



## ORIGINAL ARTICLE

# Discrete Neural Signatures of Basic Emotions

Heini Saarimäki<sup>1,2</sup>, Athanasios Gotsopoulos<sup>1</sup>, Iiro P. Jääskeläinen<sup>1</sup>,  
 Jouko Lampinen<sup>1</sup>, Patrik Vuilleumier<sup>3,4</sup>, Riitta Hari<sup>1</sup>, Mikko Sams<sup>1</sup>, and  
 Lauri Nummenmaa<sup>1,5</sup>

<sup>1</sup>Department of Neuroscience and Biomedical Engineering and, <sup>2</sup>Advanced Magnetic Imaging (AMI) Centre, Aalto NeuroImaging, School of Science, Aalto University, FI-00076 Espoo, Finland, <sup>3</sup>Department of Neuroscience, University Medical Center and, <sup>4</sup>Department of Neurology, University Hospital, University of Geneva, 1211 Geneva, Switzerland, and <sup>5</sup>Turku PET Center and Department of Psychology, University of Turku, FI-20014 Turku, Finland

Address correspondence to Heini Saarimäki, Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, P.O. Box 12200, FI-00076 Aalto, Finland. Email: heini.saarimaki@aalto.fi

## Abstract

Categorical models of emotions posit neurally and physiologically distinct human basic emotions. We tested this assumption by using multivariate pattern analysis (MVPA) to classify brain activity patterns of 6 basic emotions (disgust, fear, happiness, sadness, anger, and surprise) in 3 experiments. Emotions were induced with short movies or mental imagery during functional magnetic resonance imaging. MVPA accurately classified emotions induced by both methods, and the classification generalized from one induction condition to another and across individuals. Brain regions contributing most to the classification accuracy included medial and inferior lateral prefrontal cortices, frontal pole, precentral and postcentral gyri, precuneus, and posterior cingulate cortex. Thus, specific neural signatures across these regions hold representations of different emotional states in multimodal fashion, independently of how the emotions are induced. Similarity of subjective experiences between emotions was associated with similarity of neural patterns for the same emotions, suggesting a direct link between activity in these brain regions and the subjective emotional experience.

**Key words:** emotion, fMRI, MVPA, pattern classification

## Introduction

Emotional systems comprise both neural and bodily states that provide immediate means for protection of the individual and that maximize adaptation to survival-salient events (Damasio 1998; Tsuchiya and Adolphs 2007). Categorical emotion models (Panksepp 1982; Ekman 1992) argue that evolution has shaped a limited set of basic emotions (anger, fear, disgust, happiness, sadness, and surprise) with distinct neural and physiological substrates to support different survival functions. These basic emotions are also characterized by distinctive feelings and culturally universal expressions (Tracy and Randles 2011). However, meta-analyses of functional magnetic resonance imaging (fMRI) studies using univariate analysis methods have failed to

establish discrete neural correlates for different basic emotions. Instead, a set of cortical midline and frontal regions, including medial prefrontal cortex (mPFC) and posterior cingulate cortex (PCC), are activated similarly during all basic emotional states (Phan et al. 2002; Murphy et al. 2003; Vytal and Hamann 2010). Moreover, some regions traditionally associated with specific emotions, such as amygdala for fear and insula for disgust (Calder et al. 2001), are in fact activated across a wide range of emotions (Sander et al. 2003). Thus, evidence for discrete brain substrates for different basic emotions has remained elusive (Kober et al. 2008; Lindquist and Barrett 2012).

It is nevertheless possible that differential activation patterns for each basic emotion exist within the cortical midline but remain uncovered with the prevailing univariate analyses. We

therefore adopted multivariate statistical pattern-recognition approach since such studies have proved successful in decoding sensory emotional signals in visual (Peelen et al. 2010; Said et al. 2010; Baucom et al. 2012) and auditory (Ethofer et al. 2009; Kotz et al. 2012) domains and across stimulus types (Peelen et al. 2010; Chikazoe et al. 2014; Skerry and Saxe 2014). These studies have focused on the perception of the emotion from emotional stimulus. Even fewer studies have aimed to classify the emotional experience or feeling, by using different emotional scenarios in the absence of direct sensory stimulation (Sitaram et al. 2011; Kassam et al. 2013). However, since these studies have focused on few a priori selected brain areas and conducted only pairwise comparisons between emotions, it still remains an open question to which extent all basic emotions would have distinct neural signatures, and which brain regions these signatures possibly include (for a review, see Krangel and LaBar 2014).

Here, we aimed at classifying the emotions induced in the participants, rather than discriminating the neural signatures of emotion-eliciting sensory stimuli, to reveal whether different basic emotions have discrete neural signatures. Emotions were induced in 3 experiments by affective movies or emotional imagery while participants' brain activity was measured with fMRI, allowing us to compare such signatures for both internally and externally triggered emotions. Using statistical pattern-recognition analysis of whole-brain as well as regional fMRI data, we show that all 6 basic emotions have distinguishable but spatially distributed neural signatures in the human brain. These emotion-specific neural signatures generalize across different emotion-eliciting conditions and also across individuals. Spatially overlapping but distinct local activation patterns for different emotions are thus a key property of the organization of emotions in the brain.

## Materials and Methods

### Participants

We conducted altogether 3 experiments where we induced emotions with affective movies or emotional imagery. These methods are widely used and reliable ways for inducing emotions in the laboratory setting (see reviews in Coan and Allen 2007). Twenty-one volunteers (12 males, ages 19–33 years, mean age 24.9 years) participated in the Movie Experiment, 14 female volunteers (age 19–30 years, mean age 23.6 years) in the Imagery Experiment, and 13 female volunteers (age 21–29 years, mean age 25.4 years) in the combined Movie–Imagery Experiment. All participants were healthy with normal or corrected-to-normal vision and gave written informed consent. The studies were run in accordance with the guidelines of the Declaration of Helsinki, and the Institutional Review Board of Aalto University had approved the study protocol.

### Design for the Movie Experiment

Fifty 10-s movie clips were chosen from a video database validated to reliably evoke basic emotions (Tettamanti et al. 2012). This database has been designed so that each video reflects prototypical and specific antecedents of the targeted emotion (e.g., Frijda 1986) and has been validated to elicit the targeted emotion with high mean accuracy (88%). Clips were chosen from 5 emotion categories (10 clips per category): disgust, fear, happiness, sadness, and neutral. Anger and surprise were not included in the study given the difficulties associated with eliciting these emotions with movies (Hewig et al. 2005). The clips were

randomly divided into 2 sets with 5 movies from each category in both sets.

During fMRI, both sets of movie clips were presented twice, thus resulting in altogether 4 runs, each of them lasting for 12 min 50 s. Each clip was preceded by a 5-s fixation cross and followed by a 15-s washout period. The clips were presented in a random order within each run. Movies were presented without sound to avoid attentional and linguistic confounds, as most movies contained English speech and the participants were native Finnish speakers. The participants were instructed to view the movies similarly as they would watch TV. No active task was required during fMRI scanning. In this and other experiments, ratings of emotional qualities of the stimuli were acquired post-experiment rather than during fMRI, as on-line reporting task is known to influence neural responses to emotional stimulation (Hutcherson et al. 2005).

The stimuli were delivered using Presentation software (Neurobehavioral Systems, Inc., Albany, CA, USA). They were back-projected on a semitransparent screen using a 3-micromirror data projector (Christie X3, Christie Digital Systems Ltd, Mönchengladbach, Germany) and from there via a mirror to the participant. After the scanning, the participants viewed the movie clips again and chose the emotion (disgust, fear, happiness, sadness, neutral, anger, surprise) that best described their feelings during each movie, and they rated the intensity (1–9) of the elicited emotion.

### Design for the Imagery Experiment

In a pilot study, we chose 36 Finnish emotion words representing 6 emotion categories: anger, disgust, fear, happiness, sadness, and surprise. Multiple synonyms for each emotion were chosen to generalize classification results to actual emotional states, rather than to single lexical units. For the pilot, 60 words were selected based on previous studies on Finnish emotion lexicon (Toivonen et al. 2012). Fifteen volunteers rated the similarity between all the 1770 possible pairs created with these emotion words using a scale ranging from 0 (completely dissimilar) to 5 (exactly similar). The mean similarity matrix across emotion concepts was then used to create a network of the emotion concepts using Gephi 0.8.2 (Bastian et al. 2009). We then applied a cluster analysis using the Louvain community detection algorithm (Blondel et al. 2008) to select 36 words that formed the most distinctive 6 categories corresponding to the 6 basic emotions (see also Supplementary Table 1): anger (furious, displeased, fierce, angry, cranky, annoyed), fear (restless, nervous, anxious, frightened, frantic, afraid), happiness (joyful, happy, merry, cheerful, delighted, pleased), sadness (sad, unhappy, sorrowful, heavy-hearted, depressed, gloomy), surprise (amazed, astonished, surprised, wondering, bemused, confused), and disgust (repelled, nauseous, bilious, disgusted, yucky, sickening).

The selected 36 emotion words were used in the subsequent fMRI experiment. One week prior to scanning, the participants were given a list of these 36 words and were asked to devise, write down, and practice their own method to elicit each emotion in the list. Sample methods of emotion elicitation (such as imagining a past event, thinking about a corresponding movie scene, or recreating the bodily state associated with the emotion) were provided but participants were free to choose whatever method they considered best for each emotion. Participants were asked to practice the imagery task at home for at least 1 h prior to the fMRI experiment.

During the fMRI scanning, each emotion word was presented once in each of the six 10-min runs. Each trial begun with a

fixation cross shown for 0.5 s, followed by the presentation of the word for 2 s, and an imagery period of 7 s. Participants were instructed to imagine the emotional state described by the emotion word they saw and to continue imagery until the subsequent inter-trial interval, which lasted randomly between 6.5 and 7.5 s. After this, the next trial was initiated. The visual stimuli were delivered as in the Movie Experiment.

After the scanning, the participants rated the pairwise similarity between all the 630 word pairs from the 36 emotion labels used in the fMRI experiment, using a scale ranging from 0 (completely dissimilar) to 5 (exactly similar). Participants also rated how difficult (1) versus easy (9) it was to induce each emotion with imagery, as well as the intensity (1–9) of the elicited emotion. Based on the average similarity ratings across participants, we calculated a mean similarity matrix between all word pairs and repeated a cluster analysis using the Louvain community detection algorithm similarly as in the pilot study to identify the most prominent emotion categories.

### Design for the Movie–Imagery Experiment

To test whether neural signatures of emotions are consistent across means by which emotions are induced, we ran a control experiment with both movie and imagery conditions (Movie–Imagery Experiment). The stimuli were selected from the same set as used in the 2 previous experiments. To make the designs of the movie and imagery conditions comparable, we chose the 4 overlapping emotion categories (disgust, fear, happiness, sadness) and selected 6 movie and 6 word stimuli per category.

Stimulus presentation in the movie condition was similar to the Movie Experiment. To match the number of data points used for classification in both conditions, we increased imagery duration slightly: each word was presented for 2 s, followed by 13 s of imagery, and another 7 s of washout. In both conditions, all the 6 exemplars of movies or words per each emotion category were presented all in one run. Altogether there were 6 runs per participant: 3 runs with movies and 3 runs with mental imagery.

### MRI Data Acquisition and Preprocessing

MRI data were collected on a 3T Siemens Magnetom Skyra scanner at the Advanced Magnetic Imaging Centre, Aalto Neuroimaging, Aalto University, using a 20-channel Siemens volume coil. Whole-brain functional scans were collected using a whole-brain  $T_2^*$ -weighted EPI sequence with the following parameters: 33 axial slices, TR = 1.7 s, TE = 24 ms, flip angle = 70°, voxel size =  $3.1 \times 3.1 \times 4.0 \text{ mm}^3$ , matrix size =  $64 \times 64 \times 33$ , FOV  $256 \times 256 \text{ mm}^2$ . A custom-modified bipolar water-excitation radio-frequency pulse was used to avoid signal from fat. High-resolution anatomical images with isotropic  $1 \times 1 \times 1 \text{ mm}^3$  voxel size were collected using a  $T_1$ -weighted MP-RAGE sequence.

Data were preprocessed using FSL 5.0 (Smith et al. 2004; Woolrich et al. 2009; Jenkinson et al. 2012). Motion was corrected using MCFLIRT (Jenkinson and Smith 2001; Jenkinson et al. 2002) and nonbrain matter was removed using BET (Smith 2002). High-pass temporal filtering was applied using Gaussian-weighted least-squares straight line fitting with sigma of 55 volumes. Participant-wise gray matter masks were generated by segmenting the  $T_1$ -weighted images into gray and white matter, plus cerebrospinal fluid, using the FAST segmentation tool (Zhang et al. 2001). The gray matter maps were subsequently transformed to  $64 \times 64 \times 33$  space to match the EPI data and thresholded using an intensity threshold  $>0.5$ . On average, the gray matter mask included 17 110 voxels.

For across-participants classification, the functional data were registered to 2-mm Montreal Neurological Institute (MNI) 152 standard space template using FLIRT (Jenkinson and Smith 2001; Jenkinson et al. 2002). The brain-extracted  $T_1$ -weighted images were first normalized to the MNI space and the normalization parameters were subsequently applied to the EPI images. All registrations were performed using 9 degrees of freedom. Since the across-participants classification was performed in standard space unlike the within-participant classification (performed in native space), we generated an average gray matter mask in MNI space. The gray matter mask was generated by segmenting, using FAST, the average normalized  $T_1$ -weighted image into gray and white matter and cerebrospinal fluid. The gray matter maps were subsequently thresholded using intensity threshold  $>0.5$  to create an average MNI mask that included 96 075 voxels.

### Multivoxel Pattern Analysis Within Participants

The classification of emotion categories within participants was performed with the Princeton multivoxel pattern analysis (MVPA) toolbox (<http://code.google.com/p/princeton-mvpa-toolbox/>, last accessed on April 21, 2015) in Matlab (2012b) using each participant's data in native space. We used the whole-brain data since recent studies have shown that emotional processing relies on large-scale cortical and subcortical circuits, rather than on isolated regions (Kober et al. 2008; Vytal and Hamann 2010; Nummenmaa, Saarimäki et al. 2014). Voxels outside gray matter were masked out and the functional data were temporally normalized to a mean of zero and unit variance in each voxel by subtracting the mean response across all categories. Feature selection was performed using ANOVA to select voxels whose activation was modulated across different emotion conditions ( $P < 0.05$ ). The feature selection preserved on average 31% of voxels in the Movie Experiment (mean 6320 voxels, SD 2100), and 38% of voxels in the Imagery Experiment (mean 5355 voxels, SD 1000). Finally, the hemodynamic lag was corrected by convolving the category regressors with the canonical double-gamma hemodynamic response function.

In the Movie Experiment, we used the 15-s washout period to train the classifier to select the correct category out of all 5 possible emotion categories (disgust, fear, happiness, sadness, neutral). This poststimulus period reflects emotional effects not directly related to the stimulus itself, and excludes sensory differences across the stimulus categories. Previous research shows that after emotion induction with movies, the emotion and corresponding brain activity can persist up until minutes if no other task/induction is introduced (Eryilmaz et al. 2011) thus confirming that this approach is technically feasible. In the Imagery Experiment, we used the 9-s period including the presentation of the emotion word and the active imagery period, and the classifier was trained to select the correct category out of possible 6 emotion categories (anger, disgust, fear, happiness, sadness, surprise).

A linear neural network classifier without hidden layers was used for classifying emotions in both experiments. The classifier starts with random weights from input  $i$  (voxels) to output  $j$  (categories). During training, the weights are adjusted for each given input with scaled conjugate gradient algorithm for optimization and mean squared error as an error function. During testing, each input is mapped to values from 0 to 1 for each output category using logistic functions. This corresponds to the confidence that the input belongs to a specific category. In the Movie and Imagery Experiments, the classifier was trained

using a leave-one-run-out procedure where training was performed with  $n - 1$  runs and testing was then applied to the remaining one run. Cross-validation was performed across all runs and the participant-wise classification accuracy was calculated as an average percentage of correct guesses across all the cross-validation runs. The chance-level performance percentage is derived as a ratio of 1 over the number of categories. Thus, the chance levels were 20% in the Movie Experiment and 16.6% in the Imagery Experiment.

To test whether classification accuracy exceeded chance level, we used permutation tests to simulate the probability distribution of the classification. Each permutation step included shuffling of category labels and re-running the classification, repeated 5000 times for each subject separately. The reported confidence limits of the chance-level accuracies correspond to the maximum over the permuted subject-wise limits (23% for  $P < 0.05$  in Movie Experiment, 19% for  $P < 0.05$  in Imagery Experiment).

To visualize the brain regions contributing most to the classifier's selection of each emotion category, voxel-wise importance values were calculated and plotted separately for each category. Importance values were calculated by defining importance  $imp = a \times w$ , where  $a$  is the activation of a voxel for a specific category and  $w$  is the trained weight from this voxel assigned to a specific category (Polyn et al. 2005). This method reveals which voxels are most important in driving the classifier's output for a specific category, and it highlights voxels that have concordant activation values and weights. Participant-wise importance maps were first calculated using the mean importance values over cross-validation runs and subsequently registered to MNI space. Then, mean importance maps were calculated across all participants for each emotion. These maps were plotted on a standard brain volume after selecting the highest 10 000 importance values (corresponding to ca. 1%). Clusters smaller than  $27 (3 \times 3 \times 3)$  voxels ( $216 \text{ mm}^3$ ) were excluded from visualizations. It should be noted that all voxels that passed the feature selection were taken into account in the classification and the importance maps simply highlight the most important clusters. Finally, to test the overlap of regions important for the classification in both experiments, we calculated the spatial correlation between the voxel importance values for the emotion categories included in both experiments (disgust, fear, happiness, sadness).

### Multivoxel Pattern Analysis Across Movie and Mental Imagery Conditions

For the combined Movie–Imagery Experiment, we trained a classifier using the 15-s washout period following the movies and the 15-s imagery periods (including the presentation of the emotion word) for the imagery part. The classifier was trained with either the movie data and tested with the imagery data, or vice versa for cross-validation, and it was trained to select the correct category out of 4 possible ones (disgust, fear, happiness, sadness). The chance level was 25% (by permuting, the significant threshold was 27.5% for  $P < 0.05$ ). Feature selection and other classifier details were the same as in the previous experiments, and feature selection preserved 20% of the voxels (mean 3100 voxels, SD 2500).

### Multivoxel Pattern Analysis Across Participants

To test whether representation of different emotions generalizes across participants, we ran whole-brain across-participants MVPAs. This analysis was performed with the same steps as the

within-participant classification but using the data registered to MNI space with 2-mm isotropic voxels. The feature selection preserved on average 19% of the voxels in Movie Experiment (mean 18 350 voxels, SD 1250), and 24% of the voxels in Imagery Experiment (mean 23 350 voxels, SD 1800). For each experiment, a linear classifier was trained using a leave-one-participant-out procedure where the training was performed with  $n - 1$  participants and the testing of the classifier with the remaining one participant. Cross-validation was then performed across all participants, and the classification accuracy was calculated as an average percentage of correct guesses across all the cross-validation runs.

### Region-of-Interest Analysis

We also applied a region-of-interest (ROI) analysis to test whether the blood oxygen level-dependent signal in any of our a priori-defined ROIs would allow a reliable classification of the emotional states when considered alone. Cortical regions showing consistent emotion-related activation in the literature were selected as candidate ROIs for coding discrete emotional content (Murphy et al. 2003; Kober et al. 2008; Vytal and Hamann 2010): orbitofrontal cortex (OFC; on average  $38 \text{ cm}^3$ ), frontal pole ( $186 \text{ cm}^3$ ), inferior frontal gyrus (IFG;  $32 \text{ cm}^3$ ), insula ( $27 \text{ cm}^3$ ), anterior cingulate cortex (ACC;  $31 \text{ cm}^3$ ), PCC ( $33 \text{ cm}^3$ ), frontal medial cortex ( $11 \text{ cm}^3$ ), precuneus ( $57 \text{ cm}^3$ ), paracingulate gyrus ( $31 \text{ cm}^3$ ), precentral gyrus ( $100 \text{ cm}^3$ ), supplementary motor cortex ( $17 \text{ cm}^3$ ), and postcentral gyrus ( $80 \text{ cm}^3$ ). The subcortical regions were amygdala (on average  $6 \text{ cm}^3$ ), nucleus accumbens ( $7 \text{ cm}^3$ ), hippocampus ( $10 \text{ cm}^3$ ), and thalamus ( $39 \text{ cm}^3$ ). Bilateral masks for these ROIs were first defined in MNI standard space using the Harvard–Oxford cortical and subcortical atlases (Desikan et al. 2006). These MNI masks were used in across-participants classification. For within-participant classification, the masks were transformed into native space using FLIRT in FSL. A classifier was then trained for each ROI separately in a similar fashion as in the whole-brain analyses. Accuracy was then averaged across the homologous areas.

### Comparison Between Behavioral and Neural Similarity of Different Emotions

To examine the correspondence between brain activation patterns and subjective feelings associated with different emotion categories, we extracted similarity matrices from fMRI data and behavioral rating of word pairs in the Imagery Experiment. To construct the neural similarity matrix, we trained a within-participant classifier to separate between brain responses to all 36 emotion words and computed the mean confusion matrix across the basic emotion categories for each participant. All other parameters remained as in the between-category classification described above. As an indicator for neural similarity, we then averaged these confusion matrices across participants and averaged the upper and lower triangles to make the matrix symmetrical and to estimate the mean confusion regardless of which category was the target and which was the guess. Likewise, the behavioral similarity matrix was computed on the basis of the average pairwise similarity ratings of emotion words across all participants. Next, we applied the Mantel test to examine the correlation between the 2 similarity matrices using an in-house algorithm (available at <http://becs.aalto.fi/~egleean/permutations.html>, last accessed on April 21, 2015). The probability distribution was obtained with permutation repeated for  $10^6$  times.

## Results

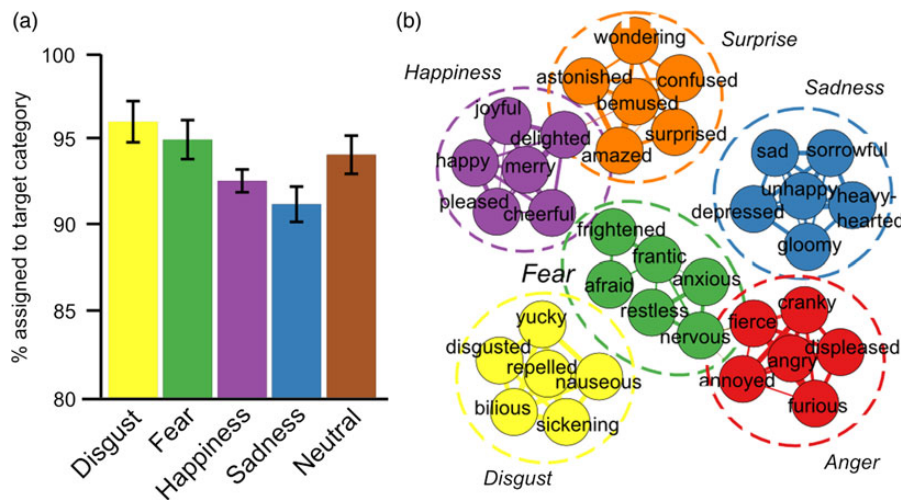
### Subjective Experience of Emotions

Behavioral ratings confirmed that the video clips used in the Movie experiment successfully induced the intended emotions in participants (Fig. 1a). Also, the intensity of the elicited emotions was high (disgust: mean 7.7, SD 0.7; fear: mean 7.7, SD 0.5; happiness: mean 6.0, SD 0.7; sadness: mean 6.2, SD 0.5) compared with the neutral state (mean 1.7, SD 0.6). For the Imagery Experiment, cluster analysis of the similarity ratings confirmed that the 36 emotion words constituted 6 discrete categories corresponding to the basic emotions, with only weak linkage between words belonging to different emotion clusters (Fig. 1b). Self-reports confirmed that participants found it easy to induce the emotions with imagery (mean 6.1, SD 2.2 on a scale ranging from 0 [difficult] to 9 [easy]). The intensity of the elicited emotions was high (anger: mean 5.8, SD 2.0; disgust: mean 6.2, SD 2.2; fear: mean 6.2, SD 2.0; happiness: mean 6.7, SD 1.4; sadness: mean 4.9, SD 1.9; surprise: mean 4.6, SD 2.1).

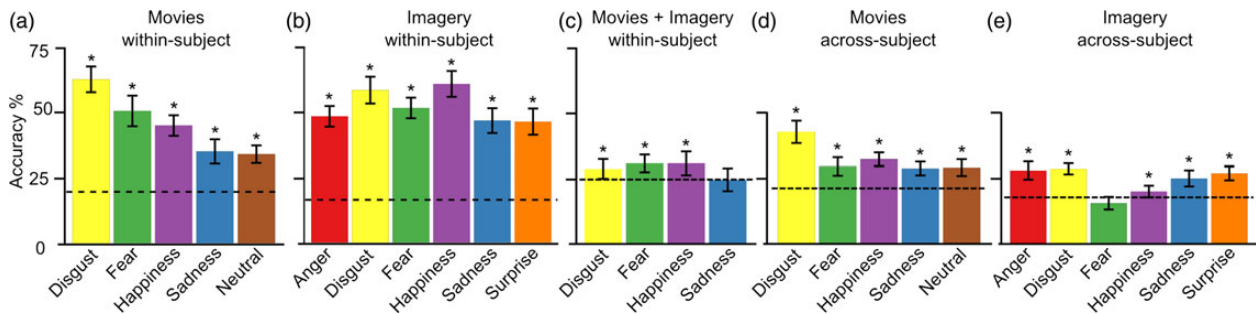
### Classification Within Participants

In the Movie Experiment, the mean within-participant classifier accuracy was 47% for distinguishing one emotion against all

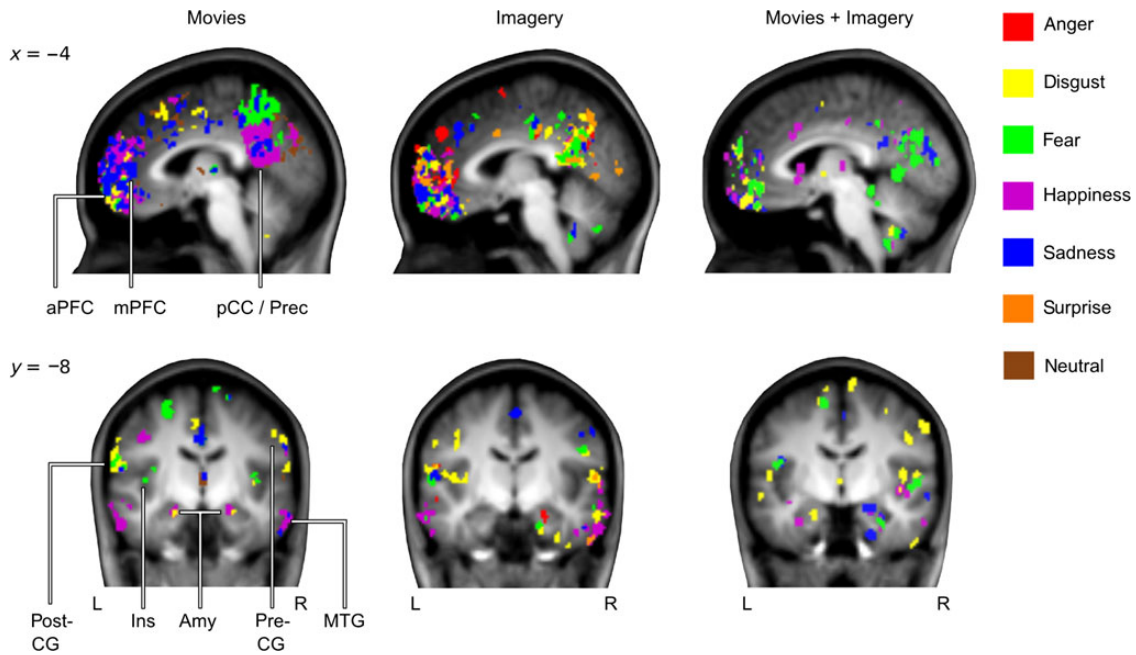
others (averaged across all categories) and the classifier was able to classify each of the 5 emotion categories statistically significantly above chance level (20%,  $P < 0.05$ ; Fig. 2a). Only in 3 out of 21 participants the average classifier accuracy remained at chance level. The left panel of Figure 3 shows the voxels with highest importance for the classification of each emotion in this experiment. The brain areas contributing to classification of all emotions predominated in frontal (dorsal, ventral, and rostral mPFC; ACC) and parietal (precuneus; PCC) midline structures. Other important regions were spread out over the cortical surface, especially in frontal lobe, including anterior prefrontal cortex (aPFC) and IFG; in motor areas including precentral gyrus; in somatosensory and visceral regions including postcentral gyrus, opercular cortices, and posterior insula; in temporal lobe including middle temporal gyrus (MTG) and temporal pole; in higher order visual regions including fusiform gyri and lateral occipital cortex (LOC); and in subcortical areas including amygdala and thalamus. Importantly, no regions were found to be important for the classification of a single emotion, although some areas were associated with a few but not all categories (e.g., insula contributed most to fear and disgust, less to other emotions; see Fig. 3). Notably, amygdala voxels contributed to the classification of happiness, fear, and disgust.



**Figure 1.** Behavioral results for Movie and Imagery Experiments. (a) Behavioral results in the Movie Experiment. Mean  $\pm$  SEM percentages of movie clips per emotion category during which the participants reported feeling the corresponding emotion. The clips were assigned to the predefined target category with 93.1% overall accuracy (chance level 20%). (b) Behavioral results in the Imagery Experiment. Mean network of basic emotion concepts based on the participants' behavioral similarity ratings. Link width denotes similarity between words.



**Figure 2.** Mean  $\pm$  SEM classification accuracy for each emotion category. (a–c) Within-participant classification. (d and e) Cross-participant classification. Dashed line represents the chance level [20% in the Movie Experiment (a and d), 16.7% in the Imagery Experiment (b and e), 25% in the Movie–Imagery Experiment (c)].



**Figure 3.** Brain regions with the largest importance for within-participant classification for each basic emotion. mPFC, medial prefrontal cortex; PCC, posterior cingulate cortex; Prec, precuneus; aPFC, anterior prefrontal cortex; LOC, lateral occipital cortex; post-CG, postcentral gyrus; pre-CG, precentral gyrus; Ins, insula; Amy, amygdala; MTG, middle temporal gyrus. Note: These importance maps are shown for visualization only. All inference is based on the classifier performance.

In the Imagery Experiment, the mean within-participant classifier accuracy was 55% and the classifier was able to classify each emotion category statistically significantly above chance level (16.7%) for all 14 participants ( $P < 0.05$ ; see Fig. 2b). The middle panel of Figure 3 shows the voxels with highest importance for the classification of each emotion in this experiment. As in the Movie Experiment, the regions contributing most consistently to the classification of all emotions predominated in midline structures of the frontal (dorsal, ventral, and rostral mPFC; ACC) and parietal (precuneus; PCC) lobes. Again, other regions supporting accurate classification covered widespread cortical areas, especially in frontal regions such as aPFC and IFG; in motor areas including precentral gyrus and supplementary motor cortex; in somatosensory regions including postcentral gyrus, opercular cortex, and posterior insula; in temporal lobe areas such as MTG and parahippocampal gyri; in higher order visual regions including LOC; and in subcortical regions including amygdala and thalamus. Again no regions were uniquely important for the classification of a single emotion.

### Classification Across Movie and Imagery Conditions

To test whether the neural signatures of emotions are independent of the emotion induction procedure, we conducted a combined Movie-Imagery experiment, with movie and imagery conditions corresponding to those of the previously described unimodal movie and imagery experiments but now presented to the same individuals and with the 4 emotion categories shared between the 2 experiments (disgust, fear, happiness, and sadness). We trained a classifier with data from one condition (either movie or imagery) and then tested it with data from the other condition. The mean within-participant classifier accuracy, averaged for movie-imagery and vice versa, was 29%. The classifier was able to classify all emotion categories except sadness statistically significantly above chance level (25%,  $P < 0.05$ ; see Fig. 2c).

We also trained and tested the classifier separately for the movie and imagery conditions to replicate the findings of the separate Movie and Imagery experiments with a single elicitation procedure. Again, the classification accuracy was high and significantly above chance-level (25%) for both movie (57%) and imagery (40%) conditions and for all emotions.

Figure 3 (right panel) shows the mean importance maps for each emotion in the cross-condition classification. As in the previous experiments, the regions contributing most consistently to the classification of all emotions were found in midline structures of the frontal (dorsal, ventral, and rostral mPFC; ACC) and parietal (PCC) lobes. Other significant voxels were also identified in frontal regions such as aPFC; motor areas including precentral gyrus and supplementary motor cortex; somatosensory regions including postcentral gyrus, opercular cortex, and posterior insula; in temporal lobe areas such as MTG and parahippocampal gyri; in higher order visual regions including LOC; and in subcortical regions including amygdala and thalamus. Again no regions were uniquely important for the classification of any single emotion.

### Classification Across Participants

To estimate whether the neural signatures of different emotions are consistent across individuals, we next carried out a similar classification analysis with data from the separate Movie and Imagery Experiments but now across participants by training the classifier with  $n - 1$  participants and testing it on the remaining participant. In both experiments, the across-participant classifier performed statistically significantly above chance level when collapsing across all emotion categories ( $P_s < 0.05$ ; 34% versus 20%; 23% versus 16.7%; for movies and imagery, respectively). When tested separately, classification accuracy was also above chance level for all emotion categories except for fear during the Imagery Experiment (Fig. 2d-e). As in the within-participant classification,

the most important regions for across-participant classification in both individual experiments included midline brain regions, such as mPFC, precuneus, and PCC (see [Supplementary Fig. 1](#)). In the Movie Experiment, additional regions included aPFC, insula, pre- and postcentral gyri, lingual gyrus, LOC, fusiform cortices, MTG, cerebellum, thalamus, and amygdala. In the Imagery Experiment, additional regions also included aPFC, insula, supplementary motor cortex, pre- and postcentral gyri, lingual gyrus, LOC, fusiform cortices, MTG, cerebellum, thalamus, and amygdala.

### Region-of-Interest Analyses

In the Movie and Imagery Experiments, the classification accuracy was above chance level for all ROIs except for the nucleus accumbens ([Supplementary Fig. 2](#)). In the Movie Experiment, the best classification accuracies across emotions were reached in frontal pole (36%), precuneus (36%), postcentral gyrus (34%), precentral gyrus (32%), and IFG (30%). In the Imagery Experiment, the highest classification accuracies were found in frontal pole (44%), precentral gyrus (34%), IFG (31%), postcentral gyrus (30%), and OFC (30%). In the Movie–Imagery Experiment, the average classification accuracy was above chance level only in frontal pole (30%) and medial frontal cortex (28%). Importantly, in all experiments the whole-brain classifiers performed significantly better (47% in the Movie Experiment, 55% in the Imagery Experiment, 29% in the Movie–Imagery Experiment) than any of the ROI classifiers alone, except for the frontal pole in Movie–Imagery Experiment, for which the classifier was equally good as the whole-brain classifier.

Since our ROIs had different sizes, we investigated the relationship between ROI size and classification accuracy by calculating the correlation between these. Larger ROIs tended to have better classification accuracies ( $r = 0.76$  in Movie Experiment,  $r = 0.86$  in Imagery Experiment;  $r = 0.61$  in Movie–Imagery Experiment); however, some large ROIs such as thalamus exhibited poor classification accuracies.

### Comparison Between Behavioral and Neural Similarity Matrices

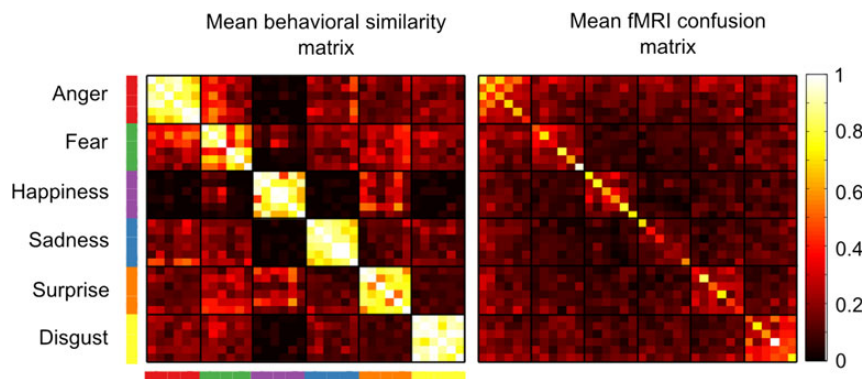
To construct the neural similarity matrix, we trained a within-participant classifier to separate between brain responses to all 36 emotion words and computed the mean confusion matrix across the basic emotion categories for each participant. The

classification accuracy was 14.1% against 2.8% chance level. The behavioral similarity matrix was extracted from similarity ratings, where participants rated the felt similarity between each pair of emotion words. The mean confusion matrix derived from the fMRI data and the mean behavioral rating similarity matrix were significantly correlated ( $r = 0.43$ ,  $P < 0.001$ ; [Fig. 4](#), [Supplementary Table 1](#)), suggesting that emotional states that are similar at the subjective level also share similar brain signatures. Participants rated emotion word pairs more similar when they belonged to the same rather than different basic emotion clusters ( $P < 0.05$ ; [Fig. 4](#), left panel). A similar predominance of confusions within a given emotion category was also observed for the fMRI classifier ([Fig. 4](#), right panel) but this pattern was generally weaker and less systematic than that found for the behavioral ratings.

### Discussion

Our results reveal that basic emotions are supported by discrete neural signatures within several brain areas, as evidenced by the high classification accuracy of emotions from hemodynamic brain signals. These emotion-specific neural signatures generalized across participants and across emotion-eliciting conditions in different modalities. Instead of engaging isolated brain regions, all basic emotions were associated with specific activation patterns within a distributed network of cortical and subcortical areas. The most consistent differential patterns were focused on the cortical midline structures and sensorimotor regions, but also extended to areas traditionally associated with emotion processing such as the insula or amygdala.

The distributed emotion-specific activation patterns may provide maps of internal states that correspond to specific subjectively experienced, discrete emotions ([Damasio and Carvalho 2013](#)). Even though our data confirm that at least the basic emotions have discrete neural bases, similar brain regions in cortical midline areas are engaged across different emotions: we found no direct one-to-one correspondence between a specific emotion and a specific brain site ([Murphy et al. 2003](#); [Kober et al. 2008](#); [Lindquist et al. 2012](#)). Critically, confusions between different—albeit closely related—emotion concepts, such as “frightened” and “restless,” were more pronounced at the subjective than neural level, suggesting that the consciously accessible feeling states are biased to be automatically interpreted as basic emotion categories, even though the underlying neural signatures remain further apart (compare behavioral similarity and neural confusion matrices in [Fig. 4](#)).



**Figure 4.** Behavioral similarity and fMRI confusion matrix from the Imagery Experiment. Left: Behavioral similarity matrix based on average ratings of experienced similarity between each pair of emotion words, ranging from 0 = no similarity to 1 = maximum similarity. Right: fMRI confusion matrix from word-by-word within-participant classification. Correct categories on the x-axis, classifier guesses on the y-axis.

## Distributed Representation of Emotions in the Brain

Results from the 3 experiments with externally and internally triggered emotions were concordant, with comparable classification accuracies for individual emotions and similar anatomical layout of importance maps focused on anterior and posterior midline regions. Together with accurate classification of emotions across different induction conditions (visual vs. imagery), these findings suggest that the involved cortical regions code emotions in a modality-independent fashion, and generalization of the neural signatures of different emotions across participants further points toward biological rather than experience-dependent brain basis of emotions. Importantly, classification accuracy was high even though the emotions induced in the participants were not necessarily completely discrete (Figs 1 and 4); it is thus likely that even more discrete neural signatures could be revealed with techniques allowing induction of purely discrete emotions.

None of the ROIs alone reached the accuracy of the whole-brain classification for any emotion, suggesting that the anatomically distributed activation patterns contain the most accurate neural signature of an individual's emotional state. This result accords with previous neuroimaging studies which typically show that joint activity from multiple regions discriminates best between different emotions (Baucom et al. 2012; Kotz et al. 2012; Kassam et al. 2013). Lesion studies have shown that deficits in different emotions (particularly of fear and disgust) are linked to damages of specific brain regions (amygdala, insula, respectively; cf. Calder et al. 2001). Nevertheless, our findings suggest that large-scale cortical and subcortical networks are crucially involved in representing all basic emotions in a distinctive, category-specific manner (see also Feinstein 2013). It must however be stressed that the present study focused only on the 6 emotions that are most widely accepted as basic. Future pattern-recognition studies should address whether also the "complex" or "social" emotions (Ekman and Cordaro 2011) would have discernible and discrete neural signatures.

## Discrete Neural Signatures for Basic Emotions

Recent data-driven meta-analysis of functional imaging studies (Kober et al. 2008) proposed a functional subdivision of emotional brain circuits into 6 groups, each responsible for processing different types of information (see also Meaux and Vuilleumier, 2015). These functional circuits were proposed to code for different components of emotions, such as attentional, motor, or mnemonic processes engaged during emotional episodes. However, previous experimental work has failed to establish different neural signatures for different emotions within these circuits. Nevertheless, the brain regions that supported the classification of emotions in the current study significantly overlap with the functional groups identified by Kober et al. (2008). In our study, the medial prefrontal and medial posterior regions (mPFC, precuneus, and PCC) contributed most significantly to classification between different basic emotions (Fig. 3). Thus, local activation patterns within these areas differ across emotions and thus presumably reflect distinct neuronal signatures for different emotions.

These midline regions are consistently activated during emotional processing in different sensory modalities (Phan et al. 2002; Kober et al. 2008; Lindquist et al. 2012) and contain emotion-specific patterns independent of the task or exact emotion categories used (Peelen et al. 2010; Chikazoe et al. 2014; Skerry and Saxe 2014; see also Kragel and LaBar 2014 for a review). The

mPFC and PCC receive inputs from insula which processes visceral information, from amygdala which codes the affective relevance of the stimulus, from medial temporal lobe areas involved in memory, and from thalamus and hypothalamus which govern arousal (Öngür and Price 2000; Kober et al. 2008; Etkin et al. 2011). Together, mPFC, precuneus, and PCC form the medial part of the default mode network (DMN), typically linked with self-referential processing (Amodio and Frith 2006; Northoff et al. 2006; Buckner and Carroll 2007). This anatomical architecture makes these midline regions a plausible candidate for integrating information about one's internal state (Klaser et al. 2011; Mar 2011) with representations from memory and personal relevance (Summerfield et al. 2009; D'Argembeau et al. 2010). The patterns of activity resulting from the binding of these various representations might constitute a core feature of an emotional state regardless of the particular emotion category, and possibly underlie the distinctive fingerprints of these states as identified by our MVPA analyses. Somatosensory and motor regions—including postcentral gyrus, posterior insula, and precentral gyrus—were also among the most important brain regions for discriminating between all emotions. This finding accords with previous work showing how different emotions elicit discernible patterns of somatic sensations (Nummenmaa, Glerean et al. 2014), and that primary somatosensory, motor and premotor cortices are reliably engaged during emotion perception (De Gelder et al. 2004; Nummenmaa et al. 2008, 2012; Pichon et al. 2008). Moreover, damage to somatosensory cortices (Adolphs et al. 2000) or their inactivation by transcranial magnetic stimulation (Pourtois et al. 2004) can cause significant deficits in the recognition of emotions. Similarly, posterior insula mediates the interoceptive awareness of one's own bodily functions (Critchley et al. 2004) and its damage may impair various components of emotion processing, including gustatory information (Calder et al. 2001) or aversive/risk information (Naqvi et al. 2007). Precentral gyrus containing the primary motor cortex is also consistently activated during emotional experience and emotion perception (De Gelder et al. 2004; Hajcak et al. 2007), and it likely plays an important role in motor preparation processes related to emotion and action tendencies (Frijda 1986; Mazzola et al. 2013).

Limbic regions, including amygdala, hippocampus, and thalamus, form an important part of the emotion network, yet our ROI analysis revealed poorer classification accuracy in limbic versus cortical components of the emotion network. Furthermore, none of the limbic ROIs was able to separate between all emotion categories. It is possible that this finding merely reflects the positive association between classification accuracy and ROI size, as the limbic ROIs were, on average, smaller than their cortical counterparts. However, follow-up analysis established that such an association is present only for the cortical ( $M_r = 0.75$ ,  $P < 0.05$ ) but not for the limbic ( $M_r = -0.05$ , ns) ROIs. Thus, mere ROI size unlikely accounts the poorer classification accuracy in the limbic ROIs, particularly as some of these—such as thalamus—were indeed relatively large. One possibility is that the limbic circuit contributes to shaping emotional states jointly with the cortical regions. The limbic regions likely govern elementary functions related to arousal, saliency, and relevance processing, which tend to be shared by all different emotions (Adolphs 2010; Damasio and Carvalho 2013; Kragel and LaBar 2014). Activity in these subcortical regions may then contribute to the generation of discrete emotional states via feed-forward connections to the frontal cortex, but the latter may also shape emotion responses through feedback interactions with limbic regions.



## Subjective Emotional Experience Is Linked with Neural Activity

Humans are usually aware of their current emotional state, which may help to fine-tune the behavior adaptively to better match to the challenges of the environment (Damasio 1996). Prior studies have linked the medial frontal cortex with the subjective feelings of emotions (Barrett et al. 2007; Etkin et al. 2011; Herbert et al. 2011; Satpute et al. 2012) and self-awareness (Lane et al. 1998; Northoff et al. 2006), and the present results establish a direct link between neural activity in these regions and emotional experience: The more similar neural signatures two emotions have, the more similar are the corresponding subjective feeling states.

Damasio et al. (2000) suggested that emotion-dependent neural patterns across regions could explain why each emotion feels subjectively different. Here, we provide direct support for this proposal by showing that distinct brain activity patterns in somatosensory regions and insula, probably code subjective feeling of bodily sensations that contribute to the generation of a distinct physiological mapping for each emotion. We propose that the joint activation of these different components is integrated in the mPFC and precuneus/PCC where distributed responses arising in the downstream brain regions are ultimately connected with the context and personal goals, presumably resulting in distinctive neural signatures that reflect the subjective experience of a specific, discrete emotion.

## Conclusions

Basic emotions have a discrete neural basis, and the neural signatures of different emotions are consistent across individuals. Basic emotions are encoded in discrete activation patterns within a widespread network of brain regions, rather than in emotion-specific brain regions or systems. We propose that the momentary subjective emotional state is the result of simultaneous activation of multiple cortico-subcortical systems, including regions processing somatosensory, motor, and self-relevant information, but also perceptual, language, memory, and executive control functions. The activation of these subcircuits is integrated in the midline frontal and parietal regions, linking emotion-driven neural and physiological changes to self-awareness.

## Supplementary Material

Supplementary Material can be found at: <http://www.cercor.oxfordjournals.org>

## Funding

This work was supported by the aivoAALTO project of the Aalto University, Academy of Finland (#265917 to L.N., #218072 to R.H., and #138145 and #276643 to I.P.J.), ERC Starting Grant (#313000 to L.N.); ERC Advanced Grant (#232946 to R.H.); and the Swiss National Science Foundation National Center of Competence in Research for Affective Sciences (#51NF40-104897 to P.V.).

## Notes

We thank Marita Kattelus for her help with the data acquisition. *Conflict of Interest:* None declared.

## References

- Adolphs R. 2010. What does the amygdala contribute to social cognition? *Ann N Y Acad Sci.* 1191:42–61.
- Adolphs R, Damasio H, Tranel D, Cooper G, Damasio AR. 2000. A role for somato sensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *J Neurosci.* 20:2683–2690.
- Amodio DM, Frith CD. 2006. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci.* 7:268–277.
- Barrett LF, Mesquita B, Ochsner KN, Gross JJ. 2007. The experience of emotion. *Ann Rev Psychol.* 58:373–403.
- Bastian M, Heymann S, Jacomy M. 2009. Gephi: an open source software for exploring and manipulating networks. *Int AAAI Conf Weblogs Soc Media.* 361–362.
- Baucom LB, Wedell DH, Wang J, Blitzer DN, Shinkareva SV. 2012. Decoding the neural representation of affective states. *Neuroimage.* 59:718–727.
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp.* 2008:P10008.
- Buckner RL, Carroll DC. 2007. Self-projection and the brain. *Trends Cogn Sci.* 11:49–57.
- Calder AJ, Lawrence AD, Young AW. 2001. Neuropsychology of fear and loathing. *Nat Rev Neurosci.* 2:342–363.
- Chikazoe J, Lee DH, Kriegeskorte N, Anderson AK. 2014. Population coding of affect across stimuli, modalities and individuals. *Nat Neurosci.* 17:1114–1122.
- Coan JA, Allen JJ. 2007. *Handbook of emotion elicitation and assessment.* New York: Oxford University Press.
- Critchley HD, Wiens S, Rotshtein P, Öhman A, Dolan RJ. 2004. Neural systems supporting interoceptive awareness. *Nat Neurosci.* 7:189–195.
- Damasio AR. 1996. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philos Trans R Soc Lond B Biol Sci.* 351:1413–1420.
- Damasio AR. 1998. Emotion in the perspective of an integrated nervous system. *Brain Res Rev.* 26:83–86.
- Damasio AR, Carvalho GB. 2013. The nature of feelings: evolutionary and neurobiological origins. *Nat Rev Neurosci.* 14:143–152.
- Damasio AR, Grabowski TJ, Bechara A, Damasio H, Ponto LL, Parvizi J, Hichwa RD. 2000. Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat Neurosci.* 3:1049–1056.
- D'Argembeau A, Stawarczyk D, Majerus S, Collette F, Van der Linden M, Feyers D, Maquet P, Salmon E. 2010. The neural basis of personal goal processing when envisioning future events. *J Cogn Neurosci.* 22:1701–1713.
- De Gelder B, Snyder J, Greve D, Gerard G, Hadjikhani N. 2004. Fear fosters flight: a mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proc Natl Acad Sci USA.* 101:16701–16706.
- Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, et al. 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage.* 31:968–980.
- Ekman P. 1992. An argument for basic emotions. *Cogn Emot.* 6:169–200.
- Ekman P, Cordaro D. 2011. What is meant by calling emotions basic. *Emot Rev.* 34:364–370.
- Eryilmaz H, Van De Ville D, Schwartz S, Vuilleumier P. 2011. Impact of transient emotions on functional connectivity during

- subsequent resting state: a wavelet correlation approach. *Neuroimage*. 54:2481–2491.
- Ethofer T, Van De Ville D, Scherer K, Vuilleumier P. 2009. Decoding of emotional information in voice-sensitive cortices. *Curr Biol*. 19:1028–1033.
- Etkin A, Egner T, Kalisch R. 2011. Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn Sci*. 15:85–93.
- Feinstein JS. 2013. Lesion studies of human emotion and feeling. *Curr Opin Neurobiol*. 23:304–309.
- Frijda NH. 1986. *The emotions*. Cambridge (UK): Cambridge University Press.
- Hajcak G, Molnar C, George MS, Bolger K, Koola J, Nahas Z. 2007. Emotion facilitates action: a transcranial magnetic stimulation study of motor cortex excitability during picture viewing. *Psychophysiology*. 44:91–97.
- Herbert C, Herbert BM, Pauli P. 2011. Emotional self-reference: brain structures involved in the processing of words describing one's own emotions. *Neuropsychologia*. 49:2947–2956.
- Hewig J, Hagemann D, Seifert J, Gollwitzer M, Naumann E, Bartussek D. 2005. A revised film set for the induction of basic emotions. *Cogn Emot*. 19:1095–1109.
- Hutcherson CA, Goldin PR, Ochsner KN, Gabrieli JD, Barrett LF, Gross JJ. 2005. Attention and emotion: does rating emotion alter neural responses to amusing and sad films? *Neuroimage*. 27:656–668.
- Jenkinson M, Bannister PR, Brady JM, Smith SM. 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*. 17:825–841.
- Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. 2012. FSL. *Neuroimage*. 62:782–790.
- Jenkinson M, Smith S. 2001. A global optimisation method for robust affine registration of brain images. *Med Image Anal*. 5:143–156.
- Kassam KS, Markey AR, Cherkassky VL, Loewenstein G, Just MA. 2013. Identifying emotions on the basis of neural activation. *PLoS One*. 8:e66032.
- Klasen M, Kenworthy CA, Mathiak KA, Kircher TT, Mathiak K. 2011. Supramodal representation of emotions. *J Neurosci*. 31:13635–13643.
- Kober H, Barrett LF, Joseph J, Bliss-Moreau E, Lindquist K, Wager TD. 2008. Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *Neuroimage*. 42:998–1031.
- Kotz SA, Kalberlah C, Bahlmann J, Friederici AD, Haynes JD. 2012. Predicting vocal emotion expressions from the human brain. *Hum Brain Mapp*. 34:1971–1981.
- Kragel PA, LaBar KS. 2014. Advancing emotion theory with multivariate pattern classification. *Emot Rev*. 6:1–15.
- Lane RD, Reiman EM, Axelrod B, Yun LS, Holmes A, Schwartz GE. 1998. Neural correlates of levels of emotional awareness: evidence of an interaction between emotion and attention in the anterior cingulate cortex. *J Cogn Neurosci*. 10:525–535.
- Lindquist KA, Barrett LF. 2012. A functional architecture of the human brain: emerging insights from the science of emotion. *Trends Cogn Sci*. 16:533–540.
- Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, Barrett LF. 2012. The brain basis of emotion: a meta-analytic review. *Behav Brain Sci*. 35:121–143.
- Mar RA. 2011. The neural bases of social cognition and story comprehension. *Annu Rev Psychol*. 62:103–134.
- Mazzola V, Vuilleumier P, Latorre V, Petito A, Gallese V, Popolizio T, Arciero G, Bondolfi G. 2013. Effects of emotional contexts on cerebello-thalamo-cortical activity during action observation. *PLoS One*. 8:e75912.
- Meaux E, Vuilleumier P. (forthcoming). Emotion perception and elicitation. In: Toga AW, editor. *Brain mapping: an encyclopedic reference*. Oxford (UK): Elsevier.
- Murphy FC, Nimmo-Smith I, Lawrence AD. 2003. Functional neuroanatomy of emotions: a meta-analysis. *Cogn Affect Behav Neurosci*. 3:207–233.
- Naqvi NH, Rudrauf D, Damasio H, Bechara A. 2007. Damage to the insula disrupts addiction to cigarette smoking. *Science*. 315:531–534.
- Northoff G, Heinzl A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J. 2006. Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage*. 31:440–457.
- Nummenmaa L, Glerean E, Hari R, Hietanen J. 2014. Bodily maps of emotions. *Proc Natl Acad Sci USA*. 111:646–651.
- Nummenmaa L, Glerean E, Viinikainen M, Jääskeläinen IP, Hari R, Sams M. 2012. Emotions promote social interaction by synchronizing brain activity across individuals. *Proc Natl Acad Sci USA*. 109:9599–9604.
- Nummenmaa L, Hirvonen J, Parkkola R, Hietanen JK. 2008. Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *Neuroimage*. 43:571–580.
- Nummenmaa L, Saarimäki H, Glerean E, Gotsopoulos A, Hari R, Sams M. 2014. Emotional speech synchronizes brains across listeners and engages large-scale dynamic brain networks. *Neuroimage*. 102:498–509.
- Öngür D, Price JL. 2000. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb Cortex*. 10:206–219.
- Panksepp J. 1982. Toward a general psychobiological theory of emotions. *Behav Brain Sci*. 5:407–422.
- Peelen MV, Atkinson AP, Vuilleumier P. 2010. Supramodal representations of perceived emotions in the human brain. *J Neurosci*. 30:10127–10134.
- Phan KL, Wager T, Taylor SF, Liberzon I. 2002. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*. 16:331–348.
- Pichon S, de Gelder B, Grezes J. 2008. Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Soc Neurosci*. 3:199–212.
- Polyn SM, Natu VS, Cohen JD, Norman KA. 2005. Category-specific cortical activity precedes retrieval during memory search. *Science*. 310:1963–1966.
- Pourtois G, Sander D, Andres M, Grandjean D, Reveret L, Olivier E, Vuilleumier P. 2004. Dissociable roles of the human somatosensory and superior temporal cortices for processing social face signals. *Eur J Neurosci*. 20:3507–3515.
- Said CP, Moore CD, Engell AD, Todorov A, Haxby JV. 2010. Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J Vis*. 10:11.
- Sander D, Grafman J, Zalla T. 2003. The human amygdala: an evolved system for relevance detection. *Rev Neurosci*. 14:303–316.
- Satpute AB, Shu J, Weber J, Roy M, Ochsner KN. 2012. The functional neural architecture of self-reports of affective experience. *Biol Psychiatry*. 73:631–638.
- Sitaram R, Lee S, Ruiz S, Rana M, Veit R, Birbaumer N. 2011. Real-time support vector classification and feedback of multiple emotional brain states. *Neuroimage*. 56:753–765.
- Skerry AE, Saxe R. 2014. A common neural code for perceived and inferred emotion. *J Neurosci*. 34:15997–16008.

- Smith SM. 2002. Fast robust automated brain extraction. *Hum Brain Mapp.* 17:143–155.
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, et al. 2004. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage.* 23:S208–S219.
- Summerfield JJ, Hassabis D, Maguire EA. 2009. Cortical midline involvement in autobiographical memory. *Neuroimage.* 44: 1188–1200.
- Tettamanti M, Rognoni E, Cafiero R, Costa T, Galati D, Perani D. 2012. Distinct pathways of neural coupling for different basic emotions. *Neuroimage.* 59:1804–1817.
- Toivonen R, Kivelä M, Saramäki J, Viinikainen M, Vanhatalo M, Sams M. 2012. Networks of emotion concepts. *PLoS One.* 7: e28883.
- Tracy JL, Randles D. 2011. Four models of basic emotions: a review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emot Rev.* 3:397–405.
- Tsuchiya N, Adolphs R. 2007. Emotion and consciousness. *Trends Cogn Sci.* 11:158–167.
- Vytal K, Hamann S. 2010. Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *J Cogn Neurosci.* 22:2864–2885.
- Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M, Smith SM. 2009. Bayesian analysis of neuroimaging data in FSL. *Neuroimage.* 45: S173–S186.
- Zhang Y, Brady M, Smith S. 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging.* 20:45–57.