doi:10.1093/scan/nsv032

SCAN (2015) 10, 1437–1448

Multivariate neural biomarkers of emotional states are categorically distinct

Philip A. Kragel and Kevin S. LaBar
Department of Psychology & Neuroscience and Center for Cognitive Neuroscience, Duke University, Durham, NC, USA

Understanding how emotions are represented neurally is a central aim of affective neuroscience. Despite decades of neuroimaging efforts addressing this question, it remains unclear whether emotions are represented as distinct entities, as predicted by categorical theories, or are constructed from a smaller set of underlying factors, as predicted by dimensional accounts. Here, we capitalize on multivariate statistical approaches and computational modeling to directly evaluate these theoretical perspectives. We elicited discrete emotional states using music and films during functional magnetic resonance imaging scanning. Distinct patterns of neural activation predicted the emotion category of stimuli and tracked subjective experience. Bayesian model comparison revealed that combining dimensional and categorical models of emotion best characterized the information content of activation patterns. Surprisingly, categorical and dimensional aspects of emotion experience captured unique and opposing sources of neural information. These results indicate that diverse emotional states are poorly differentiated by simple models of valence and arousal, and that activity within separable neural systems can be mapped to unique emotion categories.

Keywords: emotion; affect; functional magnetic resonance imaging; pattern classification; multi-voxel pattern analysis; Bayesian model comparison

INTRODUCTION

It has long been hypothesized that emotions involve coordinated physiological, psychological and behavioral changes that support survival (Plutchik and Kellerman, 1980). Emotion directs attention, modulates memory, biases perception and guides decision making. In humans, these changes are often concomitant with emotional feelings that inherently reflect their utility in maintaining homeostasis (Damasio and Carvalho, 2013) and are shaped by both biological and social influences (Barrett, 2012). Accordingly, emotional states are multifaceted phenomena that are supported through the integration of information pertaining to the environment, homeostatic state and relevant adjustments in behavior.

There is contention regarding how emotions are represented within the brain: either as discrete categories or points in a continuous dimensional space (Barrett and Wager, 2006; Hamann, 2012; Lindquist *et al.*, 2012). Although both types of models posit that emotional states result from processes mediated by the central nervous system, they characterize their emergence via different mechanisms. Dimensional models propose that emotional states differ fundamentally along a small number of factors, e.g. pleasantness and activation (Russell, 1980; Feldman Barrett and Russell, 1999), although alternative interpretations and more complex models exist (Watson and Tellegen, 1985; Mehrabian, 1996; Fontaine *et al.*, 2007). Differences in these core dimensions, along with past experience and contextual factors, are thought to contribute to the construction of distinct emotional states (Russell, 2003; Barrett, 2006; Lindquist, 2013). Functional magnetic resonance imaging (fMRI) studies have identified putative neural correlates of some of these dimensions, implicating medial prefrontal structures in the processing of valence and the amygdala in mediating arousal or emotional salience (Hamann *et al.*, 2002; Anderson *et al.*, 2003; Winston *et al.*, 2005; Lewis *et al.*, 2007; Wilson-Mendenhall *et al.*, 2013). Categorical emotion models (Izard, 1971; Panksepp, 1982; Johnson-Laird and Oatley, 1992; Ekman and Cordaro, 2011) alternatively hypothesize that a finite number of independent emotions, often thought to derive from survival-related functions, are sufficient in characterizing the structure of emotion. Meta-analyses suggest discrete emotional experiences can be discriminated on the basis of neural activation in humans (Murphy *et al.*, 2003; Vytal and Hamann, 2010), although neural substrates specific to any given emotion remain elusive (Lindquist *et al.*, 2012).

Given the lack of consensus from univariate studies regarding the representation of emotion in the brain, multivariate methods have been adopted in attempts to identify more specific mappings between patterns of neural activation and a variety of affective states (for a review, see Kragel and LaBar, 2014), Brain states (measured with fMRI) have been classified along both dimensional (e.g. Rolls *et al.*, 2009; Baucom *et al.*, 2012) and categorical (e.g. Sitaram *et al.*, 2011; Kassam *et al.*, 2013) models of emotion. Although such research has demonstrated the capacity of multivariate methods to isolate patterns of neural activity associated with different affective states, several factors limit the conclusions that can be drawn from prior research. As prior studies have focused on either categorical or a dimensional framework for classification, evidence favoring either model is difficult to ascertain. Most studies only utilize a single stimulus modality (e.g. visual images, mental imagery, facial expressions or vocalizations), which limits the generalizability of results. Assessment of classification models in the literature thus far tends to focus on predictive capacity and fails to characterize the localization and spatial characteristics of emotion-predictive patterns (i.e. does each emotion have a unique substrate? Are they organized modularly or are they more widely distributed?). Further, studies commonly sample a limited number of emotions, which fail to adequately sample the dimensions of valence and arousal. Finally, and perhaps most importantly, no prior fMRI pattern classification research has examined on-line measures of subjective experience. This is a critical issue, as it is essential to determine whether specific emotions were elicited by the stimuli presented, and whether the information decoded by the pattern classifiers contributes to the experience of emotions as distinct entities or as points along broader affective dimensions.

Received 30 October 2014; Revised 13 February 2015; Accepted 19 March 2015
Advance Access publication 25 March 2015

The authors thank Bruce Friedman and Ivan Nyklíček for providing copies of stimulus materials, and Sean Polyn for assistance regarding data analysis. They declare no competing financial interests. P.A.K. was supported by the E. Bayard Halsted fellowship. This work was supported in part by the National Institutes of Health (R21 MH098149 to K.S.L.).

Correspondence should be addressed to Kevin S. LaBar, Center for Cognitive Neuroscience, B203 LSRC Building, Research Drive, Box 90999, Durham, NC 27708-0999, USA. E-mail: klabar@duke.edu.

In this work, we tested how multi-voxel patterns of blood-oxygen level-dependent (BOLD) response predict discrete emotional states and whether the characteristics of these patterns better conform to categorical or dimensional models of emotion. Because successful classification of emotional states could equally support either model (as both models commonly predict that emotional states are experienced discretely), we examined the manner in which emotions were misclassified to test whether emotions are better characterized along valence and arousal dimensions or as independent emotion categories. To induce discrete emotions, we presented participants cinematic film clips and instrumental music immediately followed by a period of online self-report to sample categorical and dimensional aspects of emotion experience during scanning. We then conducted multivariate pattern classification followed by Bayesian model comparison to assess the discriminability and organization of emotional states in the brain.

MATERIALS AND METHODS

Participants

Thirty-two healthy, right-handed individuals ($M_{\text{age}} = 26$ years, age range = 19–38 years, 13 males) completed the study. Two additional participants were run (one male, one female), but technical issues precluded analysis of their data. In particular, one participant moved excessively during scanning (exceeding 5 mm of movement within a single run), and technical issues with behavioral response hardware prevented recording of self-report in one individual. The proportion of males and females was not found to significantly differ in the final sample ($\chi^2 = 1.125$, $P = 0.289$). All participants provided written informed consent to participate in accordance with the Duke University Institutional Review Board. All participants received $20 per hour as monetary compensation for completion of the study.

Experimental paradigm

During scanning, participants engaged in an emotion induction procedure (Stephens *et al.*, 2010; Kragel and LaBar, 2013) adapted from standardized instrumental music (Nyklicek *et al.*, 1997) and cinematic film clips (Gross and Levenson, 1995). Inducing emotions with two stimulus modalities permits generalization beyond stimulus features and enables classification across stimulus types. Although both types of stimuli contained auditory information, the nature of this information differed as the music clips did not contain speech. In this task, participants were presented with either a film or music clip (with a mean duration of 2.2 minutes), immediately followed by a 23-item affect self-report scale (Nyklicek *et al.*, 1997; Stephens *et al.*, 2010) lasting 1.9 min followed by a 1.5 min washout clip to minimize overlapping responses. On-line measurement of self-report, which has been shown to minimally alter fMRI responses to emotionally engaging stimuli (Hutcherson *et al.*, 2005), was critical in quantifying dimensional and categorical aspects of emotional experience, as post-scan assessments rely on memory judgments which may inaccurately reflect moment-to-moment fluctuations in affect.

The stimuli used were the same as those in Kragel and LaBar (2013), which have been shown to adequately sample affective space to evaluate dimensional models and to produce discrete experience of contentment, amusement, surprise, fear, anger, sadness and a neutral state (included to serve as an experimental control; Figure 1). Given that prior validation using this procedure reliably induced discrete emotional experience and differentiated autonomic responding (Stephens *et al.*, 2010; Kragel and LaBar, 2013), these emotions can be considered as a priori ground truth for classification. These emotions were chosen because they vary in both valence and arousal, whereas using the standard basic emotions (i.e. happiness, surprise, fear, anger, disgust, sadness and neutral) would bias sampling toward negative emotions with high arousal. Two music and two film clips were used to induce each

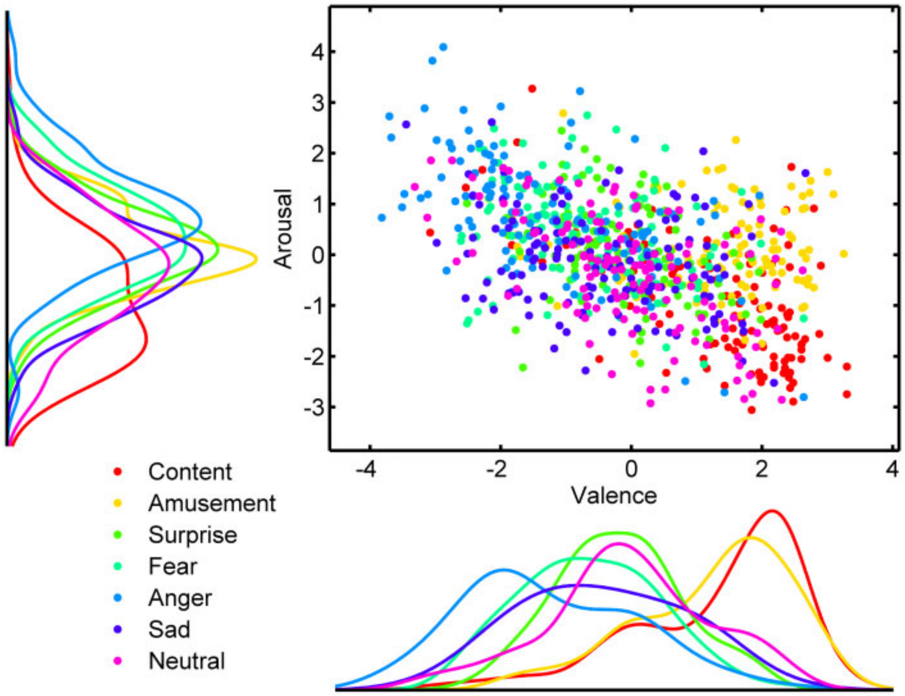

Fig. 1 Dimensional sampling from self-report of emotion. Scatterplot shows the distribution of self-reported emotion along dimensions of arousal and valence (standardized within subjects). The *x*-axis is the average of valence-related self-report items ('good', 'positive' and 'pleasant' minus the scores for 'bad', 'negative' and 'unpleasant') and the *y*-axis is the average of arousal-related items ('agitated', 'active' and 'excited' minus 'calm', 'passive' and 'relaxed'). Marginal histograms depict distributions of valence and arousal for all emotions. The distributions generally conform to dimensional models of emotion: contentment and amusement oppose fear, anger and sadness along the valence dimension whereas contentment and sadness oppose fear, anger, surprise and amusement in terms of arousal.

target emotion, resulting in a total of 28 unique stimuli—each presented a single time. Participants viewed stimuli in one of four pseudorandom counterbalanced orderings, which alternated between positive/neutral and negative valence to minimize the persistence of mood across trials. Preceding each emotion induction stimulus, participants were presented with a standardized washout video consisting of colored bars and patterned crosshairs to reduce carry-over between stimuli (Gross and Levenson, 1995).The self-report phase included items querying both categorical and dimensional aspects of emotion experience, facilitating comparisons between discrete and dimensional emotion models. The item labels on the questionnaire consisted of the following: content, amused, surprised, fearful, angry, sad, neutral, good, bad, positive, negative, calm, agitated, pleasant, unpleasant, passive, active, relaxed, excited, indifferent, intensity, enjoyment and familiarity. By sampling a number of emotions and affective dimensions on every trial, the self-report procedure minimizes conceptual biases toward one characterization of emotion over the other. Participants were instructed to 'Select the number on the scale that best describes how you felt during the music/film clip that you just listened to/viewed. If the word does not at all describe how you felt during the clip, select "1." If the word very accurately describes how you felt, select "4" and so forth'. Thus, the instructions emphasized experienced affect during the induction rather than sensory evaluative processes regarding the intended emotion based on physical properties of the stimuli.

Each experimental session consisted of seven runs of data acquisition, each comprising two trials and lasting a total of approximately 12 min. The total duration of the scanning session was accordingly 2 h. Films and music were presented on two separate days within a 2 week period in a counterbalanced order due to the length of this within-subjects procedure, to minimize the amount of instructions, and to provide seamless transitions between trials.

Presentation of stimuli and acquisition of behavioral responses was controlled using Cogent 2000 software (Wellcome Department of Imaging Neuroscience, http://www.vislab.ucl.ac.uk/cogent.php). Participants viewed stimuli on mirrors aligned with an LCD screen upon which images were projected from a stimulus control computer. Audio stimulation was presented using MR-compatible headphones (Resonance Technology, Northridge California). Responses were made by pressing the corresponding key on a fiber optic response box connected to the stimulus control computer using a universal serial bus interface (Current Designs, Philadelphia, PA, USA).

Image acquisition

Scanning was performed on a 3 Tesla General Electric MR 750 system with 50-mT/m gradients and an eight-channel head coil for parallel imaging (General Electric, Waukesha, WI, USA). High-resolution images were acquired using a 3D fast SPGR BRAVO pulse sequence: repetition time (TR) = 7.58 ms; echo time (TE) = 2.936 ms; image matrix = 256^2; $\alpha = 12^\circ$; voxel size = 1 × 1 × 1 mm; 206 contiguous slices, for coregistration with the functional data. These structural images were aligned in the near-axial plane defined by the anterior and posterior commissures. Whole-brain functional images were acquired using a spiral-in pulse sequence with sensitivity encoding along the axial plane (TR = 2000 ms; TE = 30 ms; image matrix = 64 x 128; $\alpha = 70^\circ$; voxel size = 3.8 x 3.8 x 3.8 mm; 34 contiguous slices). The first five images of each run were excluded from analyses to ensure the magnet had reached steady state.

Preprocessing and estimating neural activation

Processing of MR data was performed using Statistical Parametric Mapping software (SPM8; Wellcome Department of Imaging Neuroscience). Functional images were spatially realigned to correct for motion artifacts (Friston *et al.*, 1995), coregistered to high resolution anatomical scans (Collignon *et al.*, 1995), and normalized to Montreal Neurologic Institute space using high-dimensional warping implemented in the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm.html). In addition to normalizing anatomical and functional data, this routine provided intensity maps of gray matter, white matter and cerebral spinal fluid for each participant. Because of the relatively long duration of stimuli, conventional high pass filtering was not performed during preprocessing data, as it would attenuate task-related signal of interest.

Whole-brain patterns of neural activation were estimated using the general linear model (GLM) approach implemented in SPM8. For each subject, BOLD responses were modeled by convolving box-car functions with a canonical hemodynamic response function separately for the two induction trials in all 14 runs. Two additional regressors modeling the washout phases and self-report trials were included in each run. To model nuisance effects, six motion parameters (roll, pitch, yaw, in addition to translation in *x*, *y* and *z* dimensions) and session constants were incorporated into the model. Following voxel-wise model estimation, parameter estimates for induction trials (28 in total) were inclusively masked with a gray matter mask constructed from the average modulated gray matter output from VBM8. This procedure produced an 896 trial by 19 962 voxel pattern matrix for classification after concatenating across all 32 participants.

Assessment of self-report

To validate that the emotion induction procedure elicited experiences that were both discrete, yet subjectively varied in terms of valence and arousal, planned *t*- and *F*-tests were conducted on self-report data. Tests for discreteness were conducted by comparing the mean of the self-report item corresponding to the targeted induction against all other categories using a paired *t*-test (e.g. the mean rating of 'amused' was compared against the mean of the other six categorical items in response to the four amusing stimuli). Separate *F*-tests were performed for valence and arousal in one-way repeated measures analyses of variance (ANOVAs) with one measure per subject for each emotion category (seven in total). Statistical thresholds were corrected for multiple comparisons by using false discovery rate (FDR) correction (Benjamini and Hochberg, 1995).

Pattern classification

Decoding of emotional states was performed via partial least squares Discriminant Analysis (PLS-DA; Wold *et al.*, 2001) using the NIPALS algorithm (Martens and Næs, 1989) as implemented in the libPLS toolbox (http://www.libpls.net). This approach identifies a linear combination of features (in this case voxels) which maximally discriminate between classes. By effectively reducing the dimensionality of input features in classification, this algorithm is particularly well suited in this study as the number of features exceeds the number of observations.

Classification of multiple categories was performed using a winner-takes-all approach, wherein one class is classified against all others. Because this approach creates an uneven proportion of classes (1:6), a weighted approach was taken for discriminant analysis to minimize bias due to class imbalance. Input data (896 total instances, 28 from each participant) were mean centered before conducting the analysis. As the emotion induction procedure implemented in this study involved only four instances of each category per subject, classification was performed across rather than within subjects. We estimated classification error using 8-fold subject independent cross-validation, wherein data from 28 subjects were used to build classification models and data from four subjects were used to test performance until all subjects have been used for testing. To protect against overfitting, our primary results used 5-fold cross-validation on the training data to select number of latent variables in each model. To determine the information content of patterns, cross-validated measures of signal detection (sensitivity, specificity and the area under the receiver-operating characteristic curve) were averaged across all categories. To make statistical inference on these signal detection measures and estimate 95% confidence intervals, 10 iterations of cross-validation were performed to estimate the standard deviation of each metric. Wilson score centers and intervals were calculated to estimate confidence intervals for accuracy across all categories (Wilson, 1927). To test the robustness of our results, we additionally performed standard 5-fold cross-validation while fixing the number of latent variables between 1 and 3 or using an inner 5-fold cross-validation for model selection. Models using three latent variables were found to minimize generalization error, although using a smaller number produced accuracy above chance levels (Supplementary Figure S1).

The significance of PLS regression coefficients, which indicate the importance of each voxel in predicting emotions, was determined by bootstrap resampling of the full dataset using 10 000 replicates and three latent variables. This large number of iterations produced fine-grained estimates of confidence intervals and *z*-statistics. Statistical maps were thresholded using a voxel-wise threshold of $P < 0.05$ and extent of 14 voxels, which were shown to control the false-positive rate of $\alpha < 0.05$ in Monte Carlo simulations (Forman *et al.*, 1995). The resultant maps thus indicate in which regions activation reliably predicts a particular emotion. To verify that these maps reflected the influence of a voxel toward classification rather than the intensity of BOLD activation, we constructed scatterplots of PLS regression coefficients against the intensity range (maximum minus minimum) of parameter estimates used as input for classification, which showed minimal bias due to the scaling of the data (Supplementary Figures S2 and S3). To confirm that these maps were not driven by modality-specific information, *z* maps were created on film and music trials separately and combined using Stouffer's method. These maps were highly correlated with those generated using the full dataset (all $r > 0.79$), indicating that bootstrapping across all trials identified regions which were generally modality-independent.

The extent of overlap between thresholded maps was determined by calculating the proportion of voxels common to each unique pair of emotions relative to the total number of voxels above threshold. The average percent overlap across all 21 unique pairs was compared using a one-sample *t*-test against the average overlap created by 10 000 iterations of randomly shuffling the location of above threshold voxels and computing the same metric of relative overlap, which produced a value of 1.78%.

To test the relationship between classification errors and the subjective experience of emotion, separate generalized linear models were created for each emotion category using ratings of categorical items (e.g. 'content', 'amused', 'surprised', etc.) as independent variables and the number of predictions for each category of emotion induction as the dependent variable. As the number of predictions constitutes count data, a binomial link function was used. Statistical significance was conducted using *t*-tests on the single parameter in each model, correcting for multiple comparisons using FDR correction.

Model comparison

The organization of information in neural activation patterns was examined relative to dimensional and categorical models of emotion. As the categorical model assumes an independent axis for each emotion, the Euclidean distance between two emotions (i, j) was calculated as

$$Dcat_{i,j} = \sqrt{\sum_{e=1}^{7} (x_{e,i} - x_{e,j})^2},$$

where $x_{e,i}$ reflects the average of self-report item *e* for instances of emotion category *i*. The distance between two emotions in an affective space of valence and arousal dimensions was defined as

$$Ddim_{i,j} = \sqrt{(x_{\text{val},i} - x_{\text{val},j})^2 + (x_{\text{arous},i} - x_{\text{arous},j})^2},$$

where $x_{\text{val},i}$ reflects the average of valence-related self-report items ('good', 'positive' and 'pleasant' minus the scores for 'bad', 'negative' and 'unpleasant') across instances of emotion category *i* and $x_{\text{arous},i}$ reflects the average of arousal related self-report items ('agitated', 'active' and 'excited' minus 'calm', 'passive' and 'relaxed') across instances of emotion category *i*. Distances were computed for all 21 unique pairwise combinations of emotions for each model.

Models were fitted using distances as independent variables in a Poisson regression, as the total number of errors in classification are count data and can be modeled as a linear combination of categorical and dimensional distance. Four models were constructed: one including only a constant term, one with a constant and categorical term, one with a constant and dimensional term, and one with constant, categorical and dimensional terms. Including both dimensional and categorical terms in the combined model allows the regression to assess unique variance for each term, serving to highlight differences between the two accounts.

Bayesian model comparison was performed to determine which model was most likely to produce the distribution of errors resulting from classification of neural data. The Bayesian information criterion (BIC) score was computed for each model (which includes a penalty term for additional parameters), and Bayes factors (*B*) were subsequently estimated using the Schwarz approximation (Schwarz, 1978). We additionally converted raw BIC scores to BIC weights to visualize the relative evidence for each model (Buckland *et al.*, 1997). To determine the relationship between classification errors and distance between instances in each model space, *t*-tests were performed on parameter estimates in the favored model. Because dimensional and categorical distances were correlated ($r = 0.778$, $P < 0.0001$), multicollinearity between regressors could drive changes in the direction of observed effects. To rule out this possibility, we additionally inspected the parameter estimates in models with single terms to ensure the direction of effects did not change when including both terms.

Univariate models

To examine brain activity that varied as a function of self-reported valence and arousal, a group-level univariate ANOVA with factors of emotion (seven levels) and stimulus modality (two levels) was performed. Contrasts for each emotion (*vs* implicit baseline) were separately created for film and music trials at the subject level and used as input. Orthogonalized contrasts corresponding to the group average self-report scores for valence and arousal were performed at each gray matter voxel. To assess the similarity of PLS regression coefficients and univariate effects, separate contrasts for each emotion *vs* all others were performed. Because multiple inputs were used from the same subjects, independence was not assumed for emotion and modality factors. A correction for non-sphericity was applied to account for potential differences in variance across levels of emotion and modality. Statistical maps were thresholded using a voxel-wise threshold of $P < 0.05$ and extent of 85 voxels, which were shown to control the false-positive rate of $\alpha < 0.05$ in Monte Carlo simulations.

RESULTS

Decoding emotional states from BOLD activation patterns

Before weighing the evidence of dimensional and categorical models, we first needed to demonstrate that discrete emotional states could be predicted on the basis of brain activation. To this end, we conducted

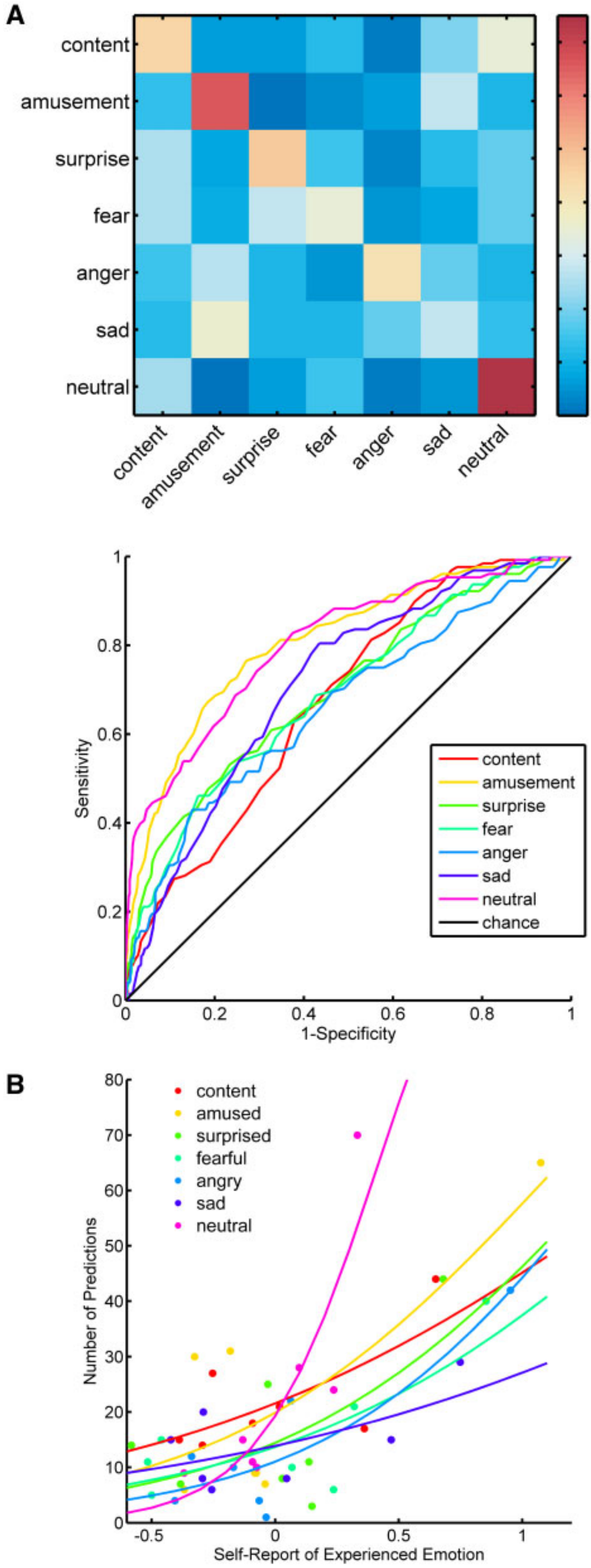

Fig. 2 Performance measures of multi-voxel pattern classification. (**A**) Confusion matrix for classifying all seven emotional states (top) and receiver-operating characteristic (ROC) curves for one-*vs*-all classifications (bottom). Increasing classification performance would yield a confusion matrix with more predictions falling along the diagonal and push ROC curves toward the upper left corner, away from chance. (**B**) Binomial regression predicting the number of instances assigned to each emotion category based on the self-reported experience of the targeted emotion. The *x*-axis reflects standardized scores of self-report for the target emotion and the *y*-axis indicates the number of trials labeled as the target emotion. Note the curves approach 18 predictions (near random assignment of 128 trials into 7 classes) as self-report scores near zero.

pattern classification using single-trial estimates of neural activation from fMRI data acquired during the presentation of film and music clips. Utilizing PLS-DA (Wold *et al.*, 2001) with 8-fold subject independent cross-validation, we classified seven emotional states (contentment, amusement, surprise, fear, anger, sadness and a neutral state) with 37.3% accuracy (where chance is 14.3%, $P < 0.001$, 95% CI [34.2%, 40.5%]) (Figure 2A). Pair-wise classification performance was above chance levels for sensitivity ($M = 0.736$, $P < 0.001$, 95% CI [0.655, 0.816]), specificity ($M = 0.560$, $P = 0.034$, 95% CI [0.506, 0.614]), and area under the receiver-operating characteristic curve (0.652, $P < 0.001$, 95% CI [0.610, 0.694]). Beyond accuracy, which is not best suited to characterize classification performance as a single number (Bradley, 1997), these measures show reliable detection of emotion-related information. High levels of sensitivity compared with specificity indicate that the classifier is better at identifying the presence of an emotion rather than the absence of an emotion, indicating that classification was not biased toward the more frequent negative class. Further, these performance metrics did not differ between film and music inductions (all P's > 0.204), suggesting that learning was not biased toward information in one modality.

To explicitly test how well the neural classification models generalized across stimulus modalities, we performed cross-classification between responses to music and films. This procedure involved training a classification model on patterns evoked using film stimuli and testing on the responses to music clips. Next, music clips were used for training models and film stimuli were used to test the performance of classification. Cross-classification accuracy of all seven emotional states was 28.38% (binomial test against chance of 14.3%, $P < 0.001$), indicating that a substantial amount of pattern-information generalized across the stimulus modality used to induce emotions.

Beyond testing performance measures for stimulus decoding, we examined whether the predictions from pattern classification tracked the subjective experience of the emotion. Binomial regression was performed to assess whether the average rating for each categorical self-report item obtained during scanning was associated with the number of predictions made during classification across all seven emotion inductions (Figure 2B). This analysis revealed the extent to which participants felt a given emotion followed the number of times that emotion was predicted during classification: 'content' ($\beta = 0.990$, $t = 4.13$, $P_{adj} < 0.0001$, 95% CI [0.519, 1.46]), 'amused' ($\beta = 1.49$, $t = 9.03$, $P_{adj} < 0.0001$, 95% CI [1.17, 1.82]), 'surprised' ($\beta = 1.49$, $t = 5.40$, $P_{adj} < 0.0001$, 95% CI [0.950, 2.03]), 'fearful' ($\beta = 1.24$, $t = 5.74$, $P_{adj} < 0.0001$, 95% CI [0.817, 1.66]), 'angry' ($\beta = 1.72$, $t = 8.04$, $P_{adj} < 0.0001$, 95% CI [1.30, 2.14]), 'sad' ($\beta = 0.79$, $t = 3.29$, $P_{adj} = 0.00099$, 95% CI [0.321, 1.26]) and 'neutral' ($\beta = 4.20$, $t = 9.01$, $P_{adj} < 0.0001$, 95% CI [3.29, 5.12]). The average coefficient of determination (R^2, computed as 1 minus the ratio of the residual sum of squares and the sum squared deviation in the dependent measure) for these regression models was 0.57 (min: 0.21 for surprise, max: 0.94 for anger), indicating that a considerable portion of the variance in the number of predictions made for different emotions could be explained by subjective ratings of the target emotion.

Given that pattern classification identified emotion-specific information in the fMRI data, we next identified the voxels in which increased levels of activation reliably predicted each emotion. We found that PLS-DA regression coefficients, which indicate the contribution of each voxel in predicting the classification outcome, were reliably greater than zero in a number of clusters distributed throughout the brain (Figure 3; Supplementary Tables S1–S7). To provide approximate localization, we computed the percent overlap of each predictive pattern with all regions from a macroscopic atlas of brain anatomy (Tzourio-Mazoyer *et al.*, 2002). Activation in a distributed set of clusters predicted states of contentment, including dorsal precuneus,

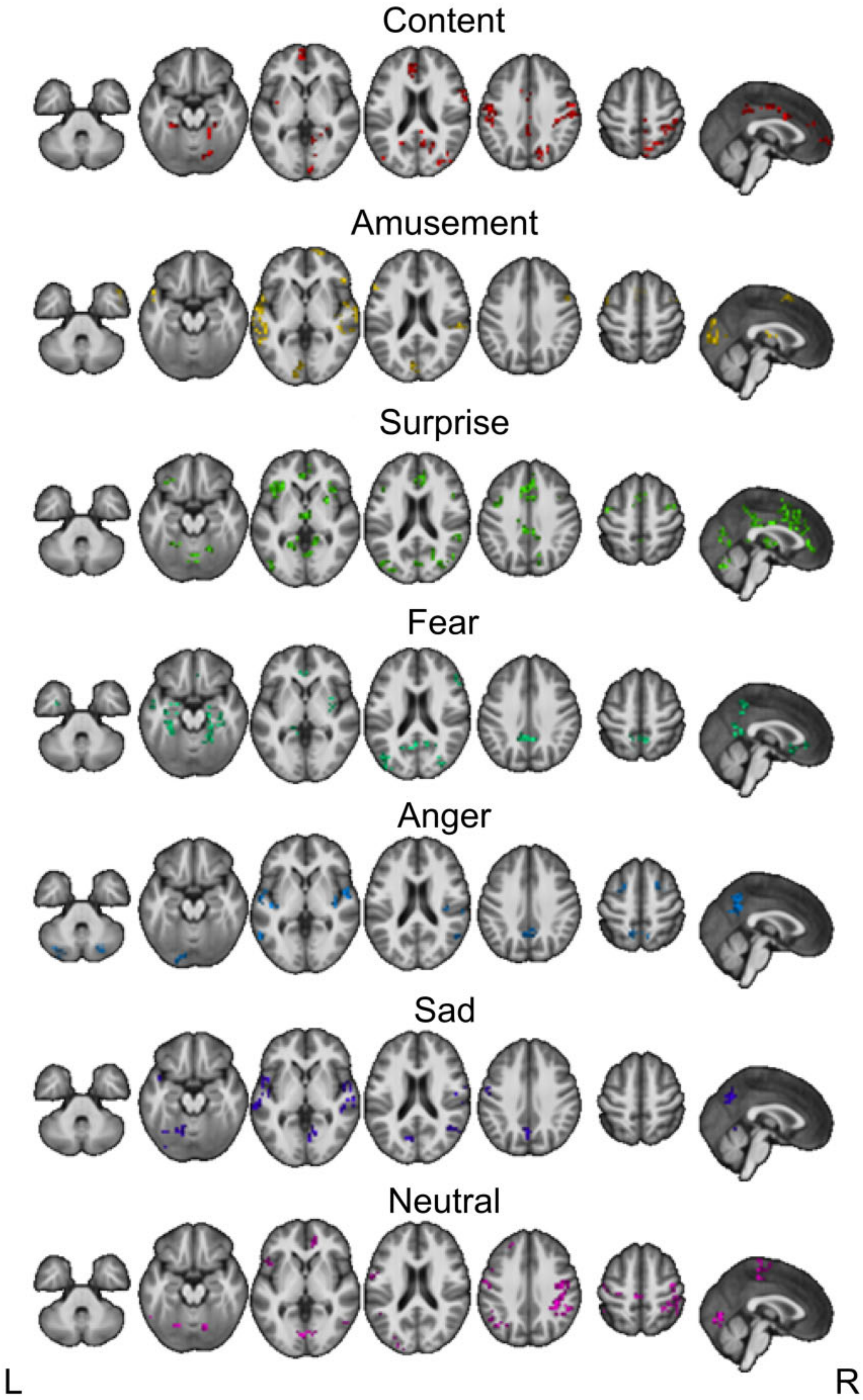

Fig. 3 Distributed patterns of neural activity predict discrete emotional states. Each map depicts voxels whose activation consistently led to predictions of the related emotion across all subjects and stimulus modalities ($P < 0.05$, FWE corrected; see Materials and Methods), overlaid on the group-average anatomical image.

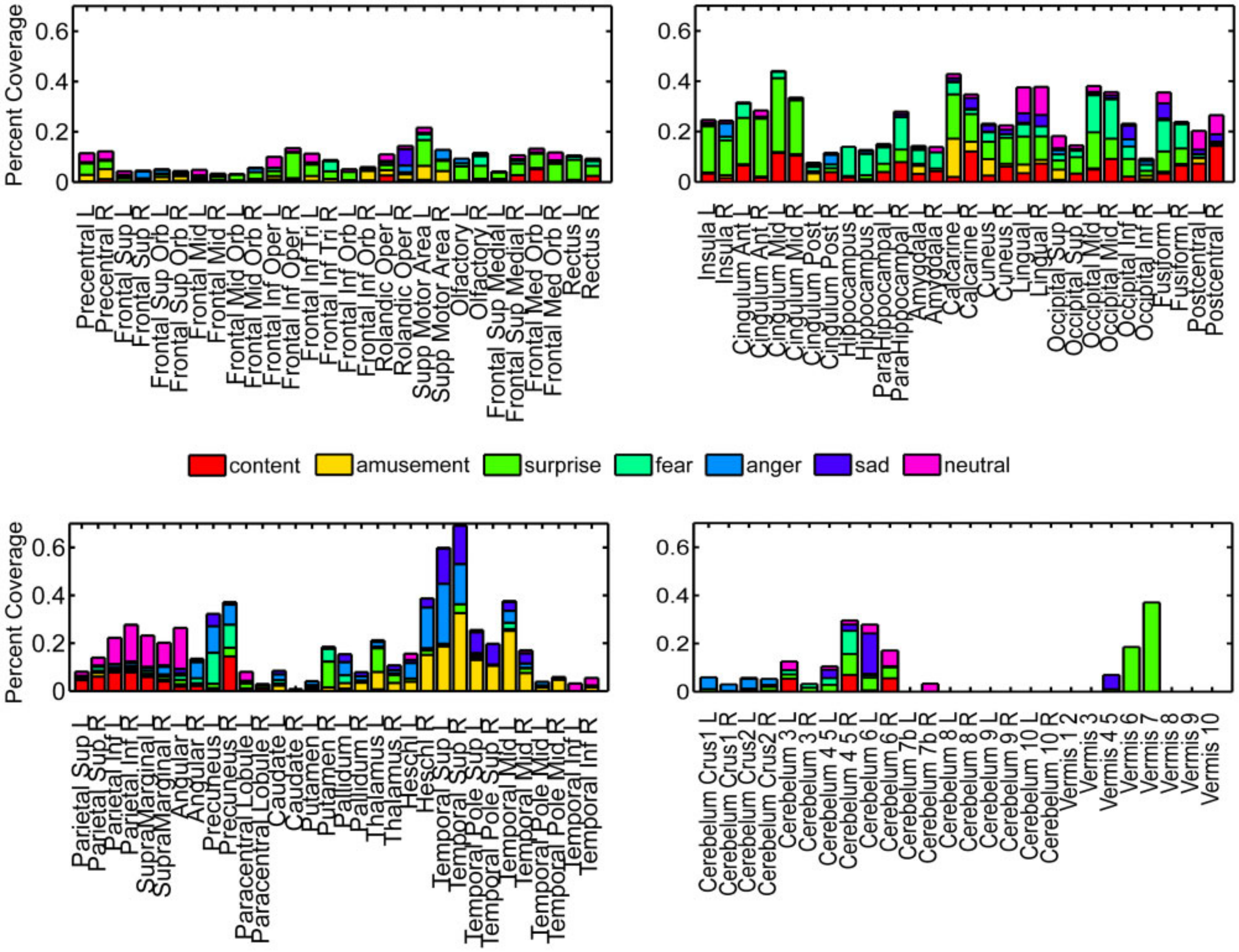

Fig. 4 Stacked bar plot showing the coverage of all regions in the AAL atlas by each emotion-predictive pattern.

bilateral postcentral gyrus and mid-cingulate gyrus. A more focal set of regions informed the classification of amusement: bilateral superior- and middle-temporal lobe, bilateral early visual cortex and bilateral supplementary motor area, among other frontal regions. Several clusters of activation led to labeling instances as surprise, including much of mid- and anterior-cingulate gyrus, bilateral anterior insula, bilateral mid-occipital cortex, bilateral thalamus and bilateral cerebellum. Predictions of fear were derived from activation within bilateral lingual gyrus, bilateral fusiform gyrus, dorsal/anterior precuneus and a number of bilateral medial temporal lobe structures: amygdala, hippocampus and parahippocampal cortex. Activation within bilateral superior temporal gyrus, ventral precuneus and right angular gyrus primarily informed predictions of anger. Patterns of activity including left cerebellum, bilateral superior temporal gyrus and bilateral temporal pole led to sad emotional states. Bilateral activation in angular gyrus, supramarginal gyrus, postcentral gyrus and lingual gyrus contributed to classification of neutral states.

Overall, we found the patterns to be distinct yet partially overlapping at this macro-scale (Figure 4), as activation within many of the same structures contributed to the prediction of several emotions. Given dimensional and categorical models predict differing degrees of differentiation in the neural substrates underlying emotions, we quantified the spatial overlap of emotion-predictive patterns. Although different emotional states were predicted by activation in some of the same regions when considered at a macro-scale of brain parcellation, the average voxel-wise overlap between patterns ($M = 2.66\%$, 95% CI [1.55%, 4.17%]) did not vary from random permutations of clusters of the same size within the same volume ($M = 1.78\%$, $P = 0.2386$, 95% CI [1.04%, 2.02%]). This result shows that despite being localized in similar macro-scale structures, the patterns were relatively non-overlapping at the voxel level. Such little overlap indicates that discrete emotional states are best differentiated on the basis of activation in a number of spatially disjoint brain regions.

Comparisons of PLS regression coefficients and analogous contrasts from univariate GLMs (i.e. each emotion *vs* the average of all others) revealed moderate spatial correspondence of parameter estimates. Pearson correlations were highly significant for estimates of contentment ($r = 0.428$), amusement ($r = 0.644$), surprise ($r = 0.368$), fear ($r = 0.828$), anger ($r = 0.531$), sad ($r = 0.612$) and neutral ($r = 0.716$) states (all $P < 0.0001$). Averaging across emotions, 5.53% of total gray matter voxels were predictive in both univariate and multivariate maps, whereas 11.05% were predictive only in multivariate maps, and 9.68% of voxels were predictive only in univariate maps. Thus, the two methods provide unique and complementary insight into neural activity associated with distinct emotions.

Testing categorical and dimensional models of emotion

In order to assess the structure of information within emotion-specific patterns, it was first necessary to validate that the emotion induction paradigm produced emotional experiences that varied along both emotion categories and affective dimensions of valence and arousal.

Planned contrasts of categorical self-report items demonstrated that participants experienced the targeted emotion to a greater extent than alternative emotions ('content': $t_{31} = 4.19$, $P_{\mathrm{adj}} = 0.0002$, 95% CI [0.870, 2.52]; 'amused': $t_{31} = 8.32$, $P_{\mathrm{adj}} < 0.0001$, 95% CI [2.16, 3.57]; 'surprised': $t_{31} = 5.70$, $P_{\mathrm{adj}} < 0.0001$, 95% CI [1.28, 2.70]; 'afraid': $t_{31} = 7.40$, $P_{\mathrm{adj}} < 0.0001$, 95% CI [1.61, 2.84]; 'angry': $t_{31} = 6.50$, $P_{\mathrm{adj}} < 0.0001$, 95% CI [1.56, 3.00]; 'sad': $t_{31} = 4.57$, $P_{\mathrm{adj}} = 0.0001$, 95% CI [1.01, 2.63]; 'neutral': $t_{31} = 2.66$, $P_{\mathrm{adj}} = 0.0123$, 95% CI [0.222, 1.68]). One-way repeated measures ANOVAs performed on dimensional self-report items revealed that emotional states could be differentiated in terms of valence ('good', 'positive' and 'pleasant' minus the scores for 'bad', 'negative' and 'unpleasant': $F_{6,186} = 50.503$, $P_{\mathrm{adj}} < 0.0001$, 95% CI [20.39, 246.7]) and arousal ('agitated', 'active' and 'excited' minus 'calm', 'passive' and 'relaxed': $F_{6,186} = 25.305$, $P_{\mathrm{adj}} < 0.0001$, 95% CI [10.217, 123.6]) dimensions. Further, plotting individual trials in an affective space organized by valence and arousal showed broad coverage and relatively uniform sampling (Figure 1), replicating prior work using the same stimuli (Kragel and LaBar, 2013). These results suggest that the instances of emotion experienced in the scanner were adequate for use in testing both categorical and dimensional models.

Next, we quantified differences in emotional experience along categorical and dimensional models by computing the Euclidean distance in self-report ratings for each of 21 unique pairs of emotional states. Distances spanned seven dimensions in the categorical model, with each dimension directly corresponding to each emotion category. The dimensional model was constructed using self-report to parameterize dimensions from negative to positive valence and low to high arousal (Kragel and LaBar, 2013). The average distance between emotions did not differ between dimensional ($M = 1.37$, 95% CI [1.05, 1.69]) and categorical models ($M = 1.55$, 95% CI [1.38, 1.72]), indicating that the emotions selected for study were equally separable along either model. To visualize the spatial homogeneity of each model, we performed hierarchical clustering of distances using Ward's method in each model's affective space. This analysis revealed a more uniform distribution of emotional states in the categorical space whereas states were more tightly clustered in the dimensional model ($F_{1,40} = 3.71$, $P = 0.0051$, 95% CI [1.50, 9.13]) (Figure 5A). These results show the models we constructed were consistent with their theoretical underpinnings: emotional states are relatively sparse and equidistant in the categorical account and more clustered and overlapping in the dimensional account.

As participants experienced emotional states that could be characterized as both independent categories and points in a low dimensional space, it is possible that pattern classifiers utilized information organized according to either or both models to classify emotional states. For instance, if the information in neural activation were to conform to a single model (e.g. the categorical model with seven axes), then the number of classification errors should decrease with distance in the model space. Alternatively, if dimensions of valence and arousal drove the classification results, then errors should decrease for inductions rated differently in terms of valence and arousal. As the dimensional and categorical models can be viewed as complementary, it is possible that increasing distance along both categorical and dimensional models could contribute to fewer errors in classification. To test these hypotheses, we conducted Bayesian model comparison (Materials and Methods; Kass and Raftery, 1995) to identify whether categorical, dimensional, or a combination of the models was most likely to produce the observed classification results.

Model comparison using Bayes factors strongly favored the combined model across all other alternatives ($B_{\text{comb-dim}} = 6.43$, $B_{\text{comb-categ}} = 2632.3$, $B_{\text{comb-con}} = 1269.7$), indicating that jointly considering dimensional and categorical conceptions of emotional states best characterized classification errors (BIC weights for individual models are plotted in Figure 5B). Testing the parameter estimate for distance in categorical space indicated that the number of errors decreased with distance ($\beta = -0.573$, $t = -2.69$, $P = 0.0073$, 95% CI $[-0.992, -0.155]$), suggesting increased experience of categorical emotions other than the target emotion increased classification errors. Surprisingly, inspection of parameter estimates for the dimensional term revealed the opposite relationship between distance in model space and classification errors ($\beta = 0.533$, $t = 4.56$, $P < 0.0001$, 95% CI [0.304, 0.762]). This positive effect remained significant when categorical distance was excluded from the model ($\beta = 0.293$, $t = 3.956$, $P < 0.0001$; Supplementary Figure S4), indicating that it was not the result of multicollinearity in the combined model. Furthermore, similar effects were observed for categorical ($\beta = -1.377$, $t = -6.63$, $P < 0.0001$, 95% CI $[-1.785, -0.970]$) and dimensional terms ($\beta = 0.5137$, $t = 4.23$, $P < 0.0001$, 95% CI [0.276, 0.752]) when predicting the distribution of errors when modality-independent cross-classification was performed. Together, these results show that instances differing more in terms of valence and arousal were more difficult to classify, which is opposite to the relationship proposed by dimensional accounts of emotion. Together, these findings demonstrate that emotional states, as represented by neural activation patterns, are most separable during categorically distinct instances and are obscured when experiences differ in terms of valence and arousal.

Examining effects of valence and arousal

Given the detrimental effect of valence and arousal on classification performance, we examined their effect on neural activity at the group level using contrasts based on self-report (Supplementary Table S8). Increasing valence was associated with activation in a number of brain regions including bilateral inferior parietal lobule, anterior dorsal cingulate, dorsomedial prefrontal cortex and left somatosensory cortex. Decreasing valence was associated with activation in posterior cingulate, precuneus, middle and lateral occipital gyri, inferior temporal gyrus and right middle frontal gyrus. Increases in self-reported arousal were associated with activation in numerous cortical and subcortical regions including thalamus, caudate, bilateral amygdala, bilateral insula, bilateral superior temporal gyrus and right inferior frontal gyrus. Decreases in arousal were associated with activity in right parietal lobe.

DISCUSSION

We have successfully predicted the occurrence of seven distinct emotional states using multivariate pattern classification across two different induction methods. In acquiring self-report ratings during scanning, we demonstrated for the first time a mapping between patterns of neural activation and an on-line assay of subjective feeling states. Relating subjective experience to neural activity is a critical advance over prior fMRI studies classifying emotional states in which experimental factors other than emotion, such as low level stimulus properties or task demands, may have contributed to classification. These findings further establish the use of multivariate classification to study emotion in brain and bodily responses (Shinkareva *et al.*, 2013; Kragel and LaBar, 2014; Nummenmaa *et al.*, 2014) and support the notion that emotions are categorically organized response patterns suited to bias behavior in functionally distinct ways.

Given the broad array of emotions induced, successful classification likely required considerable information beyond valence (or arousal) to differentiate emotional states. Our analysis of regression coefficients revealed that this information was contained within diverse patterns of activation, spanning a number of cortical and subcortical brain regions. For instance, states of contentment and amusement were

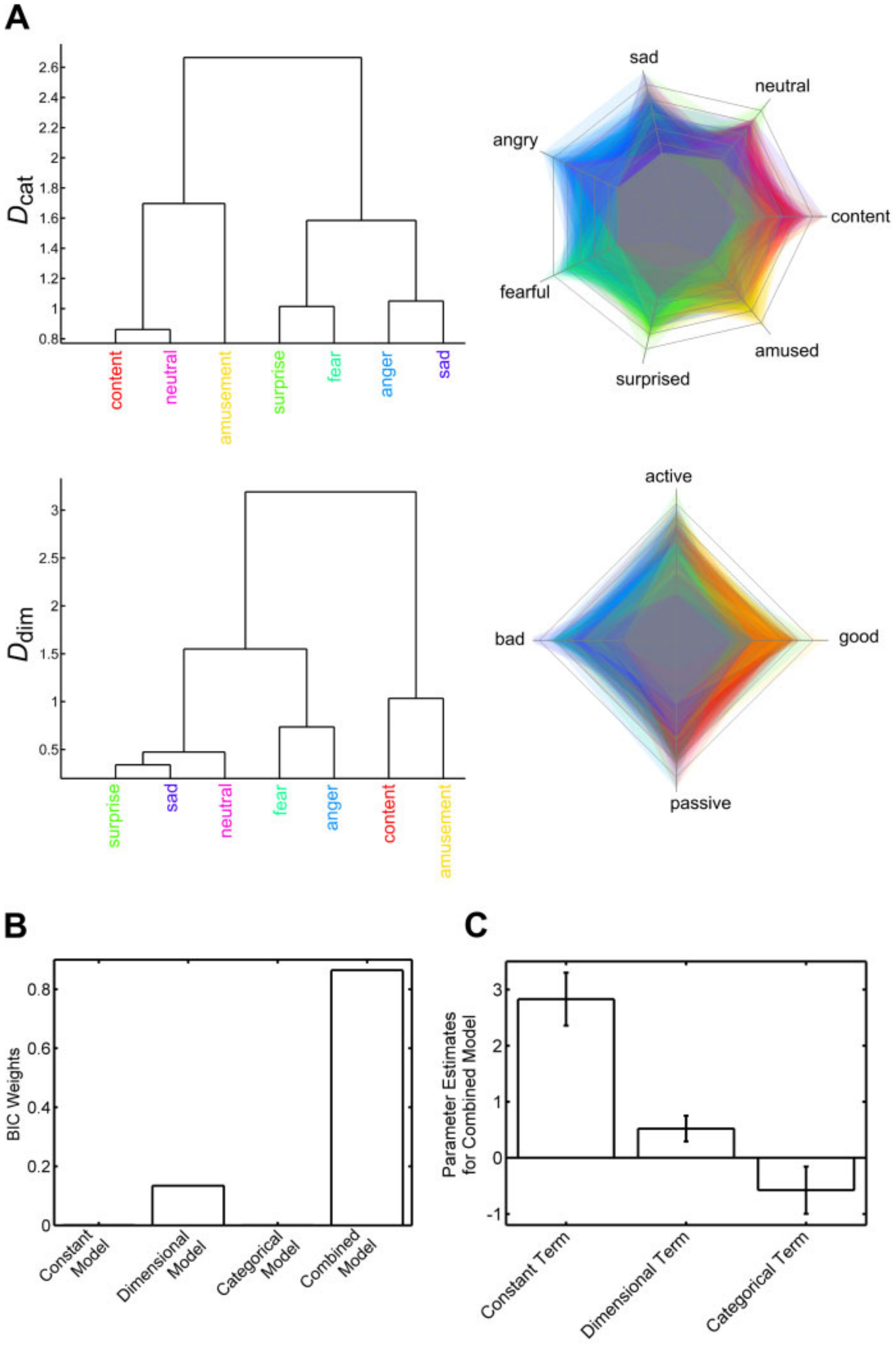

Fig. 5 Categorical and dimensional models of emotion. (A) Hierarchical clustering of self-report (left) and radar plots of categorical and dimensional terms (right). Note the increased distance between positive and negative emotions in the dimensional model and the reduced overlap between emotions in the categorical model. (B) BIC weights suggest that a combined model with both categorical and dimensional terms was most likely to produce the observed distribution of classification errors (using subject-independent cross-validation). (C) Parameter estimates from the combined model show that errors are more frequent as distance increases along valence and arousal, and that errors are less frequent as categorical distance increases. Error bars reflect 95% confidence intervals.

both rated as highly pleasant and positive, yet quantitatively disjoint patterns of activity formed their prediction. Maps for contentment included precuneus, medial prefrontal, cingulate and primary somatosensory cortices—regions implicated in self-referential processing (Northoff *et al.*, 2006) and topographic representation of bodily sensations. Predictions of amusement were largely informed by activation in temporal cortex, supplementary motor area and thalamus, which have been linked to processing of humor, laughter and smiling (Wild *et al.*, 2003, 2006). Although the regions contained within these patterns are not individually specific to either emotional state (or emotion in general), accumulating information across these disparate regions specified emotions as they were experienced as distinct categories.

Despite engaging partially overlapping neural substrates at the macro-scale, emotion-predictive patterns were largely non-overlapping at the voxel level. Such separability of emotional states at the voxel level may explain why meta-analytic works (Phan *et al.*, 2002; Murphy *et al.*, 2003; Vytal and Hamann, 2010) have associated neural activity with discrete emotions (e.g. correspondence between activation within the amygdala and fear or dorsal anterior cingulate and happiness), yet have failed to consistently identify emotion-specific neural substrates

(Lindquist *et al.*, 2012). The specificity and spatial separability of patterns identified in this study suggest that methodological limitations inherent to univariate meta-analyses (variability of spatial normalization, focus on individual brain regions and necessity of spatial smoothing) may have complicated the localization of emotion-specific substrates in prior work.

We found that univariate and multivariate maps exhibited moderately similar spatial organization, although a sizeable number of voxels were differentially predictive across the two methods. Because multivariate classifiers pool information carried across multiple voxels and are generally more sensitive than univariate approaches (Jimura and Poldrack, 2012), the multivariate patterns presented here complement and extend prior univariate findings. Further, it is important to note that PLS estimates better reflect the underlying latent structure in the fMRI data compared with ordinary least squares estimates (i.e. conventional GLMs), as they are less influenced by error variances at each voxel (Burnham *et al.*, 2001). Comparisons and interpretations of the present results against prior and future univariate studies should accordingly be made with these methodological differences in mind.

Through Bayesian model comparison, we demonstrated that integrating categorical and dimensional configurations of affect best modeled the information contained in neural patterns during the experience of discrete emotional states. Although ratings made using categorical labels were sufficient to differentiate the emotional states, self-reports of emotion could also be distinguished along dimensions of valence and arousal, concordant with prior research on the similarity of emotion words (Bush, 1973; Russell, 1980), emotional experiences (Barrett and Russell, 1999) and facial expressions of emotion (Schlosberg, 1952; Susskind *et al.*, 2007). In this sense, classification of neural activation could be driven by either dimensional or categorical aspects of experience (Kragel and LaBar, 2013). Strikingly, this was not the case in the neural data: the optimal model showed that emotional states that differed the most in terms of valence and arousal were the most difficult to predict, whereas differences on categorical aspects of experience led to better classification. Neural responses related to valence or arousal may have obscured patterns of neural activation that more effectively differentiate emotional states.

From a machine learning perspective, neural variance related to valence or arousal may have acted more as noise (which led to poorer classification) as opposed to informative features that differentiate emotions. This is not to say that neural correlates of valence and arousal are unstructured or unrelated to the experience of emotion. Rather, valence or arousal-related activity likely co-occurs with emotional experiences but poorly discriminates among emotions similar along such dimensions, leading to more classification errors. For example, our univariate analyses revealed activation in middle and superior temporal gyrus associated with increasing arousal (consistent with Mourao-Miranda *et al.*, 2003; Kensinger and Schacter, 2006; Mather *et al.*, 2006), whereas classification models identified distinct patterns of activation in this region that contributed to prediction of amusement, anger and sadness. Widespread activation associated with arousal may have masked information specific to each emotion, possibly explaining higher rates of confusion between these emotions. Such opposition between categorical and dimensional representation is incompatible with models suggesting emotions emerge from undifferentiated states of valence and arousal, as such accounts suggest differences in affective dimensions should aid in discrimination among emotions.

Although the present findings refute the idea that valence and arousal serve to differentiate emotional states as reflected in patterns of neural activation, such affective dimensions have been linked to distinct neural systems (Anderson *et al.*, 2003; Small *et al.*, 2003; Lewis *et al.*, 2007; Lang and Bradley, 2010; Chikazoe *et al.*, 2014). This body of research localized the neural basis of these affective properties using pre-specified regions of interest. It is possible that using whole-brain data for classification of emotional state in this study overlooked information represented locally in these regions, due to the presence of other signals that better distinguished emotions. Alternatively, the sampling of emotions could have prevented differences in valence and arousal from informing classification. As multiple emotions were similar in terms of valence and/or arousal (e.g. fear and anger or content and neutral states), pattern classifiers may have underutilized activation related to these dimensions if it was present in the data. For these reasons, the present results should not be taken as *de facto* evidence against dimensional representations of affect in the brain. Rather, these findings implicate separable neural substrates in multifaceted categorical experience of emotion beyond a continuous mapping of valence and arousal.

We found that neutral states were among those most accurately classified. This finding is consistent with recent studies examining the perception of facial expressions (Said *et al.*, 2010) and bodily sensations (Nummenmaa *et al.*, 2014). Neutral states are likely to have high classification accuracy because they are more likely to differ from all other emotions along either dimensional or categorical factors. In a circumplex dimensional account, for instance, the neutral origin is equidistant from emotions organized along the perimeter of affective space. Emotions that are clustered about the perimeter are thus more difficult to differentiate, as they share components of valence and arousal. Additionally, neutral states can be differentiated on the basis of activation at the superordinate non-emotional *vs* emotional level, whereas discrete emotions must be classified only at the subordinate level as they all contain emotional content.

The present findings extend related research classifying emotional states on the basis of brain activity. In recent work by Kassam *et al.* (2013), states of anger, disgust, envy, fear, happiness, lust, pride, sadness and shame were self-induced by method actors during the presentation of an emotional words (e.g. afraid or frightened). These states were identified at accuracies well above chance levels of 0.50 using classification models both within (mean rank accuracy of 0.84) and across subjects (mean rank accuracy of 0.70), and an exploratory factor analysis revealed that neural activation during the self-induction period was organized along dimensions interpreted as valence, arousal, sociality and lust. Although the present results are consistent with their findings of above chance accuracy levels, we did not find evidence indicating that neural activation was organized along dimensions of valence and arousal. It is possible that the presence of emotional words, which have a semantic structure well characterized along such dimensions (Storm and Storm, 1987) and have been shown to elicit responses in frontal and limbic structures (Lewis *et al.*, 2007), may explain this difference in results.

As emotion induction trials lasted several minutes, the activation patterns we classified may have been driven by multiple processes which ultimately lead to categorically experienced states. The acquired time-course of neural activation likely reflected perceptual evaluation of stimuli, physiological expression of emotion and construction of emotional experiences. Functional neuroimaging studies of emotion perception (Peelen *et al.*, 2010; Said *et al.*, 2010) have implicated the superior temporal sulcus in representing the emotional content of stimuli categorically. Psychophysiological research has demonstrated categorical specificity in peripheral physiological responses (Stephens *et al.*, 2010; Kragel and LaBar, 2013), which have been shown to bidirectionally predict subjective feelings (Friedman *et al.*, 2014). Given the numerous and interrelated processes that unfold during an emotional episode, it is unlikely that the patterns we identified perfectly capture the essence of an emotion, but are better characterized as an amalgamation of components which make emotions unique. Future research examining which aspects of the task are essential for categorical experience of emotion will be necessary to further evaluate theories concerning the neural basis of emotions, and whether they sufficiently meet the criteria for being considered biologically basic.

In conclusion, we have found mappings between neural activation patterns and categorically distinct emotional experiences, raising numerous questions central to the representation of emotion in the brain. The patterns identified here clarify the relationship between brain activity and the experience of a subset of emotions, but they do not fully define the emotional experience of participants. Quantifying other factors, such as motivational state or social context, could further elucidate the structural organization of emotions in the brain. Assessing generalization to other forms of induction is another important avenue to further validate and refine neural representations of emotion. Examining whether the experience of emotion can be modeled at more fine-grained intervals through parametric manipulations and continuous prediction models (e.g. SVR, LASSO-PCR, or PLS regression, see Marquand *et al.*, 2010; Wager *et al.*, 2013; Favilla *et al.*, 2014) may yield more sensitive accounts of emotion experience. Beyond informing neural and theoretical models of emotion, biomarkers of emotional states could improve our understanding of mood and affective disorders, aid in diagnosis, and serve as novel outcome measures for interventions. Key diagnostic criteria for these disorders include subjective feelings of fear, anxiety and sadness. Such feelings are complex and difficult to quantify objectively. Therefore, examining the prevalence and congruence of these patterns in healthy and clinical populations may further clarify the order and disorder of emotion in the human brain.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

CONFLICT OF INTEREST

None declared.

REFERENCES

Anderson, A.K., Christoff, K., Stappen, I., et al. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience*, *6*(2), 196–202.

Barrett, L.F. (2006). Solving the emotion paradox: categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1), 20–46.

Barrett, L.F. (2012). Emotions are real. *Emotion*, *12*(3), 413–29.

Barrett, L.F., Russell, J.A. (1999). The structure of current affect: controversies and emerging consensus. *Current Directions in Psychological Science*, *8*(1), 10–4.

Barrett, L.F., Wager, T.D. (2006). The structure of emotion—evidence from neuroimaging studies. *Current Directions in Psychological Science*, *15*(2), 79–83.

Baucom, L.B., Wedell, D.H., Wang, J., Blitzer, D.N., Shinkareva, S.V. (2012). Decoding the neural representation of affective states. *Neuroimage*, *59*(1), 718–27.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate—a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B—Methodological*, *57*(1), 289–300.

Bradley, A.P. (1997). The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145–59.

Buckland, S.T., Burnham, K.P., Augustin, N.H. (1997). Model selection: an integral part of inference. *Biometrics*, *53*(2), 603–18.

Burnham, A.J., MacGregor, J.F., Viveros, R. (2001). Interpretation of regression coefficients under a latent variable regression model. *Journal of Chemometrics*, *15*(4), 265–84.

Bush, L.E. (1973). Individual differences multidimensional scaling of adjectives denoting feelings. *Journal of Personality and Social Psychology*, *25*(1), 50–7.

Chikazoe, J., Lee, D.H., Kriegeskorte, N., Anderson, A.K. (2014). Population coding of affect across stimuli, modalities and individuals. *Nature Neuroscience*, *17*(8), 1114–22.

Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G. (1995). Automated multi-modality image registration based on information theory. *Information Processing in Medical Imaging*, *3*, 263–74.

Damasio, A., Carvalho, G.B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nature Reviews. Neuroscience*, *14*(2), 143–52.

Ekman, P., Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364–70.

Favilla, S., Huber, A., Pagnoni, G., et al. (2014). Ranking brain areas encoding the perceived level of pain from fMRI data. *Neuroimage*, *90*, 153–62.

Feldman Barrett, L., Russell, J.A. (1999). The structure of current affect: controversies and emerging consensus. *Current Directions in Psychological Science*, *8*(1), 10–4.

Fontaine, J.R.J., Scherer, K.R., Roesch, E.B., Ellsworth, P.C. (2007). The world of emotions is not two-dimensional. *Psychological Science*, *18*(12), 1050–7.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance in Medicine*, *33*(5), 636–47.

Friedman, B.H., Stephens, C.L., Thayer, J.F. (2014). Redundancy analysis of autonomic and self-reported, responses to induced emotions. *Biological Psychology*, *98*, 19–28.

Friston, K.J., Ashburner, J., Frith, C.D., Poline, J.B., Heather, J.D., Frackowiak, R.S.J. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, *3*(3), 165–89.

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

Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends in Cognitive Sciences*, *16*(9), 458–66.

Hamann, S.B., Ely, T.D., Hoffman, J.M., Kilts, C.D. (2002). Ecstasy and agony: activation of the human amygdala in positive and negative emotion. *Psychological Science*, *13*(2), 135–41.

Hutcherson, C.A., Goldin, P.R., Ochsner, K.N., Gabrieli, J.D., Barrett, L.F., Gross, J.J. (2005). Attention and emotion: does rating emotion alter neural responses to amusing and sad films? *Neuroimage*, *27*(3), 656–68.

Izard, C.E. (1971). *The Face of Emotion*. New York: Appleton-Century-Crofts.

Jimura, K., Poldrack, R.A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia*, *50*(4), 544–52.

Johnson-Laird, P.N., Oatley, K. (1992). Basic emotions, rationality, and folk theory. *Cognition & Emotion*, *6*(3–4), 201–23.

Kass, R.E., Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–95.

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

Kensinger, E.A., Schacter, D.L. (2006). Processing emotional pictures and words: effects of valence and arousal. *Cognitive, Affective & Behavioral Neuroscience*, *6*(2), 110–26.

Kragel, P.A., LaBar, K.S. (2013). Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions. *Emotion*, *13*(4), 681–90.

Kragel, P.A., LaBar, K.S. (2014). Advancing emotion theory with multivariate pattern classification. *Emotion Review*, *6*(2), 160–174.

Lang, P.J., Bradley, M.M. (2010). Emotion and the motivational brain. *Biological Psychology*, *84*(3), 437–50.

Lewis, P.A., Critchley, H.D., Rotshtein, P., Dolan, R.J. (2007). Neural correlates of processing valence and arousal in affective words. *Cerebral Cortex*, *17*(3), 742–8.

Lindquist, K.A. (2013). Emotions emerge from more basic psychological ingredients: a modern psychological constructionist model. *Emotion Review*, *5*(4), 356–68.

Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., Barrett, L.F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences*, *35*(3), 121–43.

Marquand, A., Howard, M., Brammer, M., Chu, C., Coen, S., Mourao-Miranda, J. (2010). Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *Neuroimage*, *49*(3), 2178–89.

Martens, H., Næs, T. (1989). *Multivariate Calibration*. Chichester, UK/New York: Wiley.

Mather, M., Mitchell, K.J., Raye, C.L., Novak, D.L., Greene, E.J., Johnson, M.K. (2006). Emotional arousal can impair feature binding in working memory. *Journal of Cognitive Neuroscience*, *18*(4), 614–25.

Mehrabian, A. (1996). Pleasure arousal dominance: a general framework for describing and measuring individual differences in temperament. *Current Psychology*, *14*(4), 261–92.

Mourao-Miranda, J., Volchan, E., Moll, J., et al. (2003). Contributions of stimulus valence and arousal to visual activation during emotional perception. *Neuroimage*, *20*(4), 1955–63.

Murphy, F.C., Nimmo-Smith, I., Lawrence, A.D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cognitive, Affective & Behavioral Neuroscience*, *3*(3), 207–33.

Northoff, G., Heinzel, A., Greck, M., Bennpohl, F., Dobrowolny, H., Panksepp, J. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage*, *31*(1), 440–57.

Nummenmaa, L., Glerean, E., Hari, R., Hietanen, J.K. (2014). Bodily maps of emotions. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(2), 646–51.

Nyklicek, I., Thayer, J.F., van Doornen, L.J.P. (1997). Cardiorespiratory differentiation of musically-induced emotions. *Journal of Psychophysiology*, *11*(4), 304–21.

Panksepp, J. (1982). Toward a general psycho-biological theory of emotions. *Behavioral and Brain Sciences*, *5*(3), 407–22.

Peelen, M.V., Atkinson, A.P., Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *The Journal of Neuroscience*, *30*(30), 10127–34.

Phan, K.L., Wager, T., Taylor, S.F., Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, *16*(2), 331–48.

Plutchik, R., Kellerman, H. (1980). *Theories of Emotion*. New York: Academic Press.

Rolls, E.T., Grabenhorst, F., Franco, L. (2009). Prediction of subjective affective state from brain activations. *Journal of Neurophysiology*, *101*(3), 1294–308.

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–78.

Russell, J.A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–72.

Said, C.P., Moore, C.D., Engell, A.D., Todorov, A., Haxby, J.V. (2010). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *Journal of Vision*, *10*(5), 11.

Schlosberg, H. (1952). The description of facial expressions in terms of 2 dimensions. *Journal of Experimental Psychology*, *44*(4), 229–37.

Schwarz, G. (1978). Estimating dimension of a model. *Annals of Statistics*, *6*(2), 461–4.

Shinkareva, S.V., Wang, J., Kim, J., Facciani, M.J., Baucom, L.B., Wedell, D.H. (2013). Representations of modality-specific affective processing for visual and auditory stimuli derived from functional magnetic resonance imaging data. *Human Brain Mapping*, *35*(7), 3558–68.

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*(2), 753–65.

Small, D.M., Gregory, M.D., Mak, Y.E., Gitelman, D., Mesulam, M.M., Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron*, *39*(4), 701–11.

Stephens, C.L., Christie, I.C., Friedman, B.H. (2010). Autonomic specificity of basic emotions: evidence from pattern classification and cluster analysis. *Biological Psychology*, *84*(3), 463–73.

Storm, C., Storm, T. (1987). A taxonomic study of the vocabulary of emotions. *Journal of Personality and Social Psychology*, *53*(4), 805–16.

Susskind, J.M., Littlewort, G., Bartlett, M.S., Movellan, J., Anderson, A.K. (2007). Human and computer recognition of facial expressions of emotion. *Neuropsychologia*, *45*(1), 152–62.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*(1), 273–89.

Vytal, K., Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *Journal of Cognitive Neuroscience*, *22*(12), 2864–85.

Wager, T.D., Atlas, L.Y., Lindquist, M.A., Roy, M., Woo, C.W., Kross, E. (2013). An fMRI-based neurologic signature of physical pain. *The New England Journal of Medicine*, *368*(15), 1388–97.

Watson, D., Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219–35.

Wild, B., Rodden, F.A., Grodd, W., Ruch, W. (2003). Neural correlates of laughter and humour. *Brain*, *126*(Pt 10), 2121–38.

Wild, B., Rodden, F.A., Rapp, A., Erb, M., Grodd, W., Ruch, W. (2006). Humor and smiling: cortical regions selective for cognitive, affective, and volitional components. *Neurology*, *66*(6), 887–93.

Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, *22*, 209–12.

Wilson-Mendenhall, C.D., Barrett, L.F., Barsalou, L.W. (2013). Neural evidence that human emotions share core affective properties. *Psychological Science*, *24*(6), 947–56.

Winston, J.S., Gottfried, J.A., Kilner, J.M., Dolan, R.J. (2005). Integrated neural representations of odor intensity and affective valence in human amygdala. *The Journal of Neuroscience*, *25*(39), 8903–7.

Wold, S., Sjostrom, M., Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–30.