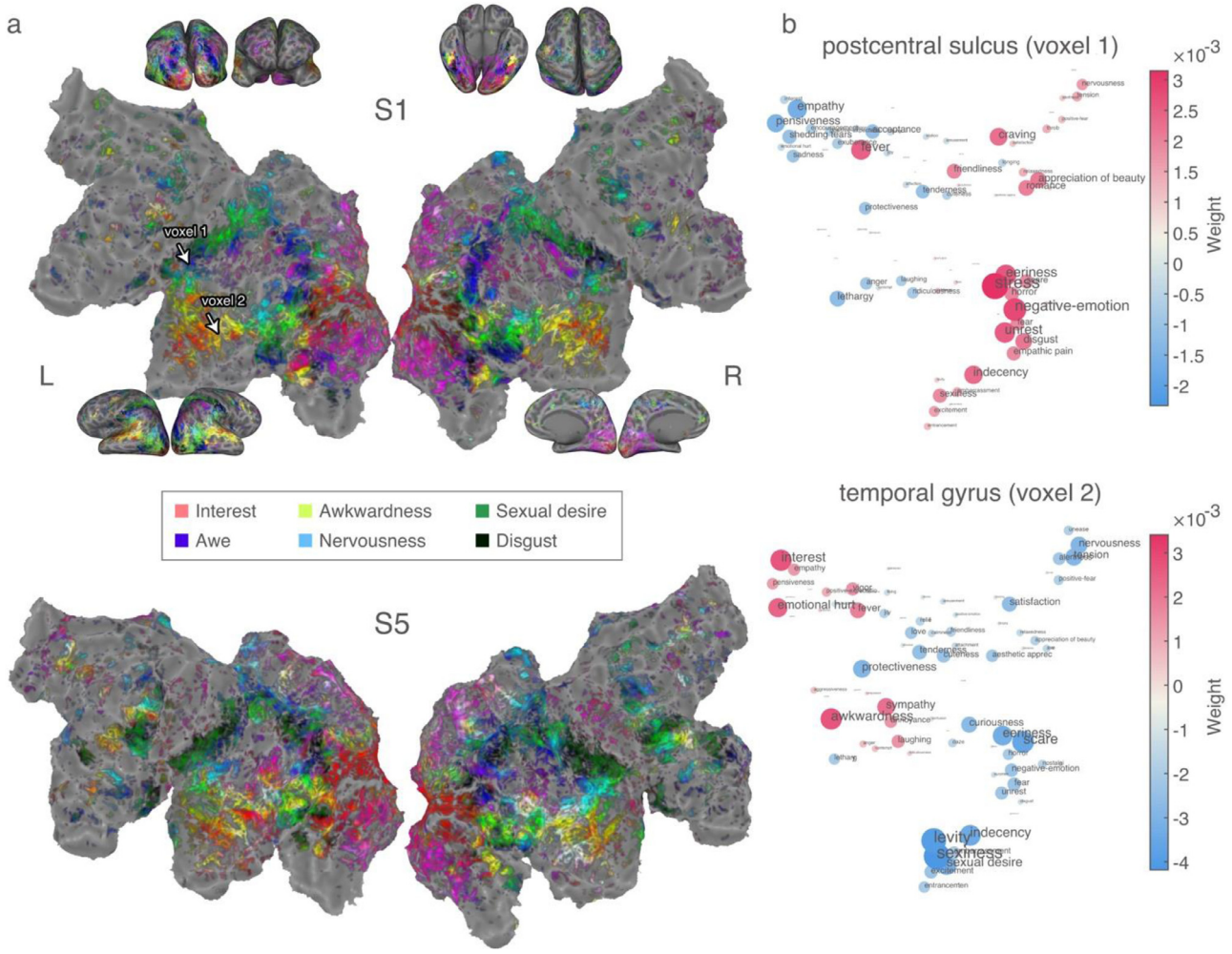


Fig. 3. Cortical map of the semantic space of emotion. **a** Cortical maps for two subjects (S1 and S5). Emotion representations were visualized by assigning RGB colors according to the three main components of the semantic space (see Materials and Methods). The legend presents six representative colors and the corresponding emotion categories. In each cortical map, we show the results only for voxels showing significantly high prediction accuracy ($p<0.0001$, uncorrected). **b** Examples of the weight distribution for two voxels in the postcentral sulcus and the temporal gyrus of a single subject (S1). The positions of the two voxels are indicated on the cortical map of S1. Red denotes positive weight and blue denotes negative weight. Positions of emotion categories were the same as the positions in Fig. 2a.

to the first to third principal components of the semantic space (see Materials and Methods). The top three components were selected for their visibility and were regarded as the representatives to demonstrated how the weights of the 25 emotion dimensions changed smoothly over the cortical surface. The cortical maps for two subjects (S1 and S5) are presented in Fig. 3a (The cortical maps for the other subjects are provided in Fig. S10). In both of the cortical maps, we found a weight transition, shown as a color transition, over the whole cortical surface. For example, different colors were assigned to the postcentral and superior temporal areas. Indeed, the weight distributions of 80 emotion categories differed between the two representative voxels obtained from these two areas (Fig. 3b). Specifically, one voxel from the postcentral sulcus contained high weights in "Disgust"-related emotions such as "empathic pain" and "unrest." The other voxel in the superior temporal gyrus contained high weights in "Interest" and "Awkwardness"-related emotions such as "interest," "empathy," "laughing," and "awkwardness." This suggests that the weight variability across voxels was visualized on the cortical map.

Then, we investigated how the weights of the emotion dimensions varied across the cortical surface, particularly in the following four areas depicting consistent color transitions across subjects: namely, the postcentral, the superior temporal, the inferior parietal areas, and the precuneus (Fig. 4b). In each of these four areas, we manually defined eight lines (line 1–8) and obtained the weights of the 25 emotion dimensions at 30 successive positions on each line. We determined the number, 30 as the number at which we could show voxel-level spatial transition of the weights. The cortical maps of the two subjects in Fig. 4a show these lines. To show spatial transition of the weights on each line, we plotted the weights obtained from each line in a sequential order of spatial coordinates on the flattened map (Figs. 4b and S11). We consistently observed specific weight transitions on each line across subjects, as suggested by the relatively small standard deviations across subjects compared to the difference among emotion dimensions. The postcentral area demonstrated high weights for negative emotions such as "Disgust" and "Nervous scared" over successive positions. "Sexual desire" also exhibited high weight and showed high correlation in the weight transition with "Disgust" (line1: $r = 0.428$, line2: $r = 0.648$). The superior temporal area showed a high weight transition from "Interest" to "Awkwardness" and "Aggressiveness," whereas "Aggressiveness" exhibited similar weight transition to "Nervous scared" (line3: $r = 0.95$, line4: $r = 0.97$), "Sexual desire" (line3: $r = 0.77$, line4: $r = 0.88$), and "Curiousness" (line3: $r = 0.90$, line4: $r = 0.95$). The inferior parietal area showed a high weight transition from "Disgust" to "Awe" and "Interest." "Disgust" demonstrated a similar weight transition to "Sexual desire" (line5: $r = 0.79$, line6: $r = 0.75$) and "Curiousness" (line5: $r = 0.95$,

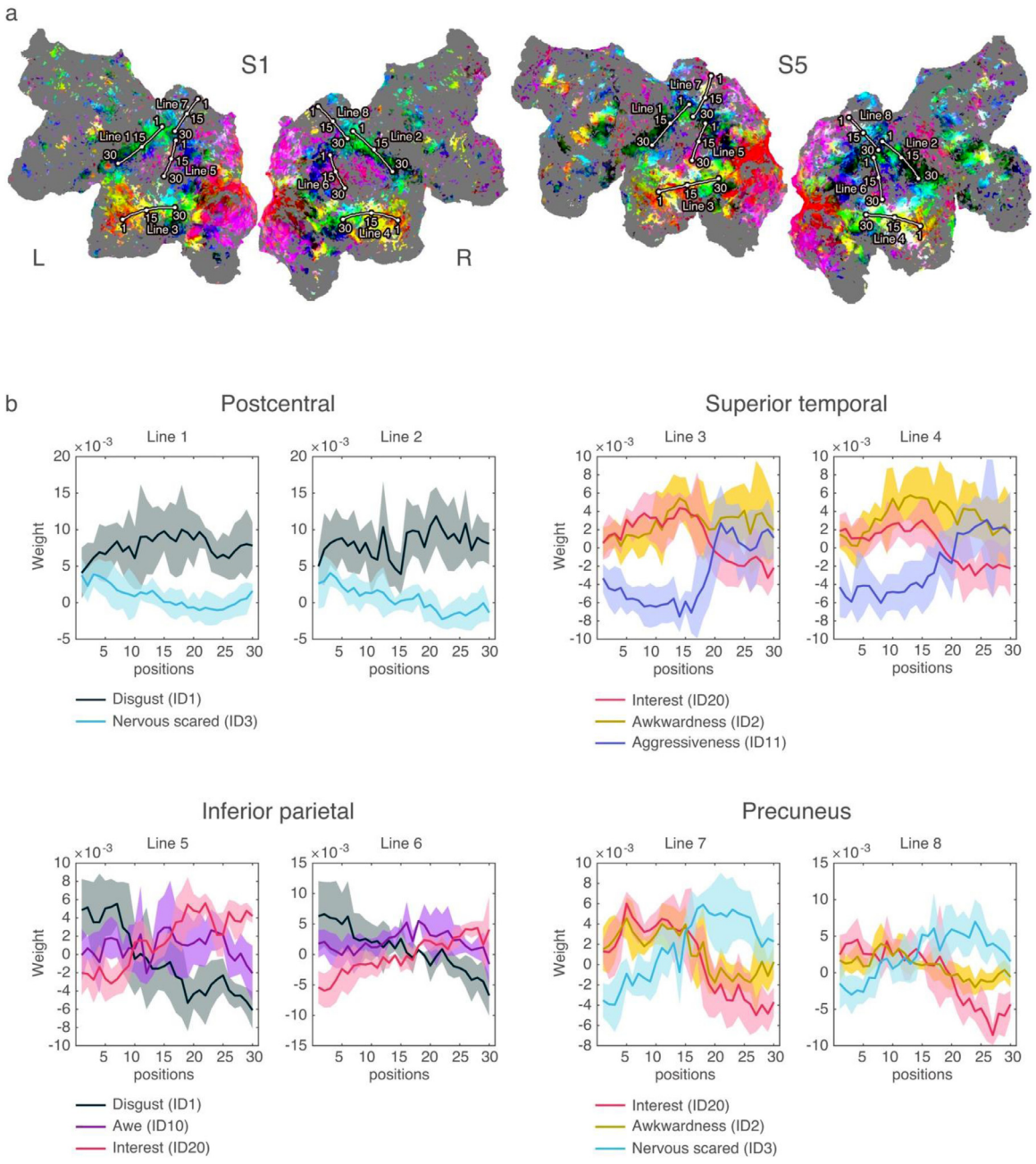


Fig. 4. Weight transition on the flattened cortical surface. **a** The positions of eight representative transitions are depicted as lines for two subjects (S1 and S5). Lines 1 and 2 were mainly located in the postcentral area, Lines 3 and 4 were mainly located in the superior temporal area, Lines 5 and 6 were mainly located in the inferior parietal area, and Lines 7 and 8 were mainly located in the precuneus. The anatomical location of each line was consistent across subjects, as supported by the results shown in Fig. S3. **b** Spatial transitions of representative (high absolute weights) emotion dimensions are plotted for each line. A bold plot denotes the mean contribution across subjects. Shaded areas denote the standard deviation of contributions across subjects. The weights of all the emotion dimensions are shown in Fig. S11.

line6: $r = 0.90$). "Interest" showed similar weight transitions to "Emotional hurt" (line5: $r = 0.93$, line6: $r = 0.77$) and "Coolness" (line5: $r = 0.95$, line6: $r = 0.89$). The precuneus showed a high weight transition from "Interest" and "Awkwardness" to "Nervous scared." "Interest" showed a similar weight transition to "Emotional hurt" (line7: $r = 0.92$, line8: $r = 0.80$). Furthermore, "Sexual desire" and "Disgust" showed high weights in the latter positions and their transitions were highly correlated (line5: $r = 0.94$, line6: $r = 0.96$). To reinforce these results, we showed that the lines were present in anatomically similar regions in all subjects, as supported by the same or neighbor labels of the Destrieux atlas (Destrieux et al., 2010) (Fig. S3)

4. Discussion

To examine how emotional states are represented in the human brain, we recorded BOLD responses to a richer variety of emotion categories and constructed a semantic space consisting of dimensions of the response patterns. We performed CCA between the emotion ratings and the BOLD responses for each subject, and found that at least 18 emotion-category dimensions were significantly correlated with the individual BOLD response patterns elicited by the movie stimuli. This is a richer variety of emotions than the number of emotion categories or affective dimensions used in most previous brain-imaging studies

(i.e., a few to 15) (Chikazoe et al., 2014; Kragel and LaBar, 2014, 2016; Saarimaki et al., 2018; Saarimäki et al., 2016; Tettamanti et al., 2012; Wager et al., 2015; Winston, 2005). This type of a data-driven approach using a large number of emotion labels is becoming important to reveal emotion representation not only in the human behavior (Cowen et al., 2019, 2018; Cowen and Keltner, 2019, 2017) but also in the human brain. A recent neuroimaging study revealed the brain coding of emotion-related visual features that were extracted using a computational model trained by a large number of images with diverse emotion categories (Kragel et al., 2019). By constructing a semantic space, we visualized emotion-category distribution in the brain representation and demonstrated a smooth transition of dimension weights in the cortical surface. No such a transition has yet been revealed as the previous studies have used far fewer emotion categories. By using a rich variety of emotions, we were able to construct a continuous emotion space and show how the dimensions of the continuous space were mapped onto the cortical surface.

4.1. Semantic space of emotion

The semantic space found in this study covers a large variety of emotion representations, and emotional states posited in the previous studies can be explained in part by these representations. We found that some of the representative clusters of emotion categories in the semantic space constructed here were also reported in the previous studies, e.g., Disgust, Joy (Ekman et al., 1969; Plutchik, 1984), Interest (Plutchik, 1984), Anxiety related to Nervousness and Appreciation of beauty related to Awe (Lazarus and Lazarus, 1996). The remaining clusters (Confusion related to Curiousness, Awkwardness, Nostalgia related to Lethargy, Sexual desire, and Excitement related to Fever) were related to emotion dimensions reported by Cowen and Keltner (2017). Inconsistent with the results of previous studies, “fear” and “disgust” were combined into the same cluster in our study; these two emotions were posited as the different categories in the previous studies (Cowen and Keltner, 2017; Ekman et al., 1969; Plutchik, 1984). Furthermore, we found that positive- and negative-emotion categories were represented separately (see Materials and Methods, Fig. S7). This organization is not contradictory to the emotion distribution in the “Valence” dimension of the core affect model (Barrett, 2006; Barrett and Bliss-moreau, 2009; Russell, 2009; Russell and Barrett, 1999). In terms of the dimensions in the semantic space, we revealed that most dimensions consisted of semantically related emotion categories that can be interpreted as a group (e.g., “tension,” “nervousness,” and “alertness” form the “Nervous scared” dimension). Some dimensions for negative emotions such as “Disgust” and “Nervous scared (Fear)” had higher average weights than those of the other dimensions such as “Joy (Happiness)” and “Anger” (Fig. S8). The higher weights in these two negative emotions may be due to their more informative nature in the brain representation relative to the other categories, which is perhaps in relation to their importance for survival (Critchley et al., 2004; Grewe et al., 2009; Randler et al., 2017).

Emotion representations in the semantic space presented here were partly consistent with the representations revealed using behavioral data. Specifically, separate representations of positive- and negative-emotion categories (see Materials and Methods, Fig. S7) have previously been suggested in a the behavioral study (Toivonen et al., 2012). The distinction between positive and negative representations was also observed in a recent study using large-scale psychological assessments (Cowen and Keltner, 2017). Qualitatively, most of the emotion dimensions of our semantic space showed similar meanings to the emotion dimensions defined using behavioral responses to large amount of movie clips (e.g., “Awkwardness,” “Sexual desire,” and “Entrancement”). However, we found some differences regarding the boundaries among negative emotions (“Empathic pain,” “Disgust,” and “Fear”). In this study, these negative emotions were combined into a single dimension while they were maintained as separate dimensions in the work of Cowen and Keltner (2017). Subtle differences have been observed in the constitution of behavior- and the brain-activity-based semantic spaces and among different modalities of stimuli (e.g., vocal bursts and photographs) in behavioral studies (Cowen et al., 2019, 2018; Cowen and Keltner, 2019, 2017). A common finding between these studies and our study was that more than 20 emotion dimensions are required to explain behaviors or brain activities.

In terms of the emotion dimensions of the semantic space, we uncovered transitions of the weights in the following areas: the superior temporal area, the inferior parietal area, the precuneus, and the postcentral area. A previous study reported a unimodal to transmodal transition of functional connectivity in the former three areas, suggesting a cortical gradient of representation from sensory information to more abstract function (Margulies et al., 2016). A similar gradient was observed during speech recognition tasks, as represented by the gradient from visual/tactile (sensory) information to emotional/social (abstract) information (Huth et al., 2016). In these areas, we found the weight transition of following two groups of emotion: One group included emotions that played an important role in species preservation including protection from pathogens (Randler et al., 2017) and reproduction (Yang et al., 2018) (e.g., “Disgust,” “Nervous scared,” and “Sexual desire”). The second group included more elusive emotions that were associated with sociality (Harlé and Sanfey, 2007) and knowledge-seeking (Tong and Jia, 2017) (e.g., “Emotional hurt” and “Interest”). On the cortical surface, our weight transition from survival-related emotions to elusive emotions spatially corresponds to the sensory-to-abstract gradient found in previous studies (Huth et al., 2016; Margulies et al., 2016). This suggests that survival-related emotions may be explained by physical information (Woertman and Van Den Brink, 2012). Conversely, elusive emotions might be associated with more complex, higher-order cognitive information. Furthermore, in the postcentral area, we found a smooth transition of “Nervous scared” and “Disgust” (including empathic pain). The previous study also reported the localization of these negative emotions in this area (Benuzzi et al., 2008; Zhao et al., 2017). However, the cortical transition of the representation has not been revealed. This area is well known to have selectivity to somatosensory information, where regions responding to upper body parts are located more ventrally (Saadon-grosman et al., 2015). The localization of negative emotions may be caused by bodily reactions to the experience of high-arousal emotions, such as goose pimples (Grewe et al., 2009), and the weight transition may be caused by differences in the body parts exhibiting responses to these emotions.

In this study, we performed CCA between the emotion ratings and the BOLD response to estimate the number of emotion-category dimensions sufficient to explain the responses, and then performed encoding modeling using a ridge regression to construct the semantic space of emotion. Some readers might wonder whether we could have used regularized CCA to obtain equivalent results. We employed our current procedure for two reasons. First, we aimed at examining representations at both individual level (e.g., Figs. 3a and S10) and group level (e.g., Fig. 2a). If we had calculated CCA for individual brains, we would have obtained the latent representation between emotion ratings and the BOLD responses that differ across subjects, and there would be no straight-forward way to summarize the estimated individual representation for the group. To estimate the group representation, we could have re-run CCA for the concatenated voxels across subjects. However, the results of the group-level CCA were not compatible with those of the individual-level CCA. The encoding modeling was advantageous in this respect because we could run both individual- and group-level analyses as we did earlier (see, for instance, Huth et al., 2012; Nakai and Nishimoto, 2020). Second, we followed a recent behavioral study of comprehensive emotions by Cowen and Keltner (2017), in which the number of components of PCA was determined using CCA. The authors performed CCA between two types of reported emotional experience (emotion category labels and 600 emotion word labels). The number of significant components was used as the number of components in PCA, which was performed on

the reported emotional experiences to construct the semantic space of emotion. We sought to obtain results which would be compatible with those of the previous study by using a similar procedure.

An important issue in this context concerns the relationship between objective information and emotion representation in the brain. In this study, we subtracted the effect of low-level visual and auditory information when estimating emotion-category weights (see Materials and Methods). However, there still remains a possibility that higher-level objective information correlated with ratings of some emotion categories, and such objective information may be needed to represent emotional states. A recent neuroimaging study suggested that emotion schemas are embedded in the human visual system by showing that brain-activity patterns in the visual cortex could predict the output units of a CNN model tuned to infer emotion categories from input images (Kragel et al., 2019). Previous brain-imaging studies have reported that emotion representation is associated with brain regions that also represent semantic information, such as areas in the temporal cortex, (Kober et al., 2008; Lindquist et al., 2015; Saarimaki et al., 2018; Saarimäki et al., 2016) similar to our finding in this study. These results may be due to not only co-occurrence between specific emotions and specific semantic labels but also the function for recognition of emotion in the temporal lobe. This is supported by studies in which the recognition of facial expressions or emotion in music is reported to be impaired in subjects with atrophy of the temporal lobe (Hsieh et al., 2012; Lindquist et al. 2014). A review study (Pessoa, 2008) suggested that core regions of emotion representation (e.g., the amygdala) play a role in integrating emotion and cognitive information and emotion representation could not be separated from cognitive representation (Schaefer et al., 2006). Such integration is considered to be important in shaping human behavior and thoughts (Pessoa, 2008).

4.2. Limitations of our study

In the present study, we demonstrated the localization of emotion categories in the whole cortex, but not in the sub-cortex. Previous brain-imaging studies have conventionally focused on localization of basic emotions in the subcortical areas (Costafreda et al., 2008; Davis and Whalen, 2001; Holland and Gallagher, 2004; Phan et al., 2002; Schienle et al., 2005) and the connected cortical areas such as the insular, cingulate (Augustine, 1996; Damasio et al., 2000; Pezawas et al., 2005), and prefrontal areas(Dixon et al., 2017; Etkin et al., 2011; Viviani, 2014; Wager et al., 2008). Recent neuroimaging studies using multi-voxel pattern analysis support emotion representation in these areas (Kragel and LaBar, 2015; Saarimäki et al., 2016) and the cingulate and prefrontal areas (Kassam et al., 2013; Satpute and Lindquist, 2019). Consistent with previous studies, we also demonstrated emotion representations in these cortical areas (the five regions listed above in Table S1. See the caption and “Ratio of emotion-related voxels in eight regions” in Supplementary methods for details of the analysis). However, we could not provide strong support for the relationships between emotion and the sub-cortical areas. In particular, the amygdala is well established to be sensitive to “fear” (Davis and Whalen, 2001; Phan et al., 2002). When we examined voxels in the amygdala, we found voxels whose responses were explainable by the emotion ratings (Table S1). However, the number of these voxels was lesser than that for most cortical areas. Therefore, we here focused on the cortex.

The weak support for the association of emotion with the amygdala in this study might have been caused by the experimental settings. In most previous studies (Costafreda et al., 2008; Schienle et al., 2005, 2001; Tabert et al., 2001), brain activities were measured when evoking a specific emotion (especially “fear”) and also under neutral conditions; and the relationships of emotion with the amygdala were examined as the difference in activity between the two conditions. In comparison, we measured brain activities when feeling certain kinds of emotions evoked by movie scenes, and the responses to each emotion were estimated without comparison to the neutral condition. Furthermore, the amygdala activity might have been influenced by a task modality. A meta-analytical study quantified the probability of detecting amygdala activity in 385 functional neuroimaging studies and observed that probability of amygdala activation tends to be lower when using audiovisual stimuli than when using other stimulus modalities, such as face or olfactory stimuli (Costafreda et al., 2008). Overall, we could not provide strong evidence for a relationship between movie-evoked emotions and the amygdala.

To obtain results that were compatible with those of the behavioral study of Cowen and Keltner (2017), similar to procedures in their study, we performed CCA between the emotion ratings and the BOLD response. However, we observed some discrepancies between our results and those of Cowen and Keltner (2017). First, there was a qualitative difference in the emotion dimensions detected by CCA. While Cowen and Keltner (2017) estimated significant dimensions associated with the reliability of ratings between two split halves, we detected the dimensions of the ratings that contribute to the brain representation. Second, while Cowen and Keltner (2017) performed CCA between split halves of the ratings by multiple annotators, we performed CCA for each of the fMRI subjects; therefore, in contrast to their study, we obtained the number of significant dimensions for each individual and the across-individual variation in these numbers made it difficult to provide the specific number of dimensions. Finally, Cowen and Keltner (2017) found that each dimension had a high factor loading for each emotion category; in our results, however, each emotion dimension had high loadings for multiple emotion categories.

In the present study, we observed a relatively large across-subject variability in the number of distinct dimensions derived from brain representation (from 18 to 36). A potential interpretation of this variability is that it reflects the emotional granularity across subjects as discussed in previous studies (Satpute and Lindquist, 2019; Tugade et al., 2004). Although we currently do not have enough samples to examine this possibility directly, our modeling framework could provide a quantitative framework for future examination of such inter-subject differences.

The movie stimuli used in our experiments were rated for each of the 80 emotion categories by four independent annotators. To examine rating reliability, we quantified the across-annotator consistency of the ratings as Pearson’s correlation coefficients (see Emotion ratings and preprocessing). The mean correlation coefficient across the 80 emotion categories was not substantially high ($r = 0.41$); thus, our emotion representation results are based on emotion ratings with a moderate degree of across-annotator consistency. For some emotion categories with low consistency (listed in Table S2), there is possibility that the normative ratings across annotators reflect the averaged emotional states where individual differences were canceled out; therefore, such ratings may be sub-optimal for the capture of individual differences in the emotional experiences of each fMRI subject.

5. Conclusion

Taken together, our results demonstrate how a rich variety of emotion categories are represented in the human brain. Our findings provide a continuous space of emotion representation in the brain and show how the space is mapped across the cortex. In the cortical map, selectivity to emotion dimensions was smoothly transited and we observed region domination of specific emotion groups in some areas (e.g., disgust and fear selectivity in the postcentral area). Here, we examined emotion representation in the brain under the assumption that audio-visual movie stimuli elicit a large variety of emotion that we experience in our daily life. Future studies could provide more generalizable emotion representations by using other types of sensory or cognitive modalities (e.g., music or free conversation).

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

CRediT authorship contribution statement

Naoko Koide-Majima: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization, Data curation. **Tomoya Nakai:** Software, Writing - review & editing. **Shinji Nishimoto:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration.

Acknowledgments

This study was partly funded by Brother Industries Ltd. and JSPS KAKENHI Grant Numbers JP15H05311, JP15H05710, JP20K07718, JP20H05023 in #4903 (Evolinguistics), JST ERATO JPMJER1801, and CREST JPMJCR18A5. Data were collected with support from the National Institute of Information and Communications Technology.

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2020.117258.

References

Augustine, J.R., 1996. Circuitry and functional aspects of the insular lobe in primates including humans. Brain Res. Rev. 22, 229–244. doi:10.1016/S0165-0173(96)00011-2.

Barrett, L.F., 2006. Solving the emotion paradox: categorization and the experience of emotion. Personal. Soc. Psychol. Rev. 10, 20–46. doi:10.1207/s15327957pspr1001_2, https://doi.org/.

Barrett, L.F., Bliss-moreau, E., 2009. Affect as a psychological primitive. Adv. Exp. Soc. Psychol. 41, 167–218. doi:10.1016/S0065-2601(08)00404-8.

Benuzzi, F., Lui, F., Duzzi, D., Nichelli, P.F., Porro, C.A., 2008. Does it look painful or disgusting ? Ask your parietal and cingulate cortex does it look painful or disgusting ? Ask your parietal and cingulate cortex. J. Neurosci. 28, 923–931. doi:10.1523/JNEUROSCI.4012-07.2008.

Chi, T., Ru, P., Shamma, S.A., 2005. Multiresolution spectrotemporal analysis of complex sounds. J. Acoust. Soc. Am. 118, 887–906. doi:10.1121/1.1945807.

Chikazoe, J., Lee, D.H., Kriegeskorte, N., Anderson, A.K., 2014. Population coding of affect across stimuli, modalities and individuals. Nat. Neurosci. 17, 1114–1122. doi:10.1038/nn.3749.

Costafreda, S.G., Brammer, M.J., David, A.S., Fu, C.H.Y., 2008. Predictors of amygdala activation during the processing of emotional stimuli: a meta-analysis of 385 PET and fMRI studies. Brain Res. Rev. 58, 57–70. doi:10.1016/j.brainresrev.2007.10.012.

Cowen, A., Sauter, D., Tracy, J.L., Keltner, D., 2019. Mapping the passions: toward a high-dimensional taxonomy of emotional experience and expression. Psychol. Sci. Public Interes. 20, 69–90. doi:10.1177/1529100619850176.

Cowen, A.S., Elfenbein, H.A., Laukka, P., Keltner, D., 2018. Mapping 24 emotions conveyed by brief human vocalization. Am. Psychol. doi:10.1037/amp0000399.

Cowen, A.S., Keltner, D., 2019. What the face displays: mapping 28 emotions conveyed by naturalistic expression. Am. Psychol. doi:10.1037/amp0000488.

Cowen, A.S., Keltner, D., 2017. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. Proc. Natl. Acad. Sci. 114, E7900–E7909. doi:10.1073/pnas.1702247114.

Critchley, H.D., Wiens, S., Rotshtein, P., Öhman, A., Dolan, R.J., 2004. Neural systems supporting interoceptive awareness. Nat. Neurosci. 7, 189–195. doi:10.1038/nn1176.

Çukur, T., Nishimoto, S., Huth, A.G., Gallant, J.L., 2013. Attention during natural vision warps semantic representation across the human brain. Nat. Neurosci. 16, 763–770. doi:10.1038/nn.3381.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 9, 179–194. doi:10.1006/nimg.1998.0395.

Damasio, A.R., Grabowski, T.J., Bechara, A., Damasio, H., Ponto, L.L.B., Parvizi, J., Hichwa, R.D., 2000. Subcortical and cortical brain activity during the feeling of self-generated emotions. Nat. Neurosci. 3, 1049–1056. doi:10.1038/79871.

Davis, M., Whalen, P.J., 2001. The amygdala: vigilance and emotion. Mol. Psychiatry 6, 13–34. doi:10.1038/sj.mp.4000812.

Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53, 1–15. doi:10.1016/j.neuroimage.2010.06.010.

Dixon, M.L., Thiruchselvam, R., Todd, R., Christoff, K., 2017. Emotion and the prefrontal cortex: an integrative review. Psychol. Bull. 143, 1033–1081. doi:10.1037/bul0000096.

Ekman, P., 2016. What scientists who study emotion agree about. Perspect. Psychol. Sci. 11, 31–34. doi:10.1177/1745691615596992.

Ekman, P., Sorenson, E.R., Friesen, W.V., 1969. Pan-cultural elements in facial displays of emotion. Science 164, 86–88. doi:10.1126/science.164.3875.86.

Ellis, D.P.W., 2009 http://www.ee.columbia.edu/~dpwe/resources/matlab/.

Etkin, A., Tobias, E., Raffael, K., 2011. Emotional processing in anterior cingulate and medial prefrontal cortex. Trends Cogn. Sci. 15, 85–93. doi:10.1016/j.tics.2010.11.004.

Grewe, O., Reinhard, K., Eckart, A., 2009. The chill parameter: goose bumps and shivers as promising measures in emotion research. Music Percept. Interdiscip. J. 27, 61–74. doi:10.1525/mp.2009.27.1.61.

Harlé, K.M., Sanfey, A.G., 2007. Incidental sadness biases social economic decisions in the ultimatum game. Emotion 7, 876–881. doi:10.1037/1528-3542.7.4.876.

Holland, P.C., Gallagher, M., 2004. Amygdala-frontal interactions and reward expectancy. Curr. Opin. Neurobiol. 14, 148–155. doi:10.1016/j.conb.2004.03.007.

Hsieh, S., Hornberger, M., Piguet, O., Hodges, J.R., 2012. Brain correlates of musical and facial emotion recognition: evidence from the dementias. Neuropsychologia 50, 1814–1822. doi:10.1016/j.neuropsychologia.2012.04.006.

Huth, A.G., De Heer, W.A., Griffiths, T.L., Theunissen, F.E., Gallant, J.L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532, 453–458. doi:10.1038/nature17637.

Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A Continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76, 1210–1224. doi:10.1016/j.neuron.2012.10.014.

Johnson-Laird, P.N., Oatley, K., 1989. The language of emotions: an analysis of a semantic field. Cogn. Emot. 3, 81–123. doi:10.1080/02699938908408075.

Kassam, K.S., Markey, A.R., Cherkassky, V.L., Loewenstein, G., Just, M.A., 2013. Identifying emotions on the basis of neural activation. PLoS One 8. doi:10.1371/journal.pone.0066032.

Kober, H., Barrett, L.F., Joseph, J., Bliss-moreau, E., Lindquist, K., Wager, T.D., 2008. Functional grouping and cortical–subcortical interactions in emotion. Neuroimage 42, 998–1031. doi:10.1016/j.neuroimage.2008.03.059.

Kragel, P.a., LaBar, K.S., 2014. Advancing emotion theory with multivariate pattern classification. Emot. Rev. 6, 160–174. doi:10.1177/1754073913512519.

Kragel, P.A., LaBar, K.S., 2016. Decoding the nature of emotion in the brain. Trends Cogn. Sci. 20, 444–455. doi:10.1016/j.tics.2016.03.011.

Kragel, P.A., LaBar, K.S., 2015. Multivariate neural biomarkers of emotional states are categorically distinct. Soc. Cogn. Affect. Neurosci. 10, 1437–1448. doi:10.1093/scan/nsv032.

Kragel, P.A., Reddan, M.C., LaBar, K.S., Wager, T.D., 2019. Emotion schemas are embedded in the human visual system. Sci. Adv. 5, eaaw4358. doi:10.1126/sciadv.aaw4358.

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis – connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 1–28. doi:10.3389/neuro.06.004.2008.

Lazarus, R.S., Lazarus, B.N., 1996. Passion and Reason: Making Sense of Our Emotions. Oxford University Press.

Lench, H.C., Flores, S.A., Bench, S.W., 2011. Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: a meta-analysis of experimental emotion elicitations. Psychol. Bull. 137, 834–855. doi:10.1037/a0024244.

Lindquist, Kristen A., Gendron, Maria, Barrett, Lisa Feldman, Dickerson, B.C., 2014. Emotion perception, but not affect perception, is impaired with semantic memory loss. Emotion 14, 375. doi:10.1038/jid.2014.371.

Lindquist, K.A., Satpute, A.B., Gendron, M., 2015. Does language do more than communicate emotion? Curr. Dir. Psychol. Sci. 24, 99–108. doi:10.1177/0963721414553440.

Maaten, L.V.D., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., 2016. Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc. Natl. Acad. Sci. 113, 12574–12579. doi:10.1073/pnas.160828211.

Mauss, I.B., McCarter, L., Levenson, R.W., Wilhelm, F.H., Gross, J.J., 2005. The tie that binds? Coherence among emotion experience, behavior, and physiology. Emotion 5, 175–190. doi:10.1037/1528-3542.5.2.175.

Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Uğurbil, K., 2010. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain FMRI. Magn. Reson. Med. 63, 1144–1153. doi:10.1002/mrm.22361.

Nakai, T., Nishimoto, S., 2020. Quantitative models reveal the organization of diverse cognitive functions in the brain. Nat. Commun. 11, 1–12. doi:10.1038/s41467-020-14913-w.

Naselaris, T., Kay, K.N., Nishimoto, S., Gallant, J.L., 2011. Encoding and decoding fMRI. Neuroimage 56, 400–410. doi:10.1016/j.neuroimage.2010.07.073.

Nishida, S., Nishimoto, S., 2018. Decoding naturalistic experiences from human brain activity via distributed representations of words. Neuroimage 1–11. doi:10.1016/j.neuroimage.2017.08.017.

Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., Gallant, J.L., 2011. Reconstructing visual experiences from brain activity evoked by natural movies. Curr. Biol. 21, 1641–1646. doi:10.1016/j.cub.2011.08.031.

Pessoa, L., 2008. On the relationship between emotion and cognition. Nat. Rev. Neurosci. 9, 148–158. doi:10.1038/nrn2317.

Pezawas, L., Meyer-Lindenberg, A., Drabant, E.M., Verchinski, B.A., Munoz, K.E., Kolachana, B.S., Egan, M.F., Mattay, V.S., Hariri, A.R., Weinberger, D.R.,

2005. 5-HTTLPR polymorphism impacts human cingulate-amygdala interactions: a genetic susceptibility mechanism for depression. Nat. Neurosci. 8, 828–834. doi:10.1038/nn1463.

Phan, K.L., Wager, T., Taylor, S.F., Liberzon, I., 2002. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. Neuroimage 16, 331–348. doi:10.1006/nimg.2002.1087.

Plutchik, R., 1991. The Emotions. University Press of America.

Plutchik, R., 1984. Emotions: a general psychoevolutionary theory. Approaches to Emotion, pp. 197–219.

Randler, C., Desch, I.H., im Kampe, V.O., Wüst-Ackermann, P., Wilde, M., Prokop, P., 2017. Anxiety, disgust and negative emotions influence food intake in humans. Int. J. Gastron. Food Sci. 7, 11–15. doi:10.1016/j.ijgfs.2016.11.005.

Rottenberg, J., Ray, R.D., Gross, J.J., 2007. Emotion elicitation using films. Handbook of emotion elicitation and assessment. Oxford University Press.

Russell, J.A., 2009. Emotion, core affect, and psychological construction. Cogn. Emot. 23, 1259–1283. doi:10.1080/02699930902809375.

Russell, J.A., Barrett, L.F., 1999. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. J. Pers. Soc. Psychol. 76, 805–819. doi:10.1037//0022-3514.76.5.805.

Saadon-grosman, N., Tal, Z., Itshayek, E., Amedi, A., Arzy, S., 2015. Discontinuity of cortical gradients reflects sensory impairment. Proc. Natl. Acad. Sci. 112, 16024–16029. doi:10.1073/pnas.1506214112.

Saarimaki, H., Ejtehadian, L.F., Glerean, E., Jaaskelainen, I.P., Vuilleumier, P., Sams, M., Nummenmaa, L., 2018. Distributed affective space represents multiple emotion categories across the brain. Soc. Cogn. Affect. Neurosci. 13, 471–482. doi:10.1101/123521.

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I.P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., Nummenmaa, L., 2016. Discrete neural signatures of basic emotions. Cereb. Cortex 26, 2563–2573. doi:10.1093/cercor/bhv086.

Satpute, A.B., Lindquist, K.A., 2019. The default mode network's role in discrete emotion. Trends Cogn. Sci. 23, 851–864. doi:10.1016/j.tics.2019.07.003.

Schaefer, A., Braver, T.S., Reynolds, J.R., Burgess, G.C., Yarkoni, T., Gray, J.R., 2006. Individual differences in amygdala activity predict response speed during working memory. J. Neurosci. 26, 10120–10128. doi:10.1523/JNEUROSCI.2567-06.2006.

Schienle, A., Schäfer, A., Stark, R., Walter, B., Vaitl, D., 2005. Relationship between disgust sensitivity, trait anxiety and brain activity during disgust induction. Neuropsychobiology 51, 86–92. doi:10.1159/000084165.

Sheline, Y.I., Barch, D.M., Donnelly, J.M., Ollinger, J.M., Snyder, A.Z., Mintun, M.A., 2001. Increased amygdala response to masked emotional faces in depressed subjects resolves with antidepressant treatment: an fMRI study. Biol. Psychiatry 50, 651–658. doi:10.1016/S0006-3223(01)01263-X.

Strapparava, C., Valitutti, A., 2004. WordNet-affect: an affective extension of WordNet. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation, 4, pp. 1083–1086.

Tabert, M.H., Borod, J.C., Tang, C.Y., Lange, G., Wei, T.C., Johnson, R., Nusbaum, A.O., Buchsbaum, M.S., 2001. Differential amygdala activation during emotional decision and recognition memory tasks using unpleasant words: an fMRI study. Neuropsychologia 39, 556–573. doi:10.1016/S0028-3932(00)00157-3.

Tettamanti, M., Rognoni, E., Cafiero, R., Costa, T., Galati, D., Perani, D., 2012. Distinct pathways of neural coupling for different basic emotions. Neuroimage 59, 1804–1817. doi:10.1016/j.neuroimage.2011.08.018.

Toivonen, R., Kivelä, M., Saramäki, J., Viinikainen, M., Vanhatalo, M., Sams, M., 2012. Networks of emotion concepts. PLoS One 7, e28883. doi:10.1371/journal.pone.0028883.

Tong, E.M.W., Jia, L., 2017. Positive emotion, appraisal, and the role of appraisal overlap in positive emotion co-occurrence. Emotion 17, 40–54. doi:10.1037/emo0000203.

Tugade, M.M., Fredrickson, B.L., Feldman Barrett, L., 2004. Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. J. Pers. 72, 1161–1190.

Viviani, R., 2014. Neural correlates of emotion regulation in the ventral prefrontal cortex and the encoding of subjective value and economic utility. Front. Psychiatry 5, 123. doi:10.3389/fpsyt.2014.00123.

Wager, T.D., Davidson, M.L., Hughes, B.L., Lindquist, M.A., Ochsner, K.N., 2008. Prefrontal-subcortical pathways mediating successful emotion regulation. Neuron 59, 1037–1050. doi:10.1016/j.neuron.2008.09.006.

Wager, T.D., Kang, J., Johnson, T.D., Nichols, T.E., Satpute, A.B., Barrett, L.F., 2015. A Bayesian model of category-specific emotional brain responses. PLoS Comput. Biol. 11, e1004066. doi:10.1371/journal.pcbi.1004066.

Weidman, A.C., Steckler, C.M., Tracy, J.L., 2017. The jingle and jangle of emotion assessment: imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. Emotion 17, 267–295. doi:10.1037/emo0000226.

Winston, J.S., 2005. Integrated neural representations of odor intensity and affective valence in human amygdala. J. Neurosci. 25, 8903–8907. doi:10.1523/JNEUROSCI.1569-05.2005.

Woertman, L., Van Den Brink, F., 2012. Body image and female sexual functioning and behavior: a review. J. Sex Res. 49, 184–211. doi:10.1080/00224499.2012.658586.

Yang, L., Comninos, A.N., Dhillo, W.S., 2018. Intrinsic links among sex, emotion, and reproduction. Cell. Mol. Life Sci. 75, 2197–2210. doi:10.1007/s00018-018-2802-3.

Zhao, K., Zhao, J., Zhang, M., Cui, Q., Fu, X., 2017. Neural responses to rapid facial expressions of fear and surprise. Front. Psychol. 8, 761. doi:10.3389/fpsyg.2017.00761.