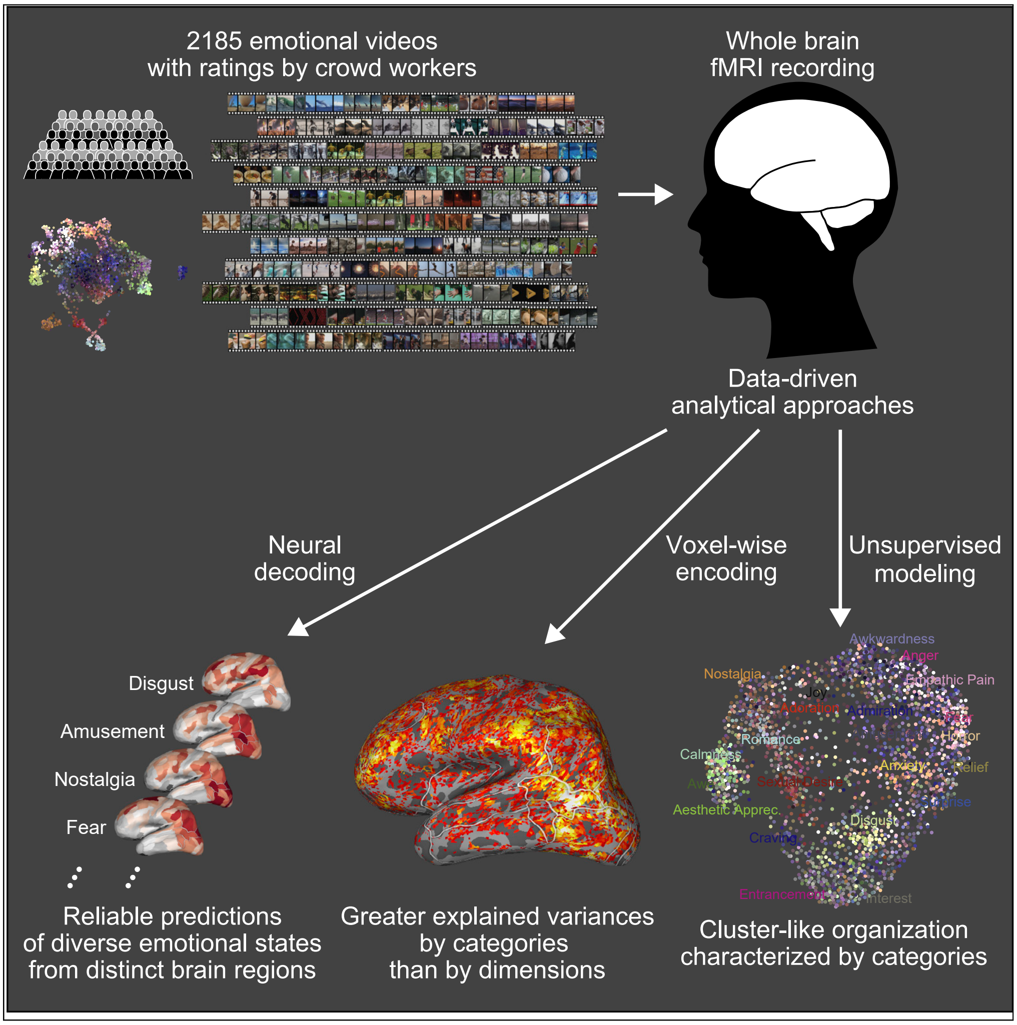
iScience

CellPress
OPEN ACCESS

Article

The Neural Representation of Visually Evoked Emotion Is High-Dimensional, Categorical, and Distributed across Transmodal Brain Regions



Tomoyasu Horikawa, Alan S. Cowen, Dacher Keltner, Yukiyasu Kamitani

horikawa-t@atr.jp (T.H.)
kamitani@i.kyoto-u.ac.jp (Y.K.)

HIGHLIGHTS

Dozens of video-evoked emotions were predicted from fMRI patterns in multiple regions

Categories better predicted cortical and subcortical responses than dimensions

Emotion-related responses had a cluster-like organization characterized by categories

Neural representation of emotion is high-dimensional, categorical, and distributed

Horikawa et al., iScience *23*, 101060
May 22, 2020 © 2020 The Author(s).
https://doi.org/10.1016/j.isci.2020.101060

iScience

CellPress
OPEN ACCESS

Article

The Neural Representation of Visually Evoked Emotion Is High-Dimensional, Categorical, and Distributed across Transmodal Brain Regions

Tomoyasu Horikawa,[1,*] Alan S. Cowen,[2] Dacher Keltner,[2] and Yukiyasu Kamitani[1,3,4,*]

SUMMARY

Central to our subjective lives is the experience of different emotions. Recent behavioral work mapping emotional responses to 2,185 videos found that people experience upward of 27 distinct emotions occupying a high-dimensional space, and that emotion categories, more so than affective dimensions (e.g., valence), organize self-reports of subjective experience. Here, we sought to identify the neural substrates of this high-dimensional space of emotional experience using fMRI responses to all 2,185 videos. Our analyses demonstrated that (1) dozens of video-evoked emotions were accurately predicted from fMRI patterns in multiple brain regions with different regional configurations for individual emotions; (2) emotion categories better predicted cortical and subcortical responses than affective dimensions, outperforming visual and semantic covariates in transmodal regions; and (3) emotion-related fMRI responses had a cluster-like organization efficiently characterized by distinct categories. These results support an emerging theory of the high-dimensional emotion space, illuminating its neural foundations distributed across transmodal regions.

INTRODUCTION

Emotions are mental states generated by the brain, constituting an evaluative aspect of our diverse subjective experiences. Traditionally, subjective emotional experiences have been described in theoretical and empirical studies using a limited set of emotion categories (e.g., happiness, anger, fear, sadness, disgust, and surprise) as proposed in early versions of basic emotion theory (Ekman and Friesen, 1969; Plutchik, 1980) and by broad affective dimensions (e.g., valence and arousal) that underpin the circumplex model of affect and the more recent core affect theory (Posner et al., 2005; Russell, 1980, 2003). Grounded in these theoretical claims, affective neuroscientists have investigated neural signatures underlying those representative emotion categories and affective dimensions (Barrett, 2017; Celeghin et al., 2017; Giordano et al., 2018; Hamann, 2012; Kragel and LaBar, 2015; Lindquist and Barrett, 2012; Saarimäki et al., 2016; Satpute and Lindquist, 2019; Wager et al., 2015), finding that distributed brain regions and functional networks are crucial for representing individual emotions. Importantly, though, studies within this tradition have limited their focus to a small set of emotions. This narrow focus does not capture the increasingly complex array of states associated with distinct experiences (Cowen et al., 2018, 2019a, 2020, Cowen and Keltner, 2017, 2018, 2019). As a result, much of the variability in the brain's representation of emotional experience likely has yet to be explained (Cowen et al., 2019a; Lindquist and Barrett, 2012).

To elucidate the structure of diverse emotional experiences, Cowen and Keltner, 2017, 2018, 2019, Cowen et al., 2018, 2019a, 2020) offered a new conceptual and methodological approach to capture the more complex ''high-dimensional emotion space'' that characterizes emotional experiences in response to a diverse array of stimuli in large-scale human behavioral experiments. In their study, Cowen and Keltner (2017) applied statistical methods to analyze reported emotional states elicited by 2,185 videos each annotated by multiple raters using self-report scales of 34 emotion categories (e.g., joy, amusement, and horror) and 14 affective dimensions (e.g., valence and arousal) derived from basic, constructivist, and dimensional theories of emotion. They identified 27 distinct kinds of emotional experience reliably associated with distinct videos. Furthermore, they found that the affective dimensions explained only a fraction of the variance in self-reports of subjective experience when compared with emotion categories, a finding that has now been replicated in studies of facial expression, speech prosody, nonverbal vocalization, and music (Cowen and

[1] Department of Neuroinformatics, ATR Computational Neuroscience Laboratories, Hikaridai, Seika, Soraku, Kyoto, 619-0288, Japan

[2] Department of Psychology, University of California, Berkeley, CA 94720-1500, USA

[3] Graduate School of Informatics, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan

[4] Lead Contact

*Correspondence: horikawa-t@atr.jp (T.H.), kamitani@i.kyoto-u.ac.jp (Y.K.)

https://doi.org/10.1016/j.isci.2020.101060

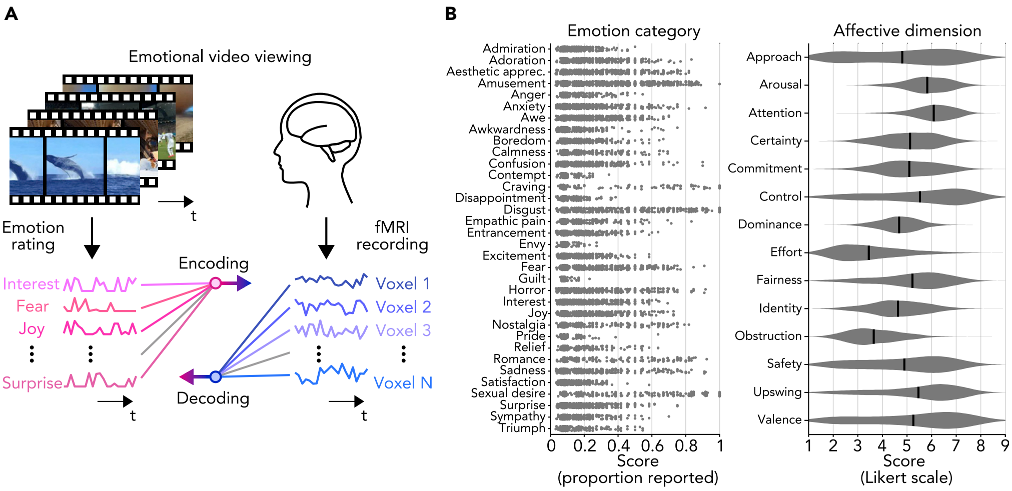


Figure 1. Decoding and Encoding Emotional Responses in fMRI Signals Evoked by Videos

(A) Schematic of the experiment and analyses. fMRI signals were recorded while subjects viewed emotional videos (with no sound). Decoding/encoding models were trained to predict scores/responses of individual emotions/voxels for presented videos from patterns of responses/scores, respectively.

(B) Score distributions of emotion categories and affective dimensions. Score distributions from a total of 2,181 unique videos are shown for 34 emotion categories and 14 affective dimensions (see Transparent Methods: "Video stimulus labeling"; Cowen and Keltner, 2017 for more details; abbreviation: Aesthetic apprec., Aesthetic appreciation). For emotion categories, each dot indicates a proportion that the category was evaluated by raters with non-zero scores for each video. For visualization purposes, only dots corresponding to videos with non-zero scores were shown with jittering. For affective dimensions, black bars indicate mean values across all videos. In the main analyses, mean scores averaged across multiple raters were used for each video.

Keltner, 2019; Cowen et al., 2018, 2019a, 2020). This work lays a foundation for examining the neural bases of emotional experience in terms of a rich variety of states, as well as whether—at the level of neural response—categories or dimensions organize the experience of emotion, and how a diverse array of states cluster in their neural representations.

The methodological criteria established in this study of emotion—the use of broad ranging stimuli sufficient to compare theoretical models of emotion—can be readily extended to the study of the neural basis of emotion. In addition, it is important for studies of the neural basis of emotion to consider activity across the whole brain, given that diverse psychological processes and distributed brain regions and functional networks are known to contribute to the representation of various emotional states (Barrett, 2017; Celeghin et al., 2017; Hamann, 2012; Lindquist and Barrett, 2012; Satpute and Lindquist, 2019). Although some recent studies have examined a richer variety of emotions than typically tested before (more than twenty), their investigations have been focused on specific brain regions (Kragel et al., 2019; Skerry and Saxe, 2015) and have relied on relatively small numbers of stimuli or conditions (tens or hundreds, but not thousands) (Koide-Majima et al., 2018; Saarimäki et al., 2018). Moreover, studies have not adequately differentiated representations of emotion from representations of sensory and semantic features that are known to be encoded throughout the cortex (Huth et al., 2012) and may contaminate neural responses to emotional stimuli (Kragel et al., 2019; Skerry and Saxe, 2015). Given these considerations, a comprehensive analysis of whole-brain responses to diverse emotional stimuli is still needed.

We sought to understand the neural underpinnings of the high-dimensional space of diverse emotional experiences associated with a wide range of emotional stimuli using the experimental resources generated in the study by Cowen and Keltner (2017). Specifically, we measured whole-brain functional magnetic resonance imaging (fMRI) signals while five subjects watched 2,181 emotionally evocative short videos. We then analyzed the measured fMRI signals using data-driven approaches, including neural decoding, voxel-wise encoding, and unsupervised modeling methodologies (Figure 1A) to investigate neural representations of emotional experiences characterized by 34 emotion categories and 14 affective dimensions with which the videos were previously annotated by a wide range of raters (Figure 1B).

Using multiple analytical approaches, we investigated different facets of the neural representation of the emotional experiences associated with the videos. Neural decoding was used to examine whether and where individual features (e.g., ratings of an emotion) are represented in the brain and provides a powerful way to identify mental states of subjects from brain activity patterns (Horikawa and Kamitani, 2017; Kragel et al., 2016; Sitaram et al., 2011). Voxel-wise encoding modeling made it possible to evaluate how specific sets of features (e.g., emotion category scores) modulate activity of individual voxels and to characterize representational properties of individual voxels by comparing encoding accuracies from different feature sets (de Heer et al., 2017; Güçlü and van Gerven, 2015; Kell et al., 2018; Lescroart and Gallant, 2018). Unsupervised modeling methods, including dimensionality reduction and clustering analyses, were used to characterize the distributional structure of brain representations of emotion in an exploratory manner (Kriegeskorte et al., 2008).

Here, we first show that dozens of distinct emotions associated with videos were reliably predicted from activity patterns in multiple brain regions, which enabled accurate identification of emotional states. Although the brain regions representing individual emotions overlapped, configurations of regions proved to be unique to each emotion and consistent across subjects. We show that voxel activity in most brain regions was better predicted by voxel-wise encoding models trained on ratings of emotion categories than on ratings of affective dimensions. Comparisons with other models trained on visual or semantic features revealed that the emotion category model uniquely explained response modulations of voxels in transmodal regions, indicating that these regions encode either emotional experience or visual affordances for specific emotional experiences (e.g., the inference that a landscape is likely to inspire awe). Finally, unsupervised modeling of emotion-related neural responses showed that distributions of brain activity patterns associated with diverse emotional experiences (or affordances) were organized with clusters corresponding to specific emotion categories, which were bridged by continuous gradients. These results provide support for the theory that emotion categories occupy a high-dimensional space (Cowen and Keltner, 2017) and shed new light on the neural underpinnings of emotion.

RESULTS

To understand neural representations of diverse emotional experiences associated with rich emotional content, we recorded whole-brain fMRI signals from five subjects while they were freely viewing emotionally evocative short videos (Figure 1A). The stimuli consisted of a total of 2,181 unique videos annotated with a total of 48 emotion ratings (Figure 1B; 34 emotion categories and 14 affective dimensions), each by multiple human raters, which were collected in a previous study (Cowen and Keltner, 2017; see Transparent Methods: ''Video stimulus labeling''). The fMRI volumes measured over the course of each video presentation were shifted by 4 s and averaged to estimate fMRI responses to individual video stimuli. We performed decoding and encoding analyses between the fMRI responses to each video and its emotion ratings by using regularized linear regression in a cross-validated manner (6-fold cross-validation; see Transparent Methods: ''Regularized linear regression analysis''). We also performed unsupervised modeling of the fMRI responses to examine the distribution of emotion-related brain activity patterns and whether their organization was better described by emotion categories or by broad affective dimensions (see Transparent Methods: ''Dimensionality reduction analysis'' and ''Clustering analysis'').

Neural Decoding Analysis to Predict Emotion Scores

We first used a neural decoding analysis to examine whether the emotions associated with presented videos could be predicted from brain activity patterns within specific regions and across the brain. We constructed a decoding model (decoder) for each emotion (34 emotion categories and 14 affective dimensions) to predict emotional experience ratings from the brain using multi-voxel activity patterns as input. The decoders of individual emotions were separately trained using activity patterns in each of multiple brain regions that include 360 cortical regions defined by a parcellation provided from the Human Connectome Project (Glasser et al., 2016; 180 cortical areas per hemisphere; HCP360) and 10 subcortical regions (e.g., amygdala and cerebellum). This procedure thus yielded a total of 370 region-wise decoders from brain activity per emotion. The performance for each emotion and brain region was evaluated by computing a Pearson correlation coefficient between predicted and true emotion scores across the 2,181 unique videos.

The decoding accuracies for the categories and dimensions obtained from multiple brain regions are shown in Figures 2A and 2B, respectively (see Figure 2C for color/shape schema of individual brain regions;

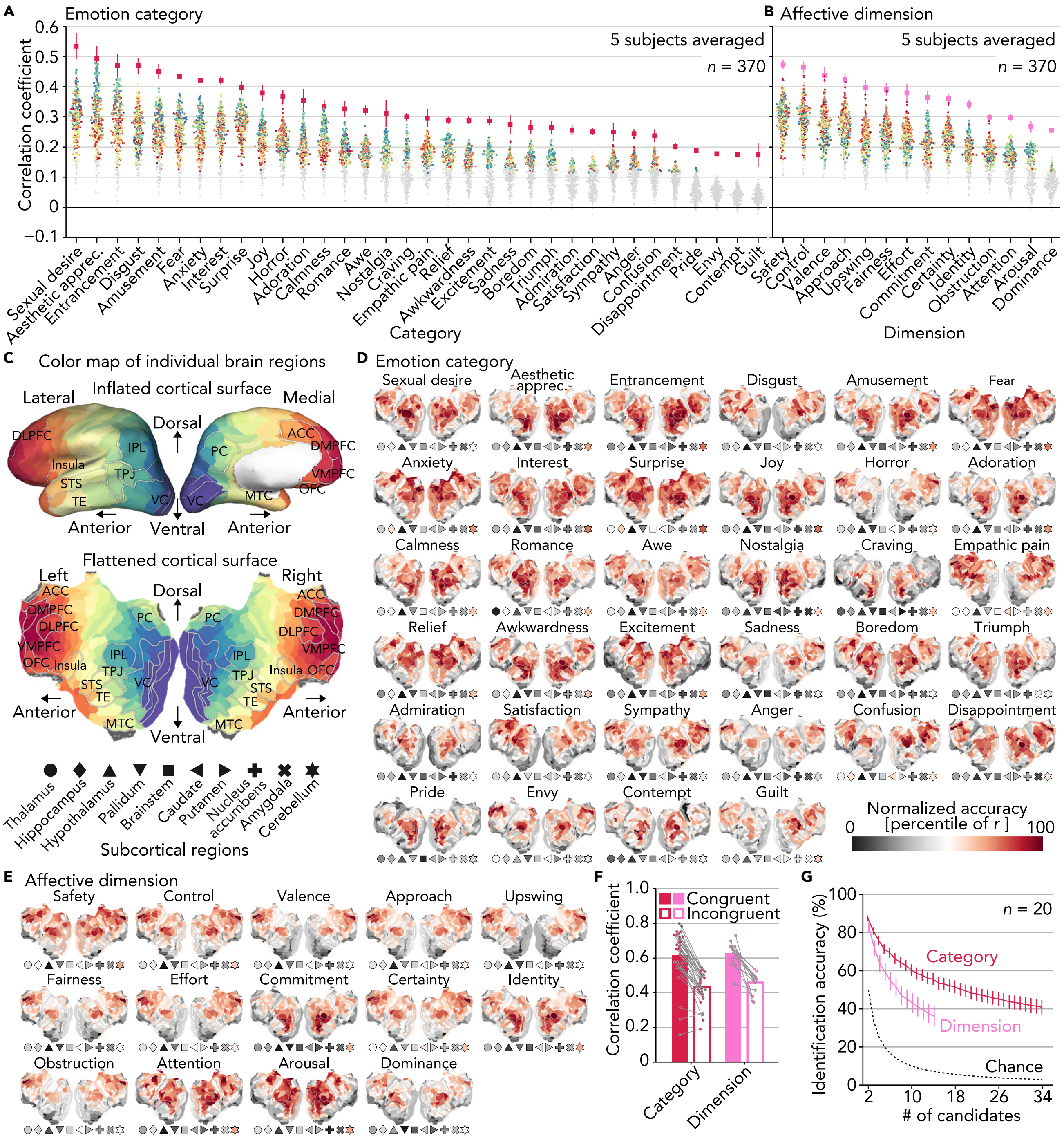


Figure 2. Neural Decoding Analysis to Predict Emotion Scores

(A) Decoding accuracy for individual categories. Dots indicate accuracies obtained from individual cortical regions of the HCP parcellation and subcortical regions ($n = 370$, five subjects averaged; see Figures S1A and S1B for results of representative cortical regions and subcortical regions). Red/pink squares indicate accuracies by ensemble decoders that aggregate predictions from multiple brain regions (error bars, 95% confidence intervals [C.I.] across subjects). Brain regions with significant accuracy from all five subjects are colored ($r > 0.095$, permutation test, $p < 0.01$, Bonferroni correction by the number of brain regions). Color and shape of each dot indicate locations of individual brain regions as (C).

(B) Decoding accuracy for individual dimensions. Conventions are the same as in (A).

(C) Color map of individual brain regions. Several brain regions are outlined and labeled (abbreviations: VC, visual cortex; TPJ, temporo-parietal junction; IPL, inferior parietal lobule; PC, precuneus; STS, superior temporal sulcus; TE, temporal area; MTC, medial temporal cortex; DLPFC/DMPFC/VMPFC, dorsolateral/dorsomedial/ventromedial prefrontal cortex; ACC, anterior cingulate cortex; and OFC, orbitofrontal cortex).

Figure 2. *Continued*

(D) Cortical surface maps of decoding accuracies for individual categories (five subjects averaged). Conventions of the marker points for the subcortical regions are the same as in (C).

(E) Cortical surface maps of decoding accuracies for individual dimensions (five subjects averaged). Conventions are the same as in (D).

(F) Correlations of region-wise decoding accuracy patterns between paired subjects. In congruent conditions, each dot indicates a mean correlation between the same emotion pairs averaged across all pairs of subjects. In incongruent conditions, each dot indicates a mean correlation between different emotion pairs averaged across all pairs of subjects and emotions.

(G) Emotion identification accuracy based on inter-individual similarities of region-wise decoding accuracy patterns (error bars, 95% C.I. across pairs of subjects [n = 20]).

also see Figures S1A and S1B for results of representative cortical regions and subcortical regions). Although scores of three emotion categories (envy, contempt, and guilt) were not reliably predicted from any brain regions consistently across five subjects, scores of the remaining emotions were predicted from brain activity patterns in at least one brain region with significant accuracy ($r > 0.095$, permutation test, $p < 0.01$, Bonferroni correction by the number of brain regions)—typically, from many different regions as the color of individual dots (or brain regions) indicated. The results established that ratings of the 27 emotion categories that were reported to be reliably elicited by video stimuli (Cowen and Keltner, 2017) were highly decodable from multiple brain regions. The poor accuracies for a few categories of emotion may be attributable in part to relatively lower proportions of videos evoking these categories (Figure 1B).

To integrate information represented in multiple brain regions, we constructed an ensemble decoder for each emotion by averaging predictions from multiple region-wise decoders selected based on decoding accuracy via a nested cross-validation procedure (red and pink squares in Figures 2A and 2B, respectively; see Transparent Methods: ‘‘Construction of ensemble decoders’’). Overall, the ensemble decoders tended to show higher accuracies than the best region-wise decoders (143/170 = 84.1% for categories, 52/70 = 74.3% for dimensions, pooled across emotions and subjects). These results indicate that information about individual emotions is represented in multiple brain regions in a complementary manner, supporting the notion of distributed, network-based representations of emotion (Barrett, 2017; Celeghin et al., 2017; Hamann, 2012; Lindquist and Barrett, 2012; Saarimäki et al., 2016; Satpute and Lindquist, 2019).

Region-Wise Decoding Accuracy Maps

To investigate which brain regions represented individual emotions, we visualized decoding accuracy maps by projecting accuracies in the 360 cortical regions onto flattened cortical surfaces alongside markers representing accuracies in the ten subcortical regions (Figures 2D and 2E; five subjects averaged). In general, brain regions showing high accuracy for specific emotions were not focal but rather distributed across multiple regions including visual, parietal, and frontal regions. For instance, scores of several basic emotions like sadness and anger were well predicted from distant regions around parietal (IPL) and frontal (DMPFC) regions (different hubs of default mode network). These results further support the notion that distributed brain regions and functional networks are recruited to represent individual emotions.

Together with the color differences of accuracy distributions across emotions in Figures 2A and 2B, visual inspections of the accuracy maps revealed different configurations of highly informative brain regions for different emotions. Interestingly, even subjectively related emotions (e.g., fear and horror, confusion and awkwardness) were often represented in markedly different configurations of brain regions, supporting a nuanced taxonomy of emotion.

Inter-Individual Consistency in Brain Regions Representing Distinct Emotions

To examine the extent to which the configurations of brain regions informative for decoding were unique to each emotion and consistent across subjects, we next tested whether emotions could be identified across subjects based on the region-wise decoding accuracy patterns of individual emotions estimated for individual subjects. A region-wise accuracy pattern for one emotion from one subject was used to identify the same emotion among a variable number of candidates (a pattern for the same emotion and randomly selected patterns for other emotions) from another subject using Pearson correlation coefficients (see Transparent Methods: ‘‘Emotion identification analysis’’). The analysis was performed for all pairs among five subjects (n = 20) separately using the accuracies for the categories and dimensions. For most of the categories and dimensions, correlations between the same emotion pairs (congruent; mean correlations, 0.615 for categories, 0.628 for dimensions) were larger than correlations between different emotion pairs (incongruent; mean correlations, 0.435 for categories, 0.458 for dimensions) (Figure 2E). Furthermore,

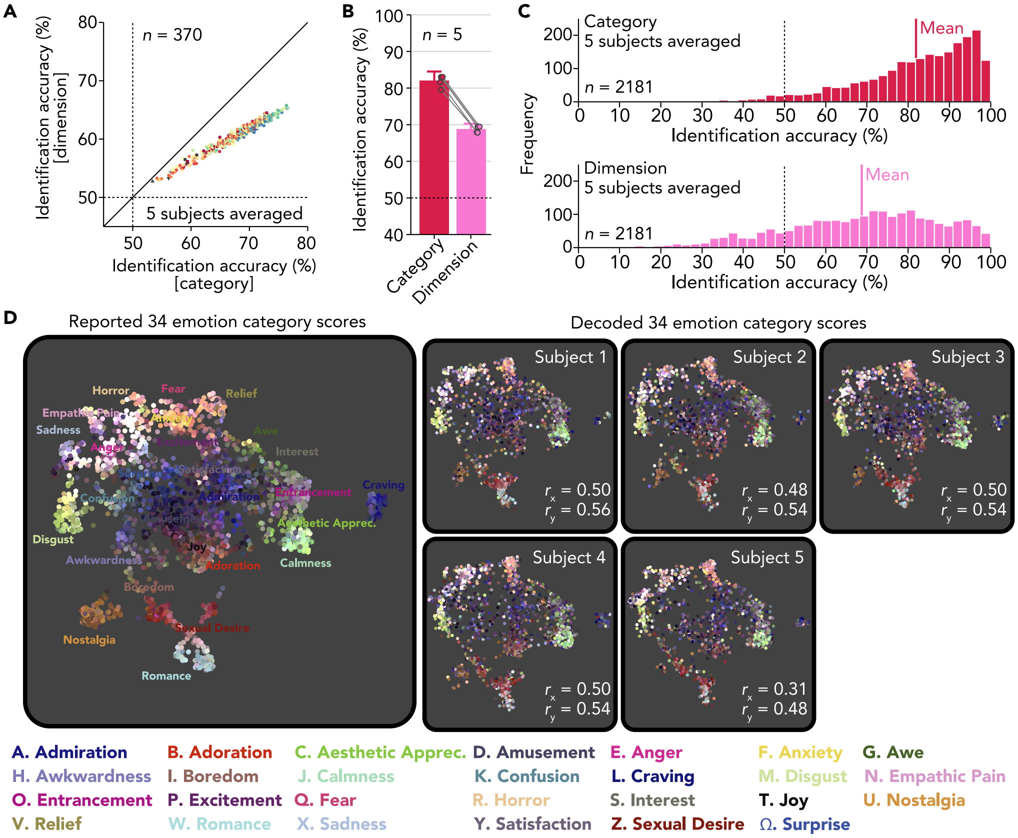


Figure 3. Identification of Videos via decoded Emotion Scores

(A) Mean video identification accuracies from region-wise decoders. Dots indicate accuracies from individual brain regions (five subjects averaged; dashed lines, chance level, 50%; see Figure S1C for individual subjects). Color and shape of each dot indicate locations of individual brain regions as Figure 2C.

(B) Mean video identification accuracies from ensemble decoders. Dots indicate accuracies of individual subjects (error bars, 95% C.I. across subjects; dashed line, chance level, 50%).

(C) Histogram of video identification accuracies for individual videos (ensemble decoder; five subjects averaged; dashed line, chance level, 50%).

(D) Two-dimensional maps of emotional experiences constructed from reported and decoded scores. Each dot corresponds to each video and is colored according to its original reported scores using a weighted interpolation of unique colors assigned to 27 distinct categories (see Cowen and Keltner, 2017 for the color schema).

mean identification accuracy averaged across all pairs of subjects far exceeded chance levels across all candidate set sizes for both of the categories and dimensions (Figure 2F; e.g., five-way identification accuracy, 71.4% for categories, 59.3% for dimensions; chance level, 20%). These results suggest that distinct emotions associated with the video stimuli were differentially represented across the brain in a consistent manner across individuals.

Identification of Videos via Decoded Emotion Scores

Given that a complex emotional state can be represented by a blend or combination of multiple emotions (e.g., anger with sadness), we next tested a decoding procedure that took into account multiple emotions simultaneously. To do so, we examined whether a specific video stimulus could be predicted, or identified, from the concatenated predictions of emotion categories or those of affective dimensions. The identification analysis was performed in a pairwise manner, in which a vector of predicted scores was compared with two candidate vectors (one from a true video and the other from a false video randomly selected from the other 2,180 videos) to test whether a correlation for the true video was higher than that for the false video (see Transparent Methods "Video identification analysis"). For each video, the analysis was performed with all combinations of false candidates (*n* = 2,180), and percentages of correct answers were used as an identification accuracy for one sample.

Identification accuracies obtained from predictions by the region-wise decoders are shown in Figure 3A (five subjects averaged, see Figure S1C for individual subjects). The results showed better identification accuracy via the categories than via the dimensions with most region-wise decoders. The superior accuracy of the categories over the dimensions was also observed with the ensemble decoders, which showed greater than 10% differences in mean accuracy in all subjects (Figure 3B; 81.9% via categories, 68.7% via dimensions, five subjects averaged; see Figure 3C for accuracies of individual videos). This tendency was preserved even when the number of emotions used for identification was matched between the category and dimension models by randomly selecting a subset of 14 categories from the original 34 categories (Figure S1D; 75.4%, five subjects averaged). These results indicate that emotion categories capture far more variance, and presumably details, in the neural representation of emotionally evocative videos than broad affective features such as valence and arousal.

Visualization of Decoded Emotional Experiences (or Affordances)

To explore the extent to which richly varying emotional states were accurately decoded, we used a nonlinear dimensionality reduction method to visualize the distribution of emotional experiences evoked by the 2,181 videos, originally represented in terms of 34 emotion categories, in a two-dimensional space (Cowen and Keltner, 2017), and tested whether the map could be reconstructed from the decoded scores. For this purpose, we used a novel dimensionality reduction algorithm, called Uniform Manifold Approximation and Projection (UMAP; McInnes et al., 2018), which enables the projection of new data points after constructing a mapping function from an independent training dataset. We first constructed a mapping function that projects 34-dimensional emotion category scores into a two-dimensional space using the original emotion category scores (Figure 3D left), and then the function was used to project decoded scores from ensemble decoders of individual subjects onto the map (Figure 3D right).

We first confirmed that the UMAP algorithm applied to the 34 emotion category scores replicated a similar two-dimensional map produced in the previous study (Cowen and Keltner, 2017), in which distinct categories were found to be organized along gradients between clusters associated with each category. Upon inspection of the maps reconstructed from decoded scores, it is apparent that distributions of colors (scores) of the original map were precisely replicated for all subjects. This was confirmed quantitatively by calculating correlations of horizontal (x) and vertical (y) data positions in the two-dimensional space between true and reconstructed maps (r = 0.46 for horizontal axis, r = 0.53 for vertical axis, five subjects averaged). The results demonstrate how precisely different emotional states were decoded from brain activity and provide a useful means to visualize decoded emotional states.

Construction of Voxel-Wise Encoding Models

We next used voxel-wise encoding models to characterize how varying features associated with the presented videos modulate activities in each voxel. We constructed an encoding model for each voxel to predict activities induced by the emotionally evocative videos using a set of features (e.g., scores for 34 emotion categories), and model performance was evaluated by computing a Pearson correlation coefficient between measured and predicted activities (see Transparent Methods: "Regularized linear regression analysis").

Encoding Models Predicting Brain Activity from Emotional Scores

We first tested how well encoding models constructed from the 34 emotion categories (e.g., joy, anger) and the 14 affective dimensions (e.g., valence, arousal) predicted brain activities induced by the presented videos (Figure 4A; Subject 1; see Figure S2A for the other subjects). We found that both the category and dimension models accurately predicted activity in many regions, indicating that a broad array of brain regions are involved in encoding information relevant to emotion, as represented by categories and affective dimensions.

In comparing the model performances, we found that on average the category model outperformed the dimension model by a considerable margin (Figure 4B; Subject 1; slope angle, 34.6°; two-tailed t test, p < 0.01 by jackknife method; see Figure S2B for the other subjects). In comparison with the dimension model, the category model predicted voxel responses significantly better in all subjects (Figure 4C, paired t test across all voxels; p < 0.01 for all subjects) and also predicted 73.0% greater number of voxels with significant accuracies (r > 0.111, permutation test, p < 0.01, Bonferroni correction by the number of voxels; five subjects averaged). In addition, when we constructed a joint encoding model by concatenating scores

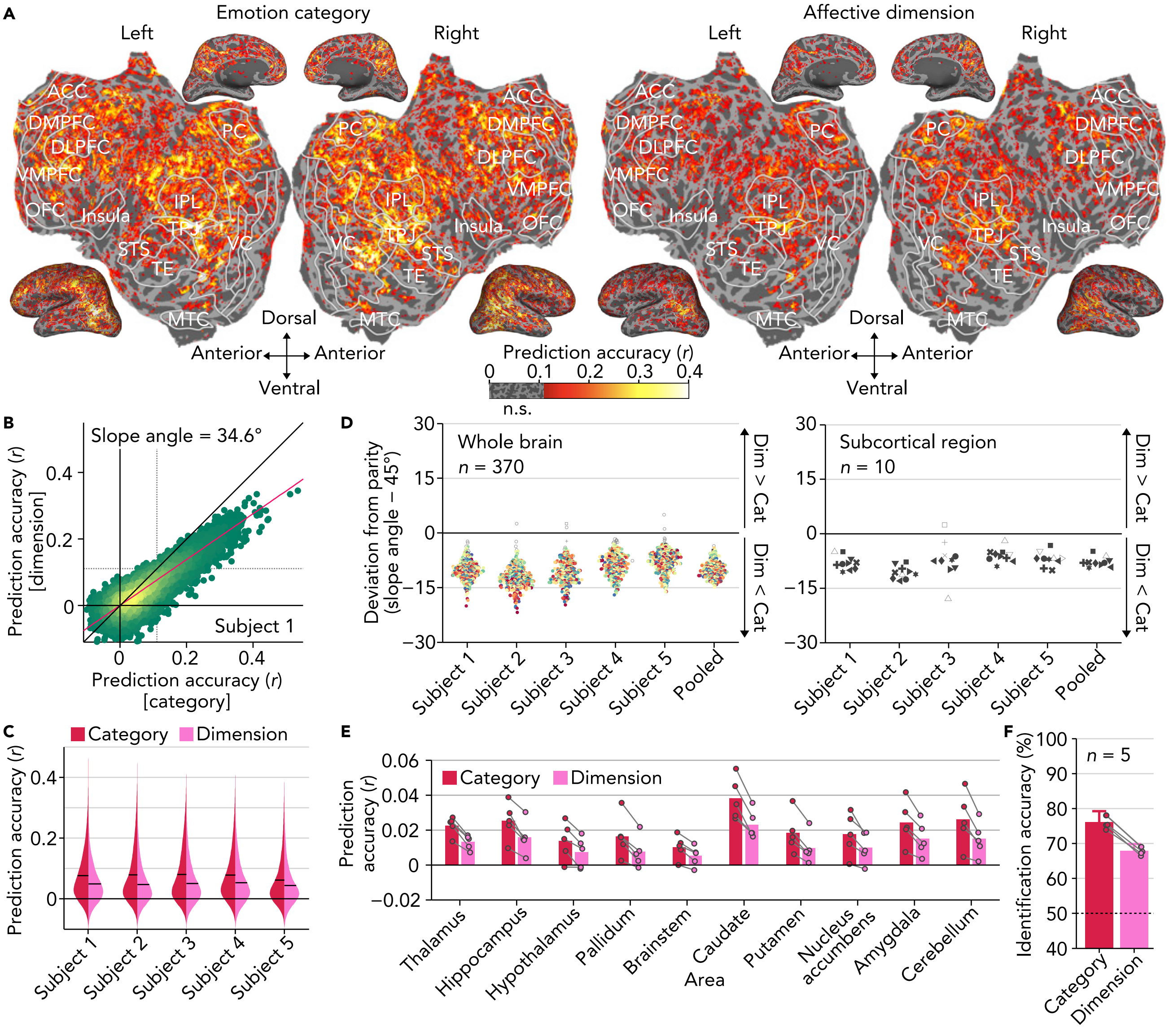


Figure 4. Encoding Models Predicting Brain Activity from Emotional Scores

(A) Prediction accuracies of emotion encoding models (Subject 1, see Figure S2A for the other subjects).

(B) Prediction accuracies of individual voxels (red line, best linear fit; dotted lines, $r = 0.111$, permutation test, $p < 0.01$, Bonferroni correction by the number of voxels; see Figure S2B for the other subjects).

(C) Distributions of prediction accuracies of all voxels (black bars, mean of all voxels).

(D) Comparisons of prediction accuracies for individual brain regions. Each dot indicates a deviation of the estimated slope angle from the parity for each brain region (baseline, 45°). Brain regions with significant deviation are colored/filled (two-tailed *t* test, $p < 0.01$ by jackknife method, Bonferroni correction by the number of ROIs, $n = 370$). Color and shape of each dot indicate locations of individual brain regions as Figure 2C.

(E) Mean prediction accuracies in individual subcortical regions. Dots indicate accuracies of individual subjects.

(F) Mean video identification accuracy via predicted brain activities. Dots indicate accuracies of individual subjects (error bars, 95% C.I. across subjects; dashed line, chance level, 50%).

of 34 categories and 14 dimensions, the increase in the number of significantly predicted voxels was marginal when compared with the category model (1.8%, five subjects averaged) but substantial when compared with the dimension model (75.9%, five subjects averaged). Together, these findings suggest that the brain representation of affective dimensions is subsumed by, and can be inferred from, the brain representation of emotion categories.

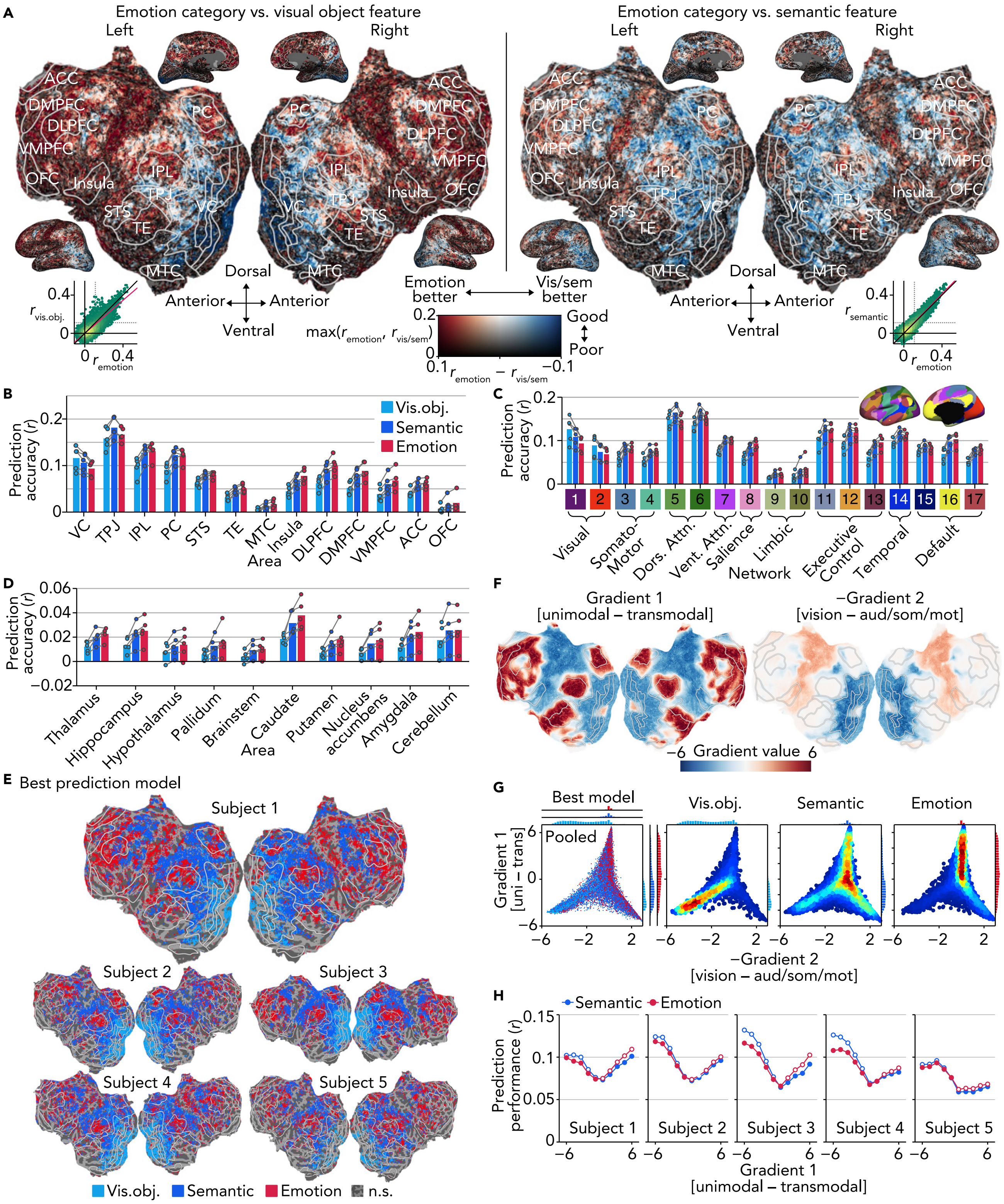


Figure 5. Disentangling Emotional, Visual Object, and Semantic Feature Encoding

(A) Differences in prediction accuracies of emotion, visual object, and semantic models (Subject 1, see Figure S4A for the other subjects). Inserted scatterplots follow the same conventions as Figure 4B.

(B) Mean prediction accuracies in individual brain regions. Dots indicate accuracies of individual subjects.

Figure 5. *Continued*

(C) Mean prediction accuracies in individual global networks. Inset offers the definitions of individual networks (see Figure S4C for a larger view). Conventions are the same as in (B).

(D) Mean prediction accuracies in individual subcortical regions. Conventions are the same as in (B).

(E) Best models among emotion (category), visual object, and semantic models (see Figure S4D for results with the dimension model).

(F) Gradient values of the first and second PG axes. The sign of second PG axis is negative, following the convention of Margulies et al. (2016).

(G) Joint and marginal distributions of the best models in PG space (five subjects pooled, see Figure S4E for individual subjects). Voxels with significant accuracy from either the visual object, semantic, or emotion models are shown (permutation test, $p < 0.01$, Bonferroni correction by the number of voxels).

(H) Mean prediction accuracies along the first PG axis. Voxels were assigned to ten bins, or levels of the first axis, such that each bin consists of a roughly equal number of voxels. White circles indicate significant performance over the opponent model (two-tailed paired *t* test, $p < 0.01$, Bonferroni correction by the numbers of PG levels and subjects).

We next explored whether there were particular brain regions in which the dimension model outperformed the category model. We divided the whole cortex into 360 regions using the HCP360 parcellation (Glasser et al., 2016) and then separately compared the performances in each of the 360 cortical and 10 subcortical regions using a slope angle of the best linear fit of the accuracies from the two models (see Transparent Methods: "Slope estimates for performance comparisons"). The distributions of the slope angle suggest that activity in nearly every region was better predicted by the category model (Figure 4D left; 368/370 for category, 0/370 for dimension; two-tailed *t* test, $p < 0.01$ by jackknife method, Bonferroni correction by the number of ROIs, $n = 370$; five subjects pooled). Notably, this tendency was even observed in ten subcortical regions, including the amygdala (see Transparent Methods: "Regions of interest [ROI]"), which are often thought to prioritize simpler appraisals of valence and arousal (Figure 4D right; 9/10 for category, 0/10 for dimension; two-tailed *t* test, $p < 0.01$ by jackknife method, Bonferroni correction by the number of ROIs, $n = 370$; five subjects pooled; see Figure 4E for mean encoding accuracy for the subcortical regions).

To further compare the performance of the category and dimension models, a video identification analysis was performed via predicted brain activities obtained from each of the two models (cf., Figure 3B; see Transparent Methods: "Video identification analysis"). We confirmed that identification accuracies from the category model were substantially better than the dimension model for all subjects (Figure 4F; 76.1% for category, 67.9% for dimension, five subjects averaged).

In sum, these encoding analyses demonstrate that emotion categories substantially outperform affective dimensions in capturing voxel activity induced by emotional videos (see Figure S3 for control analyses of potential confounds). Taken together, these results build on previous behavioral findings (Cowen and Keltner, 2017), which showed that self-reported emotional experience was more precisely characterized by a rich array of emotion categories than by a range of broad affective dimensions proposed in constructivist and componential theories to be the underlying components of emotional experience (Barrett, 2017; Posner et al., 2005; Smith and Ellsworth, 1985). In particular, using a wide range of analytic approaches, we repeatedly found that emotion categories capture nuanced, high-dimensional neural representations that cannot be accounted for by a lower-dimensional set of broad affective features.

Disentangling Emotional, Visual Object, and Semantic Feature Encoding

Given that our video stimuli consisted of scenes and events that also included many visual objects and semantic features, we next sought to disentangle brain representations of emotion from those correlated features. For this analysis, we constructed two additional encoding models from outputs of a pre-trained deep neural network for object recognition ("visual object model"; Simonyan and Zisserman, 2014) and from semantic features (or concepts; e.g., cats, indoor, and sports) annotated by human raters via crowd-sourcing ("semantic model"; see Transparent Methods: "Video stimulus labeling" for these features). We then evaluated model performances for each model and compared them with those of the "emotion model" constructed from the emotion category scores.

As found in previous studies that used similar visual (Güçlü and van Gerven, 2015) and semantic (Huth et al., 2012) encoding models, our visual object model and semantic model were also effective in predicting voxel activity elicited by video stimuli even outside of the visual cortex (Figure 5A; Subject 1; see Figure S4A for the other subjects). Although the accuracies from these two models positively correlated with those from the emotion model (inset in Figure 5A), the emotion model consistently outperformed the other two models in certain brain regions (Figure 5B), including IPL, TE, and several frontal regions (DLPFC, DMPFC, VMPFC, and OFC), which are known to represent multimodal emotion information (Chikazoe et al., 2014;

Skerry and Saxe, 2014) and abstract semantic features (Huth et al., 2016; Skerry and Saxe, 2015), and in multiple functional networks (Figure 5C; Yeo et al., 2011), including the salience network (8), subparts of the limbic network (10), executive control network (13), and default mode network (16 and 17), which is also consistent with previous reports (Barrett, 2017; Barrett and Satpute, 2013; Lindquist and Barrett, 2012; Satpute and Lindquist, 2019). Furthermore, although accuracies were not as high, activity in the subcortical regions also tended to be better predicted by the emotion models than by the other two models (Figure 5D). This analysis demonstrates that brain regions preferentially encoding emotion are broadly distributed in a similar manner across subjects (Figure 5E; see Figure S4D for evaluation based on slope estimates, cf., Figure 4C).

Visual Object, Semantic, and Emotion Encoding Models along the Cortical Hierarchy

To better understand which brain regions preferentially represent emotion, we drew on the concept of the ''principal gradient (PG),'' which is estimated from functional connectivity patterns (Margulies et al., 2016) and describes global hierarchical gradients in cortical organization (Huntenburg et al., 2018). Margulies et al. (2016) used a dimensionality reduction method to explore principal components of variances in cortical connectivity patterns and found that the first component differentiates between unimodal and transmodal brain regions (Figure 5F left) and the second component differentiates between brain regions that represent different modalities (from visual to auditory and somatomotor cortices, Figure 5F right). The meta-analysis by Margulies et al. (2016) and a review study by Huntenburg et al., 2018 suggest that different regions along the PGs are responsible for broadly different brain functions.

Here, we tested empirically whether the PGs could account for the spatial arrangement of voxels preferentially representing visual object, semantic, and emotion information using encoding performances of models constructed from those different features. We plotted voxels predicted with significant accuracy by the visual object, semantic, or emotion models within the two-dimensional space defined by their values along the first and second PGs (Figure 5G, five subjects pooled, see Figure S4E for individual subjects). Voxels best predicted by each model occupied different locations in the PG space. Most importantly, along the first PG axis, voxels best predicted by the visual object, semantic, and emotion models were each densely distributed in regions with low, medium, and high values, respectively (ANOVA, interaction between frequency and level of the first PG axis, $p < 0.01$). In terms of raw prediction accuracy, voxel activity was better predicted by both the emotion and semantic models at low and high values of the first PG than at the midpoint. However, at low/high values of the first PG, the semantic/emotion model outperformed the emotion/semantic model, respectively (Figure 5H; ANOVA, interaction between accuracy and level of the first PG axis, $p < 0.01$ for all subjects). Because emotional responses are correlated with semantic features high in emotional affordance (e.g., injury, sexual activity), any brain regions predicted better by the emotion model than by the semantic model, even marginally, should be considered to encode emotional experience (or affordances for specific emotional experience). Hence, the present findings suggest that brain regions high along the first gradient axis—that is, transmodal regions—are central to the brain's representation of emotion.

It is important to note that the results presented thus far were replicated using alternative analytic approaches, including the decoding (Figures S5A and S5B) and representational similarity analyses (Figures S5C and S5D; Kriegeskorte et al., 2008). These analyses showed higher identification decoding accuracies (cf., Figure 3A) and representational similarities across the brain for the category than the dimension model (Figures S5A and S5C). In both cases, a highly similar set of brain regions was better fit by emotion models than by visual object model or semantic model (Figures S5B and S5D). These analyses demonstrate that our results establishing the primacy of emotion categories in neural representation are highly robust to alternative methodological choices.

Taken together, these results characterize how regions representing emotion are spatially arranged within the brain. Although emotion-related representations are broadly distributed across the cortex, they are centered in regions that lie on the far end of a hierarchical gradient ranging from unimodal to transmodal regions. The representational shift along this gradient from visual to semantic to emotion-related representations may speak to the high levels of information integration and feature abstraction necessary for an emotional response.

Distribution of Brain Activity Patterns Induced by Emotional Videos

Although the above analyses addressed how particular dimensions, or kinds, of emotion are represented in the brain, the results thus far are not sufficient to reveal the distribution of neural responses to emotional stimuli along these dimensions, that is, how different kinds of emotion are distributed within a high dimensional space (Cowen and Keltner, 2017). To address the distribution of emotion-related brain activity, we performed unsupervised modeling using nonlinear dimensionality reduction and clustering analyses.

We first visually inspected the distribution of brain activity patterns induced by the 2,181 emotional videos using the UMAP algorithm that we previously used to visualize the category emotion scores (cf., Figure 3D). In doing so, we selected voxels best predicted by either of the category or dimension encoding models with significant accuracy (cf., Figure S4D, including voxels in both cortical and subcortical regions), to focus on activity primarily representing emotion-related information. We then used activity patterns of those selected voxels as inputs to the UMAP algorithm, thus projecting a total of 2,181 brain activity patterns into a two-dimensional space (see Transparent Methods: ''Dimensionality reduction analysis''). The resulting maps colored based on ratings of the videos in terms of 27 emotion categories reveal a number of brain activity clusters that correspond to experiences of specific emotions (Figure 6A; five subjects averaged; e.g., sexual desire, disgust, and aesthetic appreciation), which were also observed even in maps of individual subjects. By contrast, maps colored by three representative affective dimensions (separately for positive and negative parts after 5 [neutral] was subtracted) do not highlight many distinct clusters (Figure 6B), indicating that the distribution of brain activity patterns was better captured by emotion categories.

We next quantified the degree to which brain activity patterns were distributed into cluster-like formations associated with specific emotions. For each subject, we applied k-means clustering to the 2,181 brain activity patterns of the voxels selected as above, without any reference to the emotion scores of the eliciting videos. We set the number of clusters to 27, given that the previous study revealed 27 distinct varieties of emotional experiences elicited using the same videos (Cowen and Keltner, 2017), and we would expect at most one cluster per distinct emotion. Note, however, that similar results were obtained using different numbers of clusters (Figure S6). To characterize the emotional properties of each cluster, we examined how brain activity patterns corresponding to videos with top 5% high scores (109 videos) for each emotion are assigned to these clusters using scores of the previously reported 27 distinct categories and the positive and negative parts of affective dimensions (absolute values after 5 [neutral] was subtracted).

Histograms of the samples assigned to the 27 clusters for each emotion are shown in Figure 6C (see Figure S6A for the other subjects), revealing an important disparity between the categories and dimensions. Namely, top samples of several categories, including entrancement, sexual desire, and disgust, tended to belong to a small number of clusters, whereas those of affective dimensions tended to be broadly assigned to many clusters. This difference was quantified in two ways. First, the histograms of each emotion—independently sorted according to their frequencies within each cluster—were steeper for the categories than dimensions (Figure 6D; five subjects averaged; ANOVA, interaction between frequency and sorted clusters, $p < 0.01$; see Figure S6B for individual subjects). Second, entropy calculated for each histogram of individual emotions tended to be low for many categories (e.g., O: entrancement, Z: sexual desire, and C: aesthetic appreciation) compared with most of the dimensions (Figure 6E, five subjects averaged, see Figure S6C for individual subjects).

Together, these results indicate that brain activity patterns induced by emotional videos have cluster-like distributions that align to a greater degree with categories than with dimensions. Meanwhile, close inspection of the map (Figure 6A) also reveals that the patterns of brain activity corresponding to certain emotions (e.g., fear and horror) seem to be not entirely discrete but overlapping in their distribution. Indeed, top samples of several categories (e.g., W: romance and Z: sexual desire; C: aesthetic appreciation and J: calmness) tended to fall into overlapping clusters (Figures 6C and S6A). This finding may shed light on the biological basis of gradients between distinct emotional experiences, originally revealed in self-report (Cowen and Keltner, 2017), by suggesting that they correspond to gradients in evoked patterns of brain activity.

DISCUSSION

We have sought to characterize the neural substrates of emotional experience by analyzing activity induced by the richest set of emotional visual stimuli investigated to date in neuroscience—2,181 emotionally evocative videos annotated with 34 emotion categories and 14 affective dimensions—using neural decoding,

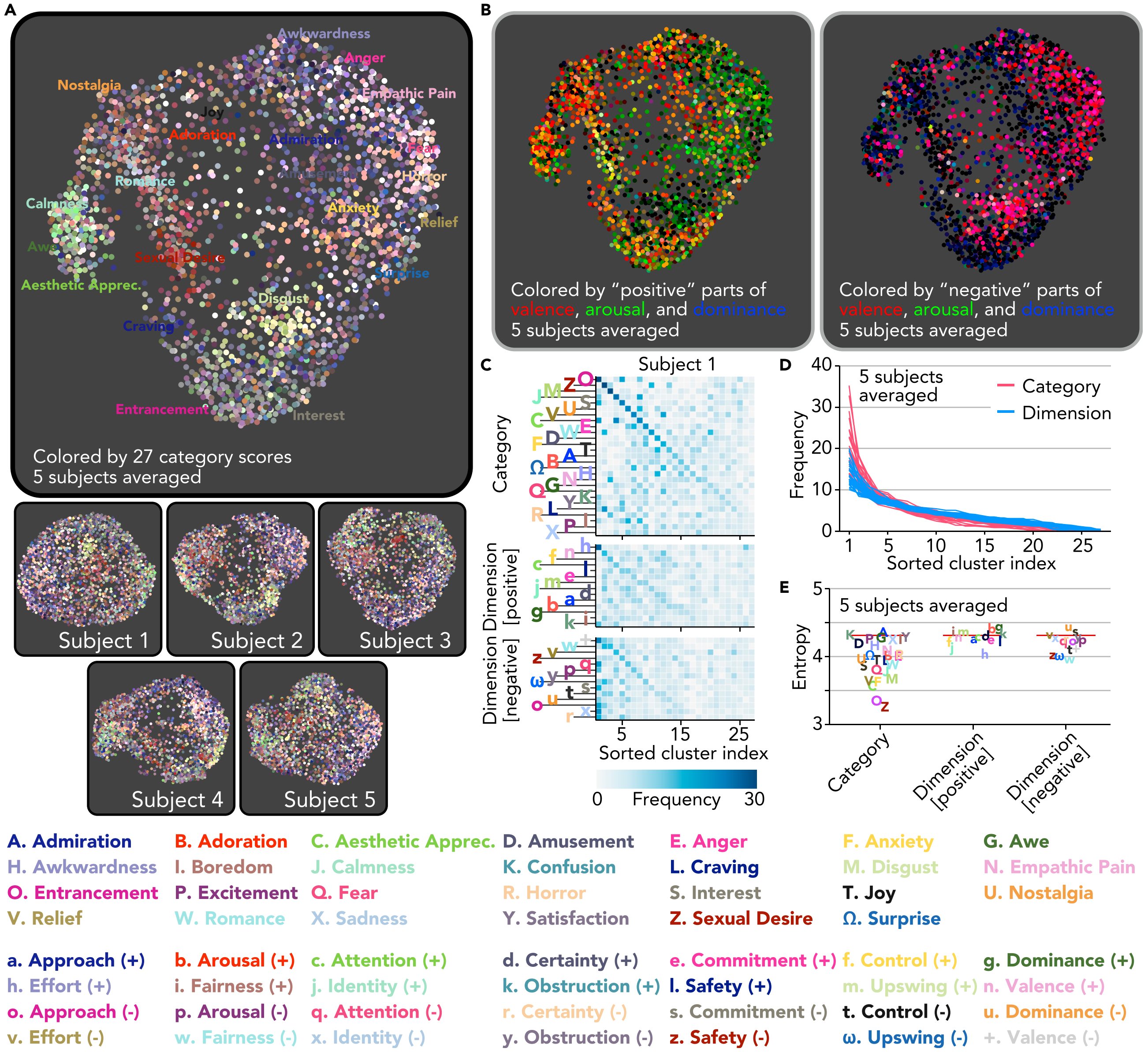


Figure 6. Distribution of Brain Activity Patterns Induced by Emotional Videos

(A) Maps of emotional experiences constructed from brain activity patterns. Each dot corresponds to each video. The same color schema as in Figure 3D is used.

(B) Maps of emotional experiences colored by positive or negative parts of valence, arousal, and dominance scores.

(C) Distributions of top 5% high score samples of each emotion on 27 clusters derived from the brain (Subject 1, see Figure S6A for the other subjects). Clusters were separately sorted for the sets of the categories and positive/negative dimensions.

(D) Sorted histograms for individual emotions. Lines indicate sorted histograms of 27 categories or 28 positive/negative dimensions (five subjects averaged, see Figures S6B and S6D for individual subjects and for results with different numbers of clusters).

(E) Entropy of the top 5% high-score sample distributions for each emotion (five subjects averaged; red lines, baseline, permutation test, $p < 0.01$, Bonferroni correction by the number of emotions; see Figures S6C and S6E for individual subjects and for results with different numbers of clusters).

voxel-wise encoding modeling, and unsupervised modeling methods. We found that ratings of individual emotions could accurately be predicted from activity patterns in many brain regions, revealing that distributed brain networks contributed in distinct ways to the representation of individual emotions in a highly consistent manner across subjects. Using voxel-wise encoding models constructed from the category and dimension ratings, we found that the brain representation of emotion categories explained greater variability in brain activity than that of affective dimensions in almost all brain regions. By comparing encoding performances between visual object, semantic, and emotion models, we confirmed that responses in a variety of brain regions were explained by features related to emotion rather than visual and semantic covariates. Those emotion-related representations were broadly distributed but commonly situated at the far end of a cortical gradient that runs from unimodal to transmodal regions. Analyses using unsupervised modeling revealed that the distribution of brain activity patterns evoked by emotional videos was structured with clusters associated with particular emotion categories bridged by gradients between related emotions. Taken together, our results demonstrate that neural representations of diverse emotional experiences during video viewing are high-dimensional, categorical, and distributed across transmodal brain regions.

Our decoding analysis revealed that brain regions that contributed to the representation of individual emotions were distinct for different emotions and similar across individuals (Figure 2). Emotions were represented not by simple one-to-one mappings between particular emotions and brain regions (e.g., fear and the amygdala) but by more complex configurations across multiple networks, consistent with notions of distributed, network-based representations of specific emotions (Hamann, 2012; Lindquist and Barrett, 2012; Saarimäki et al., 2016). In previous studies, emotion representations were found largely in transmodal brain regions, which are activated by tasks that rely upon social cognition, autobiographical memory, reward-based decision making, and other higher cognitive functions (Margulies et al., 2016). One exception from our analysis is that somatosensory regions were also found to encode empathic pain (Figure 2), consistent with notions that our empathy for others' injuries arises in part via mental simulation (Bufalari et al., 2007). Overall, these findings may reflect how experiences of individual emotions involve multiple interacting systems that automatically enter distinct states in response to significant threats and opportunities in the environment (Cowen, 2019; Seo et al., 2019). Given that these systems encode emotional experience (or affordances) even during passive viewing of videos, these findings support the view that emotions rely on system-wide adaptations that proactively recruit psychological faculties likely to support adaptive behavioral responses even before a decision is made to respond.

Our findings reconcile a history of observations that localized brain lesions can have highly specific effects on emotional behavior (e.g., Buchanan et al., 2004; Calder et al., 2000; Ciaramelli et al., 2013; Kim et al., 1993; Scott et al., 1997) with findings that these behaviors lack simple one-to-one mappings onto activity in localized brain regions (Lindquist and Barrett, 2012). If a given emotion is represented in a specific mode of activity within a network of interacting brain regions, damage to various hubs within this network would have the potential to disrupt specific behaviors associated with that emotion in differential ways (a premise further supported by findings from systems neuroscience in animal models, e.g., Kim et al., 1993; Seo et al., 2019). Our findings reveal how a set of transmodal brain regions centered in the default mode network undergo different modes of activity for different emotions. Intriguingly, the default mode network can in this way be considered to have not just one mode but many modes of activity corresponding to different emotions, with varying subregions of the network being differentially involved in each emotion (Figure 2).

Whether the neural encoding of specific emotion categories (e.g., ''anger'') can be reduced to the encoding of few broad affective dimensions such as valence, arousal, and certainty is an active area of debate that informs theories of emotion (Lindquist and Barrett, 2012; Kragel et al., 2019; Saarimäki et al., 2018). Our encoding and decoding analyses demonstrate that emotion categories reliably capture neural representations that are more specific than those captured by a wide range of affective dimensions throughout the brain (Figure 4). These results converge with recent behavioral studies revealing how emotion categories organize reported emotional experiences and emotion recognition from nonverbal signaling behavior (Cowen et al., 2019a; Cowen and Keltner, 2017). Previous studies have pitted categorical and dimensional emotion models against one another (Hamann, 2012), but done so with only a few emotions (e.g., happiness, anger, fear, sadness, disgust, and surprise) or two dimensions (valence, arousal), conflating distinct emotional experiences within and across studies. In the current study, we drew upon recent advances in the range of emotions considered to be distinct and fitted models accounting for a fuller array of appraisal dimensions of theoretical relevance to enable a more robust and thorough comparison of categorical and dimensional models of emotion representation. Our results provide strong support for a categorical approach to emotional experience, or theories that propose neural adaptations for a large set of appraisal dimensions specific enough to account for dozens of distinct emotion categories (e.g., Skerry and Saxe, 2015), models of which would consequently bear the same information as our category model (although further work would be needed to establish such a comprehensive set of appraisals; see Cowen et al., 2019b for Discussion). Additionally, using unsupervised modeling methods, we found that the distribution of emotion-related brain activity patterns in regions encoding emotion had cluster-like formations corresponding to particular emotion categories but not broad appraisals (Figure 6A), further supporting the conceptualization of the emotional response in terms of self-reported experiences such as "anxiety," "disgust," and "entrancement." Taken together, these results point to a high-dimensional, categorical space in the neural representation of emotion.

As done in previous studies using voxel-wise modeling in specific sensory regions (de Heer et al., 2017; Güçlü and van Gerven, 2015; Kell et al., 2018; Lescroart and Gallant, 2018), we compared performances of encoding models constructed from different features (emotion scores, visual object features, and semantic features) and uncovered a distributed set of brain regions that encode emotional experience (Figure 5). The spatial arrangement of these emotion-related brain regions overlapped with that of the default mode network (Figure 5C) and corresponded to the far end of a gradient from unimodal to transmodal regions (Figures 5G and 5H) that are located at geodesically distant points from primary sensory cortices (Margulies et al., 2016). Our analysis in a single study revealed a gradual representational shift along the unimodal-transmodal gradient from visual to semantic to emotion information, perhaps reflecting a global hierarchy of integration and feature abstraction from sensory inputs (Huntenburg et al., 2018).

One issue worth noting is that, because each stimulus was presented only once, we were unable to estimate the proportion of systematic variance in voxel responses that was left unexplained by our models. Many voxel-wise encoding modeling studies have relied on repeated stimulus presentations to determine the upper bound of model prediction accuracies (Huth et al., 2016; Kell et al., 2018; Lescroart and Gallant, 2018). However, responses to emotional stimuli are likely to change with repeated exposure to the stimuli, given that many emotional responses (e.g., surprise) can be affected by past exposure and expectations. For this reason, we performed our analyses on data from single trials. Nevertheless, the model prediction accuracies we observed were moderately high by fMRI standards. This might be attributable in part to the potency of naturalistic stimuli in evoking widespread brain responses (Hamilton and Huth, 2018) and, more specifically, to the emotional intensity of many of the video stimuli we used (Cowen and Keltner, 2017).

On this point, it is worth considering how investigations of brain representations of emotion may depend on the richness of the experimental stimuli used. The 2,181 videos studied here captured a wide spectrum of psychologically significant contexts. By contrast, most previous work has investigated neural underpinnings of emotion using a far narrower range of eliciting stimuli and a small set of emotions (Barrett, 2017; Celeghin et al., 2017; Giordano et al., 2018; Hamann, 2012; Kragel and LaBar, 2015; Lindquist and Barrett, 2012; Saarimäki et al., 2016; Satpute and Lindquist, 2019; Wager et al., 2015). Only recently have several studies investigated neural representations of emotion with richer models of emotion, but they have still relied on a relatively narrow range of stimuli (Koide-Majima et al., 2018; Kragel et al., 2019; Saarimäki et al., 2018; Skerry and Saxe, 2015). In assembling the present stimulus set, Cowen and Keltner (2017) drew on the recent availability of highly evocative stimuli on the internet to capture more extensive, diverse, and poignant stimuli, although it is worth noting that these stimuli are still only visual in nature. Future studies could build on this work by including auditory content, content with personal significance to the subject (e.g., imagery of loved ones), or tasks that include behavioral involvement of the subject (e.g., games, social interaction). As the field of affective science moves forward, the present study reveals how a more extensive analysis of whole-brain responses to a much richer and more extensive set of experimental conditions, open to diverse analytical approaches, can yield more definitive results.

This approach allowed rich and nuanced emotion categories to be accurately predicted from brain activity patterns (Figures 2A and 2B) and even made it possible to accurately identify eliciting stimuli based on patterns of decoded emotion scores (Figure 3). Previous decoding studies of emotion have relied on discrete classification into a small set of emotion categories (Kragel et al., 2016; Kragel and LaBar, 2015; Saarimäki et al., 2016, 2018; Sitaram et al., 2011), which may conflate representations of diverse emotional states grouped into each category (Clark-Polner et al., 2017). By predicting continuous intensities of many distinct emotions simultaneously, a more comprehensive picture of emotional states can be characterized. We visualized these nuanced emotional states by combining decoded scores of richly varying emotion categories with the UMAP algorithm (Figure 3D). This approach can provide a tool for various applications, such as product evaluation in neuro-marketing (Nishida and Nishimoto, 2018). Furthermore, in light of recent findings of a broad overlap between neural representations of perceived and internally generated experiences (Horikawa et al., 2013; Kragel et al., 2016; Tusche et al., 2014), decoders trained for perceived emotions may generalize to emotional states that occur spontaneously during dreams and mind wandering, with potential clinical applications (Kragel et al., 2016; Sitaram et al., 2011).

A critical question in neuroscience is where emotions evoked by different modalities of sensory input are encoded in the brain (Chikazoe et al., 2014; Kragel et al., 2016; Skerry and Saxe, 2014). Although here we used only visual stimuli (with no sound), we found that emotion-related representations were not primarily found in the visual cortex (Kragel et al., 2019) but in many transmodal and frontal regions that have been implicated in representing other modalities of emotionally evocative stimuli and tasks, including verbal stimuli (Huth et al., 2016; Skerry and Saxe, 2015), visual and gustatory stimuli (Chikazoe et al., 2014), and volitional and spontaneous imagery (Tusche et al., 2014). A few studies have investigated commonalities in the neural representation of emotion across multiple stimulus modalities but have focused on narrower ranges of stimuli and emotions (e.g., basic emotions: Peelen et al., 2010; valence: Chikazoe et al., 2014; Skerry and Saxe, 2015). As behavioral studies generate richer and diverse experimental resources for investigating multiple modalities of emotions (e.g., resources for emotional responses to facial expression, speech prosody, nonverbal vocalization, music, games, and social interaction; Cowen and Keltner, 2019; Cowen et al., 2018, 2019a, 2020), cross-modal neural representations of emotions should be investigated more deeply with richer models of emotion.

Limitations of the Study

One limitation of the present study was that the emotion ratings we used were averages from multiple raters and not from the subjects of our fMRI experiment. Although our decoding and encoding models were able to establish systematic statistical relationships between those ratings and the brain activity patterns of our subjects, the reported feelings of those raters may not completely match those of our subjects. Hence, to the extent that neural representations of emotion are subject to individual differences in culture, demographic variables (gender, age), and personality, future studies might uncover more robust neural representations by incorporating these variables.

Another potential concern regarding the present investigation is the confounding of first-person and third-person emotions. The brain regions in which the category model outperformed the visual object and semantic models (e.g., IPL, VMPFC) overlapped with brain regions that have been implicated in a ‘‘theory of mind network’’ (Skerry and Saxe, 2015). It is possible that the neural representations uncovered here reflect information not only regarding first-person emotional experiences but also regarding predictions of how others might feel. Although the observed overlap may also indicate common neural substrates for first- and third-person emotions, future experiments are needed to dissociate these phenomena using more carefully designed experiments, perhaps including psychophysiological measures indicative of first-person emotional experience.

Finally, as we allowed the subjects to freely view presented video stimuli, differential gaze patterns for each emotional experience might have some influence on our brain data. Although the neural representation of gaze is known to be highly localized in specific brain regions (Connolly et al., 2005; Sestieri et al., 2008; Culham et al., 1998), modeling effects of emotion on gaze patterns would help to more precisely characterize neural representations associated with subjective feelings of emotions.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND CODE AVAILABILITY

The experimental code and data that support the findings of this study are available from our repository (https://github.com/KamitaniLab/EmotionVideoNeuralRepresentation) and open data repositories (OpenNeuro: https://openneuro.org/datasets/ds002425; Mendeley Data: https://doi.org/10.17632/jbk2r73mzh.1 figshare: https://doi.org/10.6084/m9.figshare.11988351.v1).

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101060.

ACKNOWLEDGMENTS

The authors thank Mitsuaki Tsukamoto, Hiroaki Yamane, and Shuntaro C. Aoki for help with data collection and software setups. This study was conducted using the MRI scanner and related facilities of Kokoro Research Center, Kyoto University. This research was supported by grants from the New Energy and Industrial Technology Development Organization (NEDO); Tateishi Science and Technology Foundation; JSPS KAKENHI Grant number JP15H05710, JP15H05920, and JP17K12771; JST CREST Grant Number JPMJCR18A5; and JST PRESTO Grant Number JPMJPR185B Japan.

AUTHOR CONTRIBUTIONS

Conceptualization, T.H. and Y.K.; Methodology, T.H.; Validation, T.H.; Formal Analysis, T.H.; Investigation, T.H.; Resources, Y.K.; Writing – Original Draft, T.H.; Writing – Review & Editing, T.H., A.S.C., D.K., and Y.K.; Visualization, T.H.; Supervision, Y.K.; Funding Acquisition, T.H. and Y.K.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 27, 2019
Revised: March 11, 2020
Accepted: April 9, 2020
Published: May 22, 2020

REFERENCES

Barrett, L.F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. Soc. Cogn. Affect. Neurosci. *12*, 1–12.

Barrett, L.F., and Satpute, A.B. (2013). Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. Curr. Opin. Neurobiol. *23*, 361–372.

Buchanan, T.W., Tranel, D., and Adolphs, R. (2004). Anteromedial temporal lobe damage blocks startle modulation by fear and disgust. Behav. Neurosci. *118*, 429–437.

Bufalari, I., Aprile, T., Avenanti, A., Di Russo, F., and Aglioti, S.M. (2007). Empathy for pain and touch in the human somatosensory cortex. Cereb. Cortex *17*, 2553–2561.

Calder, A.J., Keane, J., Manes, F., Antoun, N., and Young, A.W. (2000). Impaired recognition and experience of disgust following brain injury. Nat. Neurosci. *3*, 1077–1078.

Celeghin, A., Diano, M., Bagnis, A., Viola, M., and Tamietto, M. (2017). Basic emotions in human neuroscience: neuroimaging and beyond. Front. Psychol. *8*, 1432.

Chikazoe, J., Lee, D.H., Kriegeskorte, N., and Anderson, A.K. (2014). Population coding of affect across stimuli, modalities and individuals. Nat. Neurosci. *17*, 1114–1122.

Ciaramelli, E., Sperotto, R.G., Mattioli, F., and di Pellegrino, G. (2013). Damage to the ventromedial prefrontal cortex reduces interpersonal disgust. Soc. Cogn. Affect. Neurosci. *8*, 171–180.

Clark-Polner, E., Johnson, T.D., and Barrett, L.F. (2017). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. Cereb. Cortex *27*, 1944–1948.

Connolly, J.D., Goodale, M.A., Goltz, H.C., and Munoz, D.P. (2005). fMRI activation in the human frontal eye field is correlated with saccadic reaction time. J. Neurophysiol. *94*, 605–611.

Cowen, A.S. (2019). Neurobiological explanation for diverse responses associated with a single emotion. Sci. eLetter. https://science.sciencemag.org/content/363/6426/538/tab-e-letters.

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

Cowen, A.S., and Keltner, D. (2018). Clarifying the conceptualization, dimensionality, and structure of emotion: response to Barrett and colleagues. Trends. Cogn. Sci. *22*, 274–276.

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

Cowen, A.S., Elfenbein, H.A., Laukka, P., and Keltner, D. (2018). Mapping 24 emotions conveyed by brief human vocalization. Am. Psychol. *74*, 698–712.

Cowen, A.S., Fang, X., Sauter, D., and Keltner, D. (2020). What music makes us feel: at least thirteen dimensions organize subjective experiences associated with music across cultures. Proc. Natl. Acad. Sci. U S A *117*, 1924–1934.

Cowen, A.S., Laukka, P., Elfenbein, H.A., Liu, R., and Keltner, D. (2019a). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. Nat. Hum. Behav. *13*, 1–16.

Cowen, A.S., Sauter, D., Tracy, J.L., and Keltner, D. (2019b). Mapping the passions: toward a high-dimensional taxonomy of emotional experience and expression. Psychol. Sci. Public Interest *20*, 69–90.

Culham, J.C., Brandt, S.A., Cavanagh, P., Kanwisher, N.G., Dale, A.M., and Tootell, R.B.H. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. J. Neurophysiol. *80*, 2657–2670.

de Heer, W.A., Huth, A.G., Griffiths, T.L., Gallant, J.L., and Theunissen, F.E. (2017). The hierarchical cortical organization of human speech processing. J. Neurosci. *37*, 6539–6557.

Ekman, P., and Friesen, W.V. (1969). The repertoire of nonverbal behavior: categories, origins, usage and coding. Semiotica *1*, 49–98.

Giordano, B.L., Whiting, C., Kriegeskorte, N., Kotz, S.A., Belin, P., and Gross, J. (2018). From categories to dimensions: spatio-temporal dynamics of the cerebral representations of emotion in voice. bioRxiv. https://doi.org/10.1101/265843.

Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Hacker, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. Nature *536*, 171–178.

Güçlü, U., and van Gerven, M.A.J. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. J. Neurosci. *35*, 100005–100014.

Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: controversies and consensus. Trends. Cogn. Sci. *16*, 458–466.

Hamilton, L.S., and Huth, A.G. (2018). The revolution will not be controlled: natural stimuli in speech neuroscience. Lang. Cogn. Neurosci. *0*, 1–10.

Huntenburg, J.M., Bazin, P.-L., and Margulies, D.S. (2018). Large-scale gradients in human cortical organization. Trends Cogn. Sci. *22*, 21–31.

Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., and Gallant, J.L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. Nature *532*, 453–458.

Huth, A.G., Nishimoto, S., Vu, A.T., and 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.

Horikawa, T., and Kamitani, Y. (2017). Generic decoding of seen and imagined objects using hierarchical visual features. Nat. Commun. *8*, 15037.

Horikawa, T., Tamaki, M., Miyawaki, Y., and Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. Science *340*, 639–642.

Kell, A.J.E., Yamins, D.L.K., Shook, E.N., Norman-Haignere, S.V., and McDermott, J.H. (2018). A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. Neuron *98*, 630–644.

Kim, J.J., Rison, R.A., and Fanselow, M.S. (1993). Effects of amygdala, hippocampus, and periaqueductal gray lesions on short- and long-term contextual fear. Behav. Neurosci. *107*, 1093–1098.

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

Koide-Majima, N., Nakai, T., and Nishimoto, S. (2018). Distinct dimensions of emotion in the human brain and their representation on the cortical surface. bioRxiv. https://doi.org/10.1101/464636.

Kragel, P.A., Knodt, A.R., Hariri, A.R., and LaBar, K.S. (2016). Decoding spontaneous emotional states in the human brain. Plos Biol. *14*, e2000106–e2000119.

Kragel, P.A., Reddan, M.C., LaBar, K.S., and Wager, T.D. (2019). Emotion schemas are embedded in the human visual system. Sci. Adv. *5*, eaaw4358.

Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P.A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron *60*, 1126–1141.

Lescroart, M.D., and Gallant, J.L. (2018). Human scene-selective areas represent 3D configurations of surfaces. Neuron *101*, 1–33.

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

Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., et al. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc. Natl. Acad. Sci. U S A *101*, 12574–12579.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold approximation and projection for dimension reduction. arXiv. https://arxiv.org/abs/1802.03426.

Nishida, S., and Nishimoto, S. (2018). Decoding naturalistic experiences from human brain activity via distributed representations of words. Neuroimage *180*, 232–242.

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

Plutchik, R. (1980). Emotion: A Psychoevolutionary Synthesis (Harper and Row).

Posner, J., Russell, J.A., and Petersona, B.S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev. Psychopathol. *17*, 715–734.

Russell, J.A. (1980). A circumplex of affect. J. Pers. Soc. Psychol. *36*, 1152–1168.

Russell, J.A. (2003). Core affect and the psychological construction of emotion. Psychol. Rev. *110*, 145–172.

Saarimäki, H., Ejtehadian, L.F., Glerean, E., Jääskeläinen, I.P., Vuilleumier, P., Sams, M., and Nummenmaa, L. (2018). Distributed affective space represents multiple emotion categories across the human brain. Soc. Cogn. Affect. Neurosci. *13*, 471–482.

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I.P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., and Nummenmaa, L. (2016). Discrete neural signatures of basic emotions. Cereb. Cortex *26*, 2563–2573.

Satpute, A.B., and Lindquist, K.A. (2019). The default mode network's role in discrete emotion. Trends Cogn. Sci. *23*, 851–864.

Scott, S.K., Young, A.W., Calder, A.J., Hellawell, D.J., Aggleton, J.P., and Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. Nature *385*, 254–257.

Seo, C., Guru, A., Jin, M., Ito, B., Sleezer, B.J., Ho, Y.Y., Wang, E., Boada, C., Krupa, N.A., Kullakanda, D.S., et al. (2019). Intense threat switches dorsal raphe serotonin neurons to a paradoxical operational mode. Science *363*, 538–542.

Sestieri, C., Pizzella, V., Cianflone, F., Romani, G.L., and Corbetta, M. (2008). Sequential activation of human oculomotor centers during planning of visually-guided eye movements: a combined fMRI-MEG study. Front. Hum. Neurosci. *1*, 1.

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv. https://arxiv.org/abs/1409.1556.

Sitaram, R., Lee, S., Ruiz, S., Rana, M., Veit, R., and Birbaumer, N. (2011). Real-time support vector classification and feedback of multiple emotional brain states. Neuroimage *56*, 753–765.

Skerry, A.E., and Saxe, R. (2014). A common neural code for perceived and inferred emotion. J. Neurosci. *34*, 15997–16008.

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

Smith, C.A., and Ellsworth, P.C. (1985). Patterns of cognitive appraisal in emotion. J. Pers. Soc. Psychol. *48*, 813–838.

Tusche, A., Smallwood, J., Bernhardt, B.C., and Singer, T. (2014). Classifying the wandering mind: revealing the affective content of thoughts during task-free rest periods. Neuroimage *97*, 107–116.

Wager, T.D., Kang, J., Johnson, T.D., Nichols, T.E., Satpute, A.B., and Barrett, L.F. (2015). A bayesian model of category-specific emotional brain responses. PLoS Comput. Biol. *11*, e1004066.

Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. *106*, 1125–1165.

iScience, Volume *23*

Supplemental Information

The Neural Representation of Visually Evoked Emotion Is High-Dimensional, Categorical, and Distributed across Transmodal Brain Regions

Tomoyasu Horikawa, Alan S. Cowen, Dacher Keltner, and Yukiyasu Kamitani

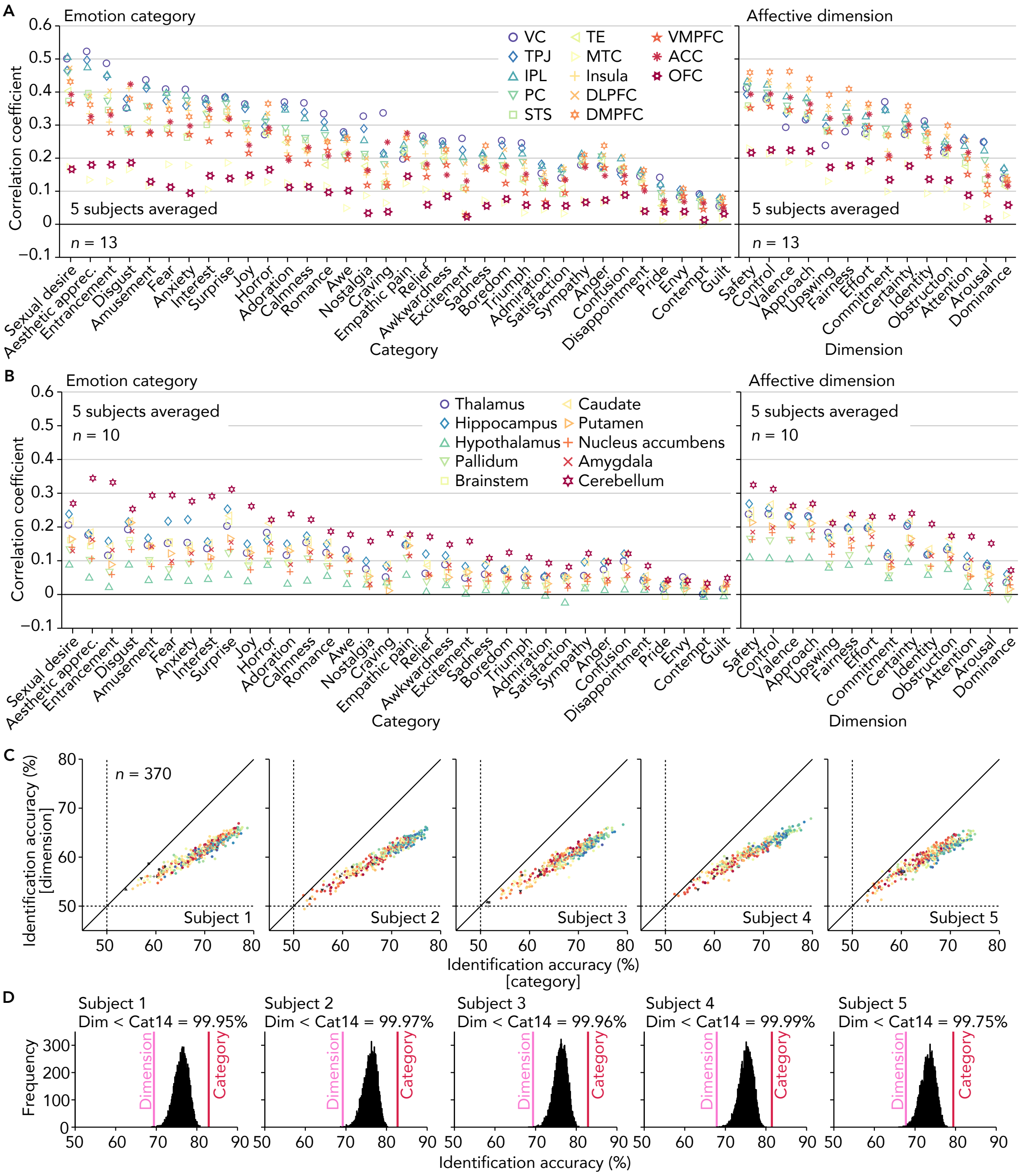


Figure S1. Performances of decoding analysis for emotion scores. (Related to Figures 2 and 3)

(A) Decoding accuracy for individual emotions predicted from activities in representative cortical regions. The decoding analysis of individual emotion scores (cf., Figures 2A and B) was performed from brain activity patterns in several cortical regions (see Transparent Methods: “Regions of interest (ROI)” for definitions of individual cortical regions). Dots indicate accuracies obtained from individual cortical regions (five subjects averaged).

(B) Decoding accuracy for individual emotions predicted from activities in subcortical regions. Conventions are the same as (A).

(C) Mean video identification accuracies from region-wise decoders of individual subjects. Conventions are the same as Figure 3A.

(D) Distributions of video identification accuracies obtained from randomly selected 14 emotion category scores. The video identification analysis by ensemble decoders (cf., Figure 3B) was performed for 10,000 times while randomly selecting different combinations of 14 emotion category scores from the original 34 emotion category scores. The identification accuracies obtained with this procedure were compared with the accuracy from the 14 affective dimensions. The estimated accuracies from more than 99% of 14 randomly selected emotion categories outperformed the accuracy from the 14 affective dimensions. The results suggest that the superiority of the 34 emotion categories over the 14 affective dimensions (Figure 3B) were not solely due to the differences of the number of emotions used for identification.

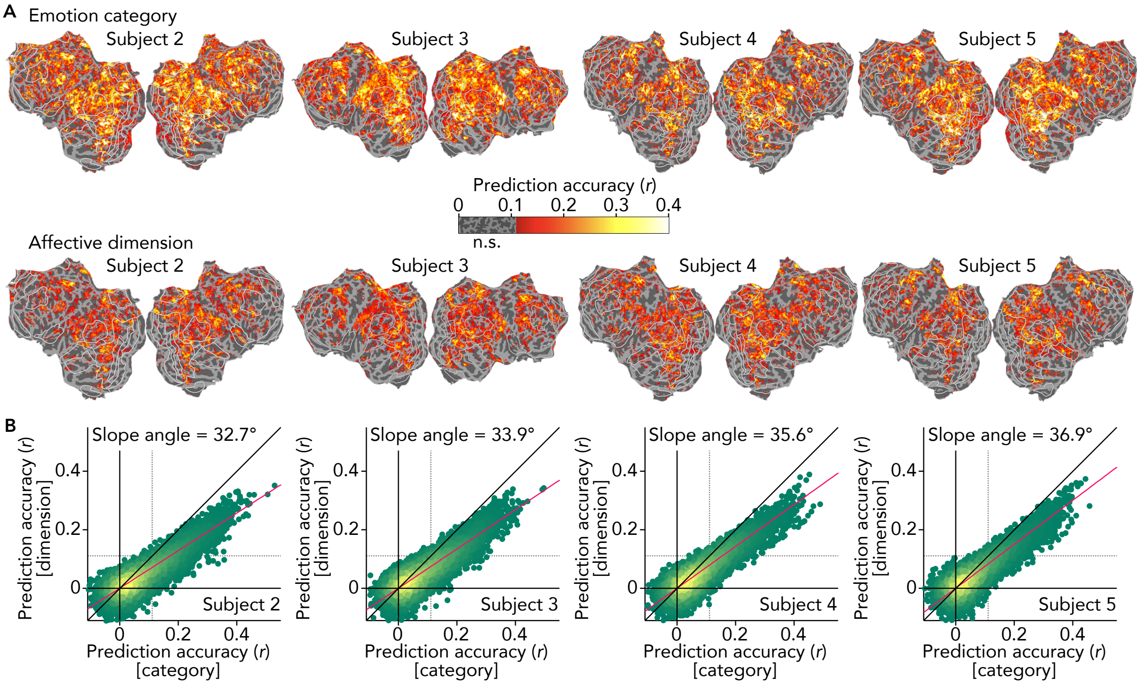


Figure S2. Performance of encoding models constructed from emotional scores for individual subjects. (Related to Figure 4)

(A) Prediction accuracies of emotion encoding models. Conventions are the same with Figure 4A.

(B) Prediction accuracies of individual voxels. Conventions are the same with Figure 4B.

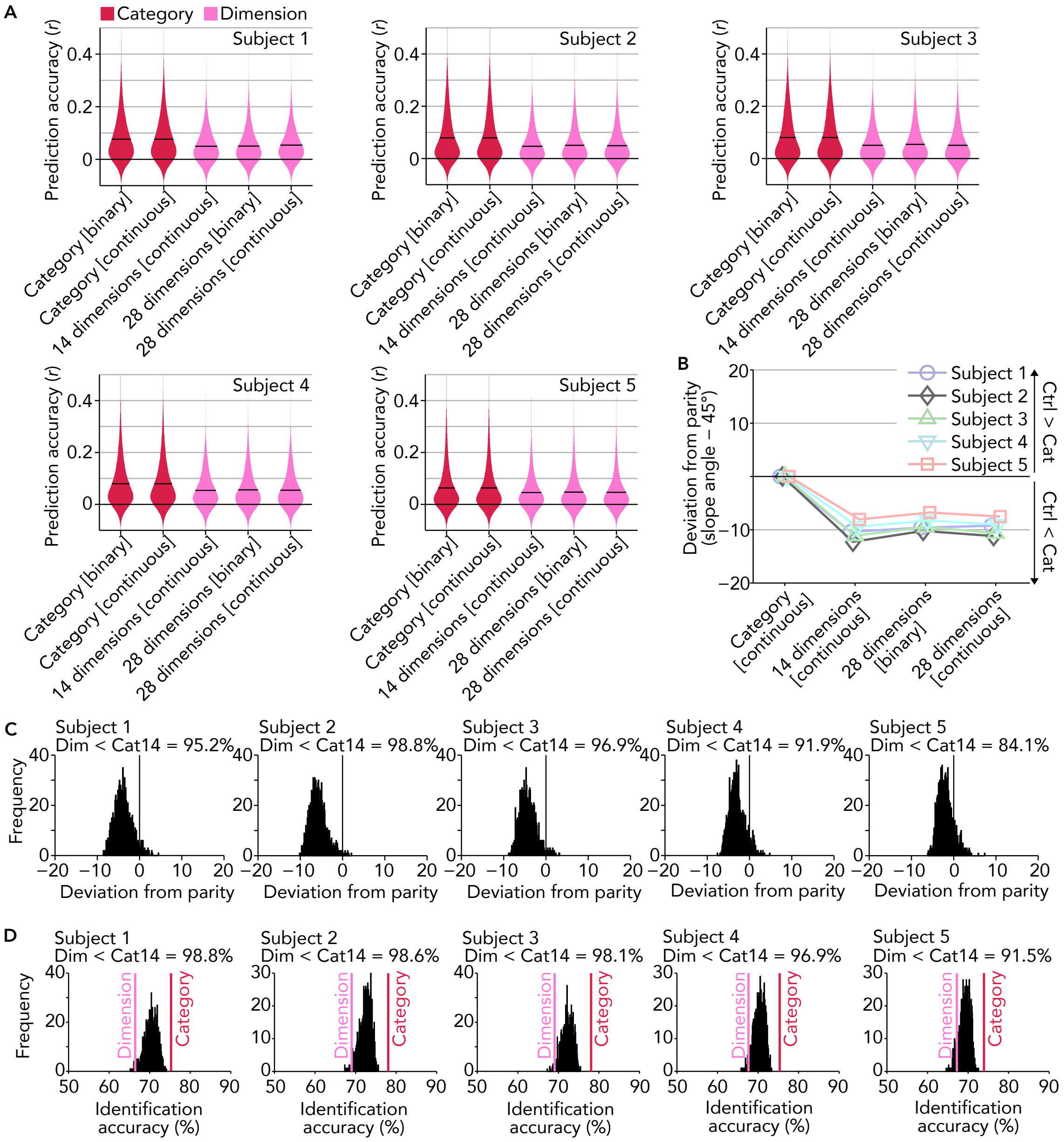


Figure S3. Control analyses for the performance comparisons between category and dimension encoding models. (Related to Figure 4)

(A) Distributions of prediction accuracies of all voxels from multiple variants of category models and dimension models. Because the methods for collecting scores of emotion categories and affective dimensions were different (see Transparent Methods: "Video stimulus labeling"), differences of encoding performances might be attributable to such differences. To compensate for the differences, we constructed encoding models from different versions of emotion category and affective dimension scores used in the main analyses, and compared encoding model performances from those multiple variants. For the emotion category scores, we have used binarized scores reported by individual human raters (originally ranged between 0 to 100) in the main analysis ("category [binary]"; cf., Figure 4). We here also tested another type of emotion category scores without binarization (mean of reported scores [ranged between 0 to 100], “category [continuous]”). In either case, high and low values of the emotion category scores can be interpreted whether that emotions exist or not. On the other hand, because scores of the affective dimensions were collected with 9-scale Likert scale (ranged between 1 to 9, “14 dimensions [continuous]”), the high and low values should be interpreted differently (e.g., positive and negative), and low values should be interpreted as strong negative emotions rather than no emotion. To compensate for the difference from the category scores, we tested other types of affective dimension scores by first subtracting 5 [neutral] from original values (-4 to 4), taking both positive and negative values separately, and concatenating the positive (0 to 4) and negative (-4 to 0) parts to yield a total of 28 unipolar affective dimension scores (“28 dimensions [continuous]”). This score conversion was done for scores of individual raters and converted scores were averaged across multiple raters. Furthermore, similar to the emotion category scores, we also tested the binarized version of affective dimension scores by binarizing zero or non-zero values of the continuous 28-dimensional scores of individual raters to zero/one values and averaging them across multiple raters (“28 dimensional [binary]”). These score conversions had little effect on encoding accuracies, showing higher mean encoding accuracies for the emotion category models than the affective dimension models.

(B) Comparisons of prediction accuracies between the emotion category models and the other control models. The encoding performance obtained from the original emotion category model was directly compared with the performances from models constructed with the other variants of emotion scores (cf., Figure S3A) based on slope angles of best linear fit estimated from encoding accuracies of all voxels across the whole brain. The results showed equivalent performances between the two category models, whereas the original category model outperformed all affective dimension models, suggesting that the differences of the data collection methods were not main factor of the superiority of the emotion category model.

(C) Distributions of deviations of slope angles from the parity compared between the dimension models and category models constructed with randomly selected 14 emotion category scores. Encoding models were constructed from 14 randomly selected emotion categories for 1,000 times, and obtained encoding accuracies were each compared with the accuracy from the 14 affective dimensions based on slope angles of best linear fit estimated from encoding accuracies of all voxels across the whole brain. The results showed that on average more than 93.4% models constructed from 14 randomly selected emotion categories outperformed the accuracy from the 14 affective dimensions (five subjects averaged).

(D) Distributions of video identification accuracies obtained from encoding models constructed from randomly selected 14 emotion categories. The video identification analysis (cf., Figure 4F) was performed with models constructed with 14 randomly selected emotion categories for 1,000 times (cf., Figure S3C). The identification accuracies obtained with this procedure were compared with the accuracy from the 14 affective dimension model (Figure 4F). The results showed that on average more than 96.8% models obtained from 14 randomly selected emotion categories outperformed the accuracy from the 14 affective dimensions (five subjects averaged). Taken together with (C), the results suggest that the superiority of the 34 emotion categories over the 14 affective dimensions were not solely due to the differences of the number of emotions used for the encoding analysis.

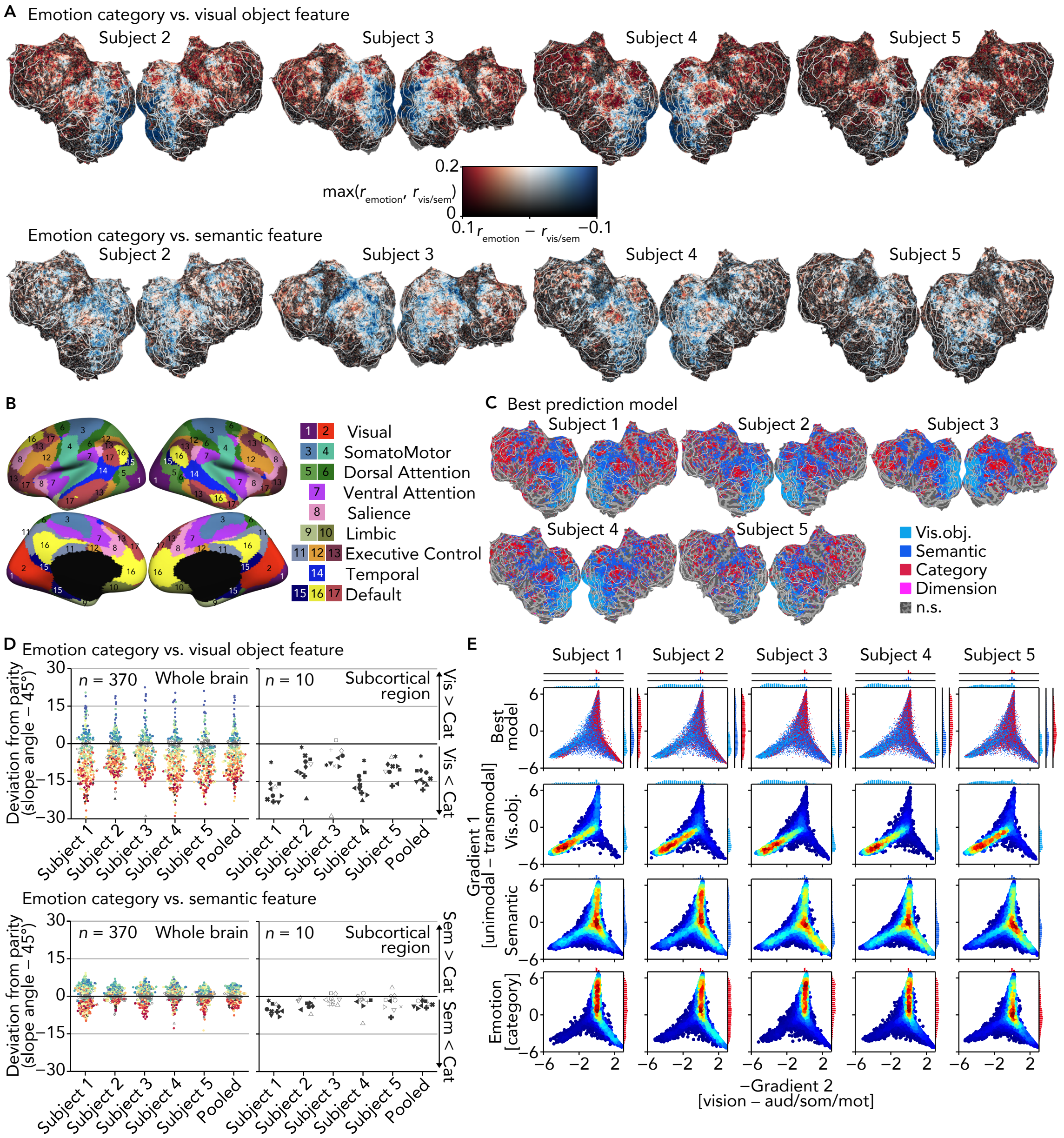


Figure S4. Comparisons of encoding accuracy based on emotion, visual object, and semantic models for individual subjects. (Related to Figure 5)

(A) Differences in prediction accuracies of emotion, visual object, and semantic models.

(B) Definition of global networks (see Yeo et al., 2011 for details).

(C) Best models among visual object, semantic, category, and dimension models.

(D) Comparisons of prediction accuracies for individual brain regions. Conventions are the same as Figure 4D. To examine the similarity of the distributions of emotion-related regions across subjects, the Pearson correlation coefficients were calculated between the estimated slope angles of all brain regions (*n* = 370) from each pair of subjects (*n* = 10). The analysis showed highly positive correlations for both comparisons between the emotion and visual object models (*r* = 0.843, averaged across pairs) and between the emotion and semantic models (*r* =

0.624, averaged across pairs), suggesting that the distributions of emotion-related brain regions were similar across subjects.

(E) Joint and marginal distributions of the best models in principal gradient space for individual subjects. Conventions are the same as Figure 5G.

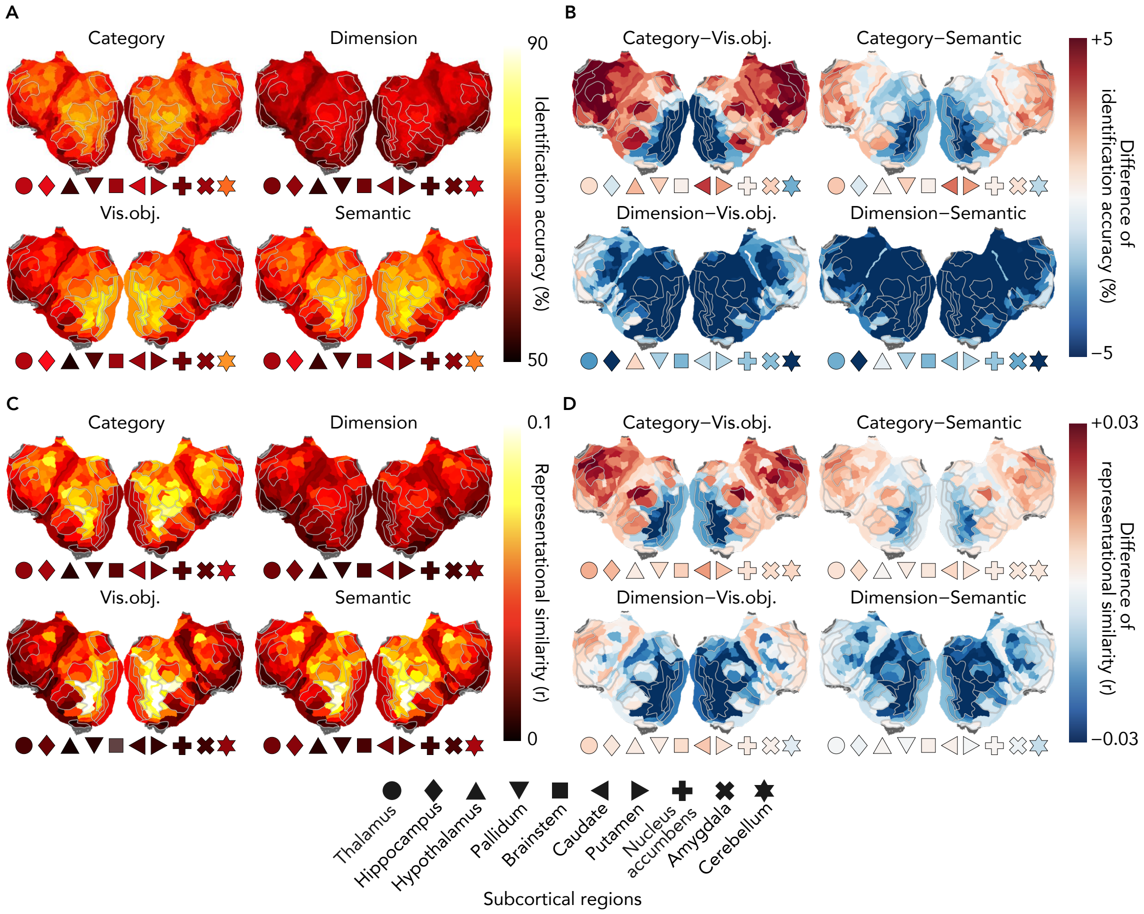


Figure S5. Performances of decoding analysis and representational similarity analysis across the whole cortex. (Related to Figure 5)

(A) Video identification accuracies via decoding analysis of individual features (five subjects averaged). The video identification analysis performed using decoded emotional scores (cf., Figures 3A and S1C) was also performed with visual object features and semantic features via decoding analysis of individual features. Mean accuracies of the identification analysis via decoded scores/features from five subjects (cf., Figure S1C) were averaged for each brain region, and were projected onto the cortical surface of Subject 1.

(B) Differences in video identification accuracies via decoding analysis between two different features. Conventions are the same as (A).

(C) Representational similarities for individual feature set. The representational similarity analysis (Kriegeskorte et al., 2008) was performed to evaluate the similarity of representational similarity matrices (RSMs) constructed from patterns of brain activities and scores/features (see Transparent Methods: “Representational similarity analysis” for details). For each set of features (e.g., 34 emotion category scores), correlation coefficients between two RSMs (one from the brain, and the other from scores/features) were calculated for individual brain regions (*n* = 370). The estimated representational similarities for individual brain regions were averaged across five subjects, and were projected onto the cortical surface of Subject 1.

(D) Differences in representational similarities between two different feature sets. Conventions are the same as (C).

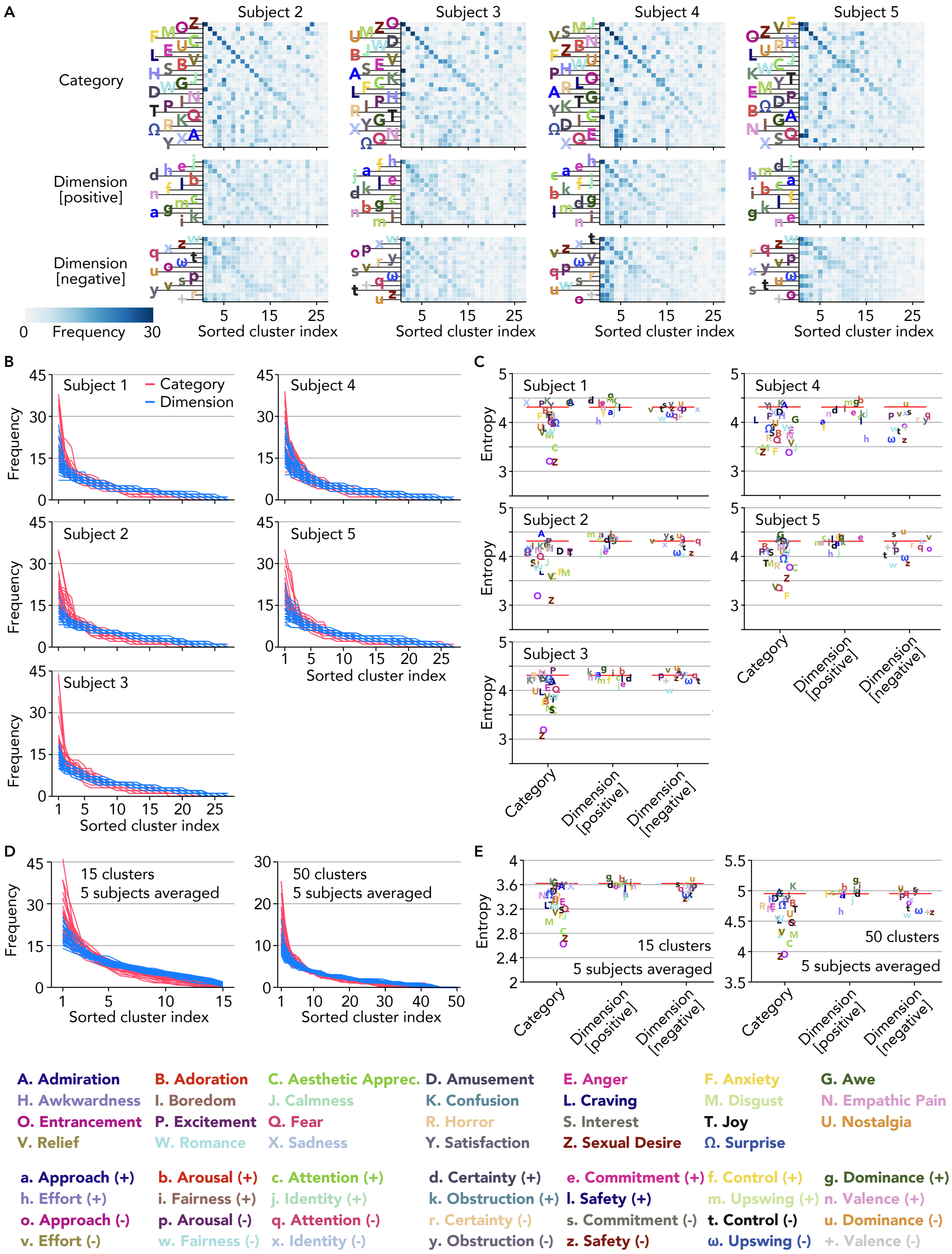


Figure S6. Clustering analysis on brain activity patterns induced by emotion evocative short videos for individual subjects. (Related to Figure 6)

(A) Distributions of top 5% high score samples of individual emotions on 27 clusters derived from brain activity patterns for individual subjects. Conventions are the same with Figure 6C.

(B) Sorted histograms of individual emotions for individual subjects. Conventions are the same with Figure 6D.

(C) Entropy of the top 5% high score sample distributions of each emotion for individual subjects. Conventions are the same with Figure 6E.

(D) Sorted histograms of individual emotions obtained with 15 and 50 clusters. Conventions are the same with Figures 6D.

(E) Entropy of the top 5% high score sample distributions of each emotion obtained with 15 and 50 clusters. Conventions are the same with Figures 6E.

Transparent Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Yukiyasu Kamitani (kamitani@i.kyoto-u.ac.jp).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects

Five healthy subjects with normal or corrected-to-normal vision participated in our experiments: Subject 1 (male, age 34), Subject 2 (male, age 23), Subject 3 (female, age 23), Subject 4 (male, age 22), and Subject 5 (male, age 27). This sample size was chosen on the basis of previous fMRI studies with similar experimental designs (Horikawa and Kamitani, 2017; Huth et al., 2016). All subjects provided written informed consent for participation in the experiments, and the study protocol was approved by the Ethics Committee of ATR.

METHOD DETAILS

Emotional movie stimuli

The stimuli consisted of sequences of emotionally evocative short videos. The videos were originally collected by Cowen and Keltner (2017). The video dataset consisted of a total of 2196 videos (downloaded at 13 September, 2017; https://goo.gl/forms/XErJw9sBeyuOyp5Q2) whose durations ranged from ~0.15 s to ~90 s. Each video was resized so that the width and height of videos were both within 12 degree (the original aspect ratio was preserved) and was visually presented at the center of the screen on gray background (no sound was delivered in our experiment, while some videos originally contained sounds).

Experimental design

All video stimuli were rear-projected onto a screen in the MRI scanner bore using a luminance-calibrated liquid crystal display projector. To minimize head movements during fMRI scanning, subjects were required to fix their heads using a custom-molded bite-bar individually made for each subject except for the case where subjects were reluctant to use the bite-bar (a subset of sessions with Subject 5). Data from each subject were collected over multiple scanning sessions spanning approximately 2 months. On each experimental day, one consecutive session was conducted for a maximum of 2 hours. Subjects were given adequate time for rest between runs (every 7–10 min) and were allowed to take a break or stop the experiment at any time.

The video presentation experiment consisted of a total of 61 separate runs. Each run comprised 36 stimulus blocks whose durations differed across blocks depending on the durations of videos presented in each stimulus block. For stimulus blocks with videos shorter than 8 s, the same video stimulus was repeatedly presented until the total presentation duration went beyond 8 s. For stimulus blocks with videos longer than 8 s, the video stimulus was presented once and followed by ~2-s rest period so that the total duration of each stimulus block can be divided by 2 s (TR). All stimulus blocks were followed by an additional 2-s rest period. Additional 32- and 6-s rest periods were added to the beginning and end of each run respectively. Consequently, the durations of individual runs ranged from 7 min 10 s to 9 min 54 s, and the total duration of all scanning sessions was about 8 hours.

To maximize subject's emotional responses to video stimuli, subjects were allowed to view video stimuli without fixation to let subjects freely focus on any details of events in videos. Subjects were requested to maintain steady fixation on the center fixation spot (0.3 × 0.3 degree) during rest periods to maintain their attention on the screen.

MRI acquisition

fMRI data were collected using a 3.0-Tesla Siemens MAGNETOM Verio scanner located at the Kokoro Research Center, Kyoto University. An interleaved T2*-weighted gradient-echo echo planar imaging (EPI) scan was performed to acquire functional images covering the entire brain (TR, 2000 ms; TE, 43 ms; flip angle, 80 deg; FOV, 192 × 192 mm; voxel size, 2 × 2 × 2 mm; slice gap, 0 mm; number of slices, 76; multiband factor, 4). T1-weighted (T1w) magnetization-prepared rapid acquisition gradient-echo (MP-RAGE) fine-structural images of the entire head were also acquired (TR, 2250 ms; TE, 3.06 ms; TI, 900 ms; flip angle, 9 deg; FOV, 256 × 256 mm; voxel size, 1.0 × 1.0 × 1.0 mm).

MRI data preprocessing

We performed the MRI data preprocessing through the pipeline provided by fMRIPrep (version 1.2.1; Esteban et al., 2019). For functional data of each run, first, a BOLD reference image was generated using a custom methodology of fMRIPrep. A deformation field to correct for susceptibility distortions was estimated based on fMRIPrep's fieldmap-less approach, and the estimated susceptibility distortion was used to calculate an unwarped BOLD reference for a more accurate coregistration with the anatomical reference. Using the estimated BOLD reference, data were motion corrected using mcflirt from FSL (version 5.0.9; Jenkinson et al., 2002) and then slice time corrected using 3dTshift from AFNI (version 16.2.07; Cox, 1996). This was followed by co-registration to the corresponding T1w image using boundary-based registration implemented by bbregister from FreeSurfer (version 6.0.1; Greve and Fischl, 2009). The coregistered BOLD time-series were then resampled onto their original space (2 × 2 × 2 mm voxels) using antsApplyTrainsforms from ANTs (version 2.1.0; Avants et al., 2008) using Lanczos interpolation.

Using the preprocessed BOLD signals, data samples were created by first regressing out nuisance parameters from each voxel amplitude for each run, including a constant baseline, a linear trend, and temporal components proportional to the six motion parameters calculated during the motion correction procedure (three rotations and three translations). The data were temporally shifted by 4 s (2 volumes) to compensate for hemodynamic delays, were despiked to reduce extreme values (beyond ± 3 SD for each run), and were averaged within each stimulus block (a video presentation period and a subsequent 2-s rest period). The data for all video stimuli were then further z-scored for each voxel. These procedures yielded a total of 2196 data samples each corresponding to each video stimulus. Because the presented video stimulus set happened to include identical videos (15 duplicates), we discarded samples that were presented later in the experiment from each of those duplicates, and used remaining 2181 unique samples in the following analyses.

For visualization of analytical results from the whole cortical areas, we visualized results using flattened cortical surfaces reconstructed from anatomical images of individual subjects. Cortical surface meshes of individual subjects were first generated from the T1w anatomical images using recon-all from FreeSurfer (version 6.0.1; Fischl, 2012). Relaxation cuts were made into the surface of each hemisphere to make flattened cortical surfaces for individual subjects. Functional data were aligned, and were projected onto the surface for visualization using Pycortex (Gao et al., 2015).

Regions of interest (ROI)

To define regions of interest (ROI) on cortical surfaces, we used two types of brain parcellations: 1) the whole cortical brain parcellation provided by the Human Connectome Project (Glasser et al., 2016), which delineated a total of 360 cortical areas (180 cortical areas per hemisphere; HCP360; cf., Figures 2 and 3), and 2) the network-based cortical parcellation estimated by intrinsic functional connectivity (Yeo et al., 2011), which delineated 17 networks on the cerebral cortex (cf., Figure 5C). To define ROI masks on individual subject’s brain, labels corresponding to brain areas or brain networks originally defined on the standard cortical surface (fsaverage) were converted to the cortical surfaces of individual subjects using FreeSurfer (version 6.0.1; Fischl, 2012). The converted labels were then reinterpolated to 2 × 2 × 2 mm voxels using flirt from FSL (version 5.0.9; Jenkinson et al., 2002). The numbers of voxels in individual ROI masks ranged from 92.4 to 2121.6 for the HCP360 ROIs (*n* = 360, median = 339.4, five subjects averaged) and from 2121.6 to 12969.8 for the Yeo's 17 network ROIs (*n* = 17, median = 8290.6, five subjects averaged).

In the analysis of individual cortical areas (Figures 5B and S1A), the visual cortex (VC), temporo-parietal junction (TPJ), inferior parietal lobule (IPL), precuneus (PC), superior temporal sulcus (STS), temporal cortex (TE), medial temporal cortex (MTC), insula, dorsolateral prefrontal cortex (DLPFC), dorsomedial prefrontal cortex (DMPFC), ventromedial prefrontal cortex (VMPFC), anterior cingular cortex (ACC), and orbitofrontal cortex (OFC) were defined based on the following sets of the HCP360 ROI labels on both left and right hemispheres: V1, V2, V3, V3A, V3B, V3CD, V4, V4t, V6, V6A, V7, V8, FST, IPS1, FFC, LO1, LO2, LO3, PH, PIT, MT, MST, VMV1, VMV2, VMV3, and VVC for VC; TPOJ1, TPOJ2, and TPOJ3 for TPJ; PFm, PGi, and PGs for IPL; 7m, v23ab, d23ab, 31pv, 31pd, 31a, and PCV for PC; STSda, STSdp, STSva, and STSvp for STS; TE1a, TE1m, TE1p, and TE2a for TE; EC, and H for MTC; AAIC, MI, PoI1, PoI2, FOP2, FOP3, Ig, and OP2-3 for insula; 9-46d, 46, a9-46v, and p9-46v for DLPFC; d32, 9m, and 10d for DMPFC; 10r, 10v, p32, and s32 for VMPFC; a24, a24pr, p24, a32pr, and p32pr for ACC; and OFC, and pOFC for OFC. Boundaries of these individual brain areas were drawn on the figures of cortical surfaces (e.g., Figures 2C and 4A). For visualization of subareas in VC, lines delineating V1, V2, V3, V4, and others were also drawn on the cortical surfaces. The numbers of voxels in individual cortical brain areas were 19721.2, 2021.2, 5031.4, 2557.0, 2483.8, 5028.8, 1386.4, 2423.8, 4141.6, 3177.4, 1876.4, 2773.0, and 2200.2 for the VC, TPJ, IPL, PC, STS, TE, MTC, insula, DLPFC, DMPFC, VMPFC, ACC, and OFC, respectively (five subjects averaged).

To define ROIs based on the levels of the principal gradient axes, we utilized the principal gradient maps provided from Margulies et al. (2016), which are also originally defined on the standard cortical surface (fsaverage). We first converted original principal gradient maps of the first and second gradient axes to the cortical surfaces of individual subjects using FreeSurfer (version 6.0.1; Fischl, 2012), and then reinterpolated to 2 × 2 × 2 mm voxels using flirt from FSL (version 5.0.9; Jenkinson et al., 2002). The resultant gradient maps registered to individual subjects' brains are shown in Figure 5F (Subject 1), and principal gradient values of individual voxels were used to generate results in Figure 5G and Figure S4E. For each subject, values of the first gradient axis assigned to individual voxels were also used to construct ROI masks that correspond to ten levels (bins) of the first axis, such that a roughly equal number of voxels were assigned to each level (cf., Figure 5H).

To define ROI masks for subcortical areas, including the thalamus, hippocampus, hypothalamus, pallidum, brainstem, caudate, putamen, nucleus accumbens (Brodmann area 34), amygdala, and cerebellum, we used anatomical masks defined by the AAL and the Talairach Daemon provided through the WFU PickAtlas (Maldjian et al., 2003; Lancaster et al, 1997; Lancaster et al., 2000). The anatomical masks, which were originally defined in the stereotaxic space, were transformed to the individual T1w anatomical images, using FreeSurfer (version 6.0.1; Fischl, 2012), and then reinterpolated to 2 × 2 × 2 mm voxels using flirt from FSL (version 5.0.9; Jenkinson et al., 2002). The numbers of voxels in individual subcortical areas were 2738.4, 2763.8, 96.0, 890.4, 5192.8, 2072.4, 2426.4, 596.6, 763.2, and 18712.0 for the thalamus, hippocampus, hypothalamus, pallidum, brainstem, caudate, putamen, nucleus accumbens, amygdala, and cerebellum, respectively (five subjects averaged).

Video stimulus labeling

Video stimuli were labeled by multiple types of scores, or features, including two types of emotion scores (34 emotion categories and 14 affective dimensions), 1000 visual object features, and 73 semantic features. Values of these labels were z-scored for each emotion/feature to remove baseline differences across emotions/features (unless otherwise stated).

Emotion scores. We used the human emotion ratings of 34 emotion categories and 14 affective dimensions, which were provided from the previous study (see Cowen and Keltner, 2017 for details). The ratings were collected using online experiments via Amazon Mechanical Turk (AMT). For the 34 emotion categories, subjects of the online experiments rated each video in terms of the degree to which it made them feel the 34 emotion categories (100-point scale). The reported scale was converted to 1, when raters scored higher than 0, such that a score for a video from an individual rater become a dichotomous yes/no response. For the 14 affective dimensions, another group of subjects rated each video in terms of its placement along 14 scales of affective dimensions (9-scale Likert scale). Each video was evaluated by multiple raters (9–17 raters), and the ratings obtained for individual emotions from multiple raters were averaged for each emotion to set one mean score for one video for one emotion.

Visual object features. For constructing visual object features, we used the Caffe implementation (Jia et al., 2014) of the VGG19 deep neural network (DNN) model (Simonyan and Zisserman, 2014), which was pre-trained to classify 1000 object categories (the pre-trained model is available from https://github.com/BVLC/caffe/wiki/Model-Zoo). The VGG19 model consisted of a total of sixteen convolutional layers and three fully connected layers. To compute outputs by the VGG19 model, all frames of videos were resized to 224 × 224 pixels and provided to the model. The outputs from the last fully connected layer (fc8, 1000 units, before softmax operation) were averaged across all frames within each video to construct a feature vector for a video.

Semantic features. We have also collected semantic ratings for the video stimuli according to 73 semantic contents associated with relatively concrete concepts in the video stimuli. The semantic features include objects, scenes, actions, and events (see below for the full list). The data collection was also conducted through online experiments via AMT with 12 raters for each video. Participants rated each video according to whether the video contains each semantic concept by dichotomous yes/no responses. The ratings from 12 raters were averaged to construct feature for individual concept for each video. The full list of the 73 semantic features is as follows: above water scenes, aquatic animals, art, automobiles, babies, birds, black people, blood, boats, bottles/cans, boys, buildings, cartoons, cats, celebrities, cities, clouds, couples, crowds of people, daytime scenes, dead bodies, dogs, elderly people, explosions, fast-moving objects, feces/urine/vomit, fire, flags, food, furniture, genitalia, girls, guns, gymnasiums, hands/feet, historical footage, hospitals, indoor scenes, injuries, insects, land animals, large animals, machines, men, mountains, naked people, nature, nighttime scenes, outdoor scenes, paper, paranormal creatures, people, planes, plants, politicians, reporters, roads, sexual activity, sharp objects, small animals, smoke, snow, soldiers, sports, stores, stunts, television, underwater scenes, vast landscapes, video games, weapons, white people, and women.

Regularized linear regression analysis

We used the L2 regularized linear regression (ridge regression) to predict stimulus labels from voxel activity patterns (decoding models) and to predict voxel activity from stimulus labels (encoding models). In the decoding analysis, voxels showing highest correlation coefficients with the target labels in the training data were provided to decoding models constructed for individual labels (with a maximum of 500 voxels). All models were evaluated using 6-fold cross-validation procedure (61 runs data were grouped into five sets of 10 runs and one set of 11 runs data). In each fold of the cross-validation, models for individual labels (decoding) and voxels (encoding) were trained with five sets of data, and the estimated models were used to predict stimulus labels (decoding) and voxel activities (encoding) for left out test set. This was repeated by rotating training-test assignments for 6 times to produce model predictions for all data samples. Then, model performance was evaluated by calculating Pearson correlation coefficients between true and predicted scores of individual labels in the decoding analysis and between measured and predicted brain activities of individual voxels in the encoding analysis. While results shown in this study are based on the 6-fold cross-validation procedure, we have confirmed that differences of the number of fold had little effect on the results.

Ridge regression uses a regularization parameter to constrain the magnitude of the weight coefficients. In decoding analysis, we individually estimated a regularization coefficient for each combination of stimulus labels, ROIs, and subjects (e.g., a single value for “fear” score predictions from V1 activities of Subject 1). In encoding analysis from each label set (category, dimension, visual object, and semantic), we estimated a single value of the regularization coefficient for all voxels in each subject. The regularization parameters were optimized based on the model performances obtained by 5-fold cross-validation procedure (inner loop) nested within training data for the outer loops of the 6-fold cross-validation. In each fold of the nested (or inner) cross-validation loop, models for individual labels (decoding) and voxels (encoding) were trained with four sets of the data (within training data) using each of 20 possible regularization coefficients (log spaced between 10 and 10000), and the estimated models were used to predict stimulus labels (decoding) and voxel activities (encoding) for left out test set (one out of five sets). This was repeated by rotating training-test assignments for 5 times to produce model predictions for all data samples within each fold of outer loops of the cross-validation. Then, model performance was evaluated by calculating Pearson correlation coefficients between true and predicted labels of individual labels in the decoding analysis and between measured and predicted brain activities of individual voxels in the encoding analysis. These procedures were also repeated for each outer loop of the cross-validation to estimate model performances for each fold of outer loops. Then, the regularization parameters producing the maximal model performances in the nested (inner) cross-validation were used for predictions of each fold of the outer cross-validation loops.

Construction of ensemble decoders

To construct a decoder that aggregates information represented in multiple brain regions (n = 370), we constructed an ensemble decoder for each emotion by averaging prediction values from multiple decoders (region-wise decoder), each of which was trained with brain activity patterns in each brain region. For each individual emotion, the brain regions used for the aggregations were selected based on the model performances evaluated in a nested cross-validated manner. For each subject and emotion, predictions from region-wise decoders that showed higher decoding accuracy than a threshold ($r > 0.095$, permutation test, $p < 0.01$, Bonferroni correction by the number of brain regions [n = 370]; see Transparent Methods: “Permutation tests” for details) were averaged to construct ensemble predictions. When no brain regions produced decoding accuracy higher than the threshold, the decoder showing the best accuracy was used as a substitute for the ensemble decoder (this was only the case for the “guilt” decoder of Subject 5).

Video identification analysis

Identification of emotional experience induced by individual video stimuli was performed via predictions from the decoding (cf., Figures 3A, B and C) and encoding (cf., Figure 4F) analyses. In the decoding and encoding analyses, scores of individual emotions or signal intensity of individual voxels were predicted from observed brain activity patterns or stimulus labels, respectively. Those predictions for individual emotions or voxels were concatenated to construct emotion score patterns or voxel activity patterns. The procedure yielded a total of 2181 predicted patterns corresponding to the presented video clips. Identification was performed in the pairwise manner, in which the video clip was identified between true and false candidates, using the predicted voxel activity pattern or emotion score pattern. The predicted pattern was compared with two candidate patterns: one for the true video and the other for a false video selected from the rest of 2180 videos. The video with a higher correlation coefficient was selected as the identified video. The analysis was repeated for all combinations of the 2181 videos. The accuracy for each video was evaluated by the ratio of correct identification.

Emotion identification analysis

To evaluate the inter-subject consistency of brain regions representing individual emotions, identification of emotions was performed via patterns of decoding accuracies obtained from multiple brain regions of different subjects (cf., Figure 2G). In the decoding analysis, we have performed decoding analysis of scores of the 34 emotion categories and 14 affective dimensions, and evaluated decoding accuracies of individual emotions from multiple brain regions (*n* = 370, including both cortical and subcortical regions). The decoding accuracies from multiple brain regions were concatenated to construct a pattern of decoding accuracies (number of elements = 370) as an emotion representation of one subject. For each pair of subjects, accuracy patterns from one subject (test subject) were compared with accuracy patterns from another subject (reference subject) using all combinations of 34 emotion categories or 14 affective dimensions. For each accuracy pattern of the test subject, the emotion whose accuracy pattern of the reference subject was most correlated with the accuracy pattern of the test subject was selected. This procedure was conducted for all combinations of five subjects (*n* = 20). The identification accuracy was evaluated with various candidate set sizes.

Slope estimates for performance comparisons

Comparisons of encoding accuracies from two models (e.g., the emotion category model and affective dimension model; Figures 4B and D) were performed based on the slope angles of best linear fit between two sets of model prediction accuracies of individual voxels. The best linear fit was estimated by Deming regression (Cornbleet and Gochman, 1979; or two-dimensional case of the total least square regression) to accounts for observation errors on both x- and y-axis (e.g., on the category model accuracy and dimension model accuracy; Figure 4B). The slope estimates were converted to angles by first calculating the arctangent of the slopes to obtain angles in radians, and then converting it to degrees. The calculated angles were further subtracted from 45 (degree) to obtain deviations from the parity (Figures 4D and S5D). Statistical significance of the slope estimates was computed based on standard errors of estimated slopes calculated by the jackknife method (two-tailed t-test). For the results of the pooled condition in Figures 4D and S5D, voxels within each brain region were aggregated from all five subjects to calculate slope estimates.

Dimensionality reduction analysis

We used Uniform Manifold Approximation and Projection (UMAP) algorithm (McInnes et al., 2018) to perform dimensionality reduction analysis. We applied the UMAP algorithm on emotion category scores (Figure 3D) and brain activity patterns (Figures 6A and B) to reduce the dimensionality of original data into two dimensions.

In the analysis with emotion category scores, we trained a mapping function from original 34 emotion category scores of 2181 video stimuli using correlation distance ($1 - r$) as the distance metric. The trained mapping function was used to project 34-dimensional category scores to two dimensions for both true (Figure 3D left) and decoded (Figure 3D right) emotion category scores.

In the analysis with brain activity patterns, we trained mapping functions based on correlation distances among brain activity patterns to 2181 video stimuli estimated from individual subjects and their average. Before applying the UMAP algorithm, voxels associated with emotions were selected based on the results of encoding analysis for individual subjects, in which voxels showing the higher accuracy from the category/dimension emotion models than the visual object and semantic models with significantly high accuracy by the category/dimension emotion models (cf., Figure S4D, including voxels in both cortical and subcortical regions) were selected. The activity patterns of selected voxels were used to construct matrices of correlation distance ($1 - r$) for individual subjects. The estimated distance matrices from individual subjects and their average (2181 × 2181 matrix) were used as inputs to the UMAP algorithm to construct two-dimensional maps of emotional experiences.

In the generated two-dimensional maps, each data sample was colored by a weighted interpolation of the unique colors assigned to 27 distinct emotion categories (Figures 3D, and 6A) or three representative affective dimensions, including valence, arousal, and dominance (Figure 6B).

Clustering analysis

We used k-means clustering algorithm to perform clustering analysis with brain activity patterns of individual subjects. The analysis was performed with activity patterns of voxels that were selected based on the results of encoding analyses (see Transparent Methods: “Dimensionality reduction analysis” for the selection criteria of emotion related voxels) using correlation distances ($1 - r$) as metric. The number of clusters (n = 27 in Figure 6) was determined based on the findings in the previous study (Cowen and Keltner, 2017; see Figures S6D and E for results with n = 15 and 50).

Representational similarity analysis

The representational similarity analysis (Kriegeskorte et al., 2008) was performed to evaluate the similarity of the representational similarity matrices (RSMs) constructed from brain activity patterns and score/feature patterns for each set of features. The RSM was calculated by Pearson correlation coefficients between patterns of voxel activities or scores/features corresponding to individual video stimuli (2181 × 2181 matrix). For calculating the representational similarity between two RSMs (e.g., one from brain activity pattern in a single ROI, and the other from emotion category scores), the off-diagonal elements (triangular part of a matrix) of the RSMs were vectorized and a Pearson correlation coefficient was calculated between those vectors from two RSMs. The analysis was performed using brain activity patterns within individual ROIs defined by the HCP360 parcellation (Glasser et al., 2016). The estimated representational similarities for individual brain regions were averaged across five subjects, and were projected on the cortical surface of Subject 1 (Figure S5).

QUANTIFICATION AND STATISTICAL ANALYSIS

Two-sided paired t-test was used to examine differences in encoding accuracy from two models (Figures 4C and 5H). ANOVA was used to examine interaction effects between encoding performances from the emotion and semantic models and the levels of principal gradients (Figures 5G and H), and to examine interaction effects between frequency and sorted clusters in the clustering analysis (Figure 6D).

Permutation tests

Statistical significance of correlation coefficients (e.g., encoding accuracy in Figure 4A) was computed by comparing estimated correlations (or accuracy) to the null distributions of correlations between two independent Gaussian random vectors of the same length (2181 elements, *n* = 100,000,000,000; Huth et al., 2016). Resulting *p*-values were corrected for multiple comparisons using the Bonferroni method.

The baseline entropy (cf., Figure 6E) was determined from null distributions of entropies (*n* = 100,000) calculated from random assignments of the same number of samples (*n* = 109) into clusters (27 clusters for Figure 6E; 15 and 50 clusters for Figures S6D and E). An entropy was calculated from a histogram by randomly assigning samples into bins. This procedure was repeated for 100,000 times to construct null distributions, and determined the baseline entropy (*p* = 0.01, Bonferroni correction by the number of emotions times the number of subjects).

DATA AND SOFTWARE AVAILABILITY

The experimental code and data that support the findings of this study are available from our repository (https://github.com/KamitaniLab/EmotionVideoNeuralRepresentation) and open data repositories (OpenNeuro: https://openneuro.org/datasets/ds002425; Mendeley Data: http://dx.doi.org/10.17632/jbk2r73mzh.1; figshare: https://doi.org/10.6084/m9.figshare.11988351.v1).

Supplemental References

- Avants, B.B., Epstein, C.L., Grossman, M., and Gee, J.C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12, 26–41.
- Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.
- Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., et al. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods 16, 111–116.
- Fischl, B. (2012). FreeSurfer. Neuroimage 62, 774–781.
- Gao, J.S., Huth, A.G., Lescroart, M.D., and Gallant, J.L. (2015). Pycortex: an interactive surface visualizer for fMRI. Front. Neuroinform. 9, 23.
- Greve, D.N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48, 63–72.
- Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841.
- Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., and Smith, S.M. (2012). FSL. Neuroimage 62, 782–790.
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. (2014). Caffe: Convolutional Architecture for Fast Feature Embedding. arXiv, arXiv:1408.5093. https://arxiv.org/abs/1408.5093.
- Cornbleet, P.J., and Gochman, N. (1979). Incorrect Least-Squares Regression Coefficients in Method-Comparison Analysis. Clin. Chem. 25, 432–438.
- Kragel, P.A., and LaBar, K.S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. Soc. Cogn. Affect. Neurosci. 10, 1437–1448.
- Kragel, P.A., Knodt, A.R., Hariri, A.R., and LaBar, K.S. (2016). Decoding Spontaneous Emotional States in the Human Brain. PLoS Biol. 14, e2000106–e2000119.
- Kragel, P.A., Reddan, M.C., LaBar, K.S., and Wager, T.D. (2019). Emotion schemas are embedded in the human visual system. Sci. Adv. 5, eaaw4358.
- Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P.A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron 60, 1126–1141.
- Lancaster, J.L., Summerin, J.L., Rainey, L., Freitas, C.S., and Fox, P.T. (1997). The Talairach Daemon, a database server for Talairach Atlas Labels. Neuroimage 5, S633.

- Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., and Fox, P.T. (2000). Automated Talairach atlas labels for functional brain mapping. Hum. Brain. Mapp. 10, 120–131.
- Maldjian, J.A., Laurienti, P.J., Burdette, J.B., and Kraft, R.A. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data Sets. Neuroimage 19, 1233–1239.