

Distributed affective space represents multiple emotion categories across the human brain

Heini Saarimäki,¹ Lara Farzaneh Ejtehadian,¹ Enrico Glerean,^{2,3,4}
Iiro P. Jääskeläinen,^{1,5} Patrik Vuilleumier,^{6,7} Mikko Sams,^{1,3} and
Lauri Nummenmaa^{1,2,8}

¹Department of Neuroscience and Biomedical Engineering, Aalto University, FI-00076 Espoo, Finland, ²Turku PET Centre, University of Turku, FI-20520 Turku, Finland, ³Department of Computer Science, Aalto University, FI-00076 Espoo, Finland, ⁴Helsinki Institute for Information Technology, Aalto University, FI-00076 Espoo, Finland, ⁵Aalto NeuroImaging, Aalto University, FI-00076 Espoo, Finland, ⁶Department of Neuroscience, University Medical Center of Geneva, CH-1211 Geneva, Switzerland, ⁷Swiss Center for Affective Sciences, Campus Biotech, University of Geneva, CH-1211 Geneva, Switzerland, and ⁸Department of Psychology, University of Turku, FI-20520 Turku, Finland

Correspondence should be addressed to Heini Saarimäki, Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, P.O. Box 12200, FI-00076 AALTO, Finland. E-mail: heini.saarimaki@aalto.fi.

Abstract

The functional organization of human emotion systems as well as their neuroanatomical basis and segregation in the brain remains unresolved. Here, we used pattern classification and hierarchical clustering to characterize the organization of a wide array of emotion categories in the human brain. We induced 14 emotions (6 ‘basic’, e.g. fear and anger; and 8 ‘non-basic’, e.g. shame and gratitude) and a neutral state using guided mental imagery while participants’ brain activity was measured with functional magnetic resonance imaging (fMRI). Twelve out of 14 emotions could be reliably classified from the haemodynamic signals. All emotions engaged a multitude of brain areas, primarily in midline cortices including anterior and posterior cingulate gyri and precuneus, in subcortical regions, and in motor regions including cerebellum and premotor cortex. Similarity of subjective emotional experiences was associated with similarity of the corresponding neural activation patterns. We conclude that different basic and non-basic emotions have distinguishable neural bases characterized by specific, distributed activation patterns in widespread cortical and subcortical circuits. Regionally differentiated engagement of these circuits defines the unique neural activity pattern and the corresponding subjective feeling associated with each emotion.

Key words: emotion; fMRI; MVPA; pattern classification

Introduction

The organization of human emotion systems is currently a topic of lively debate (Hamann, 2012; Lindquist et al., 2012; Kragel and LaBar, 2014; Meaux and Vuilleumier, 2015; Saarimäki et al., 2016;

Adolphs, 2017; Barrett, 2017). This discussion revolves around the number of distinct emotion systems and the organization of human emotion circuits in the brain. Most research on specific emotion categories has focused on ‘primary’ or ‘basic’ emotions

Received: 8 June 2017; Revised: 25 February 2018; Accepted: 28 March 2018

© The Author(s) (2018). Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

(usually anger, fear, disgust, happiness, sadness and surprise). According to the basic emotion theories, emotions have been shaped during the evolution to serve distinct survival functions via distinct neural circuits and physiological systems (Panksepp, 1982; Ekman, 1992, 1999; Damasio, 1999). In turn, constructivist theories consider specific emotions emerging from the interaction of fundamental processes shared across psychological domains and, thus, not unique to emotions alone (see, e.g. Russell, 2003; Lewis and Liu, 2011; Hamann, 2012; Cunningham, 2013). Basic emotion theories emphasize similarities in emotion mechanisms across individuals, while constructivist theories emphasize inter- and within-individual variability. Several human neuroimaging studies support the view that at least the canonical basic emotions have distinguishable neural bases, as they are associated with discernible neural activity patterns as measured by BOLD-functional magnetic resonance imaging (fMRI) (e.g. Kragel and LaBar 2014; Saarimäki et al., 2016; for a review, see Kragel and LaBar, 2016; Nummenmaa & Saarimäki, 2017).

Yet, a wide array of other emotions, including 'secondary' or 'social' emotions (see reviews and proposed taxonomies in Damasio, 1999; Adolphs, 2002a), also serve adaptive survival functions and are characterized by distinctive facial expressions (Baron-Cohen et al., 2001; Shaw et al., 2005), bodily sensations (Nummenmaa et al., 2014a), and neural activity patterns (Kassam et al. 2013; Kragel and LaBar, 2015). Nevertheless, the psychological and neural mechanisms of these non-basic emotions, as well as their commonalities or differences relative to basic emotions, remain largely unresolved (Ekman, 1999; Ekman and Cordaro, 2011; Adolphs, 2002b). These emotions may involve more elaborate cognitive representations acquired through experience, education, and social norms (Panksepp and Watt, 2011), and hence, recruit brain systems partly distinct from those implicated in more 'primitive' and possibly innate basic emotions. It is thus possible that also non-basic emotions may have distinguishable neural bases which would, however, be discernible from that of basic emotions. Accordingly, the specific number of basic emotions has been debated and the distinction between basic and non-basic emotions has been questioned (e.g. Russell, 2003). For example, Adolphs (2002b) proposed that affective processes could be understood at various taxonomies depending on the level of analysis. The basic emotion conceptualization would be particularly relevant for primary-process core affects governed by phylogenetically old emotion circuits especially within the subcortical structures (Panksepp and Watt, 2011).

A set of core emotion processing regions is consistently engaged during multiple emotions. These include cortical midline regions (Peelen et al., 2010; Chikazoe et al., 2014; Trost et al., 2012), somatomotor regions (Adolphs et al., 2000; de Gelder et al., 2004; Nummenmaa et al., 2008, 2012; Pichon et al., 2008), as well as amygdala, brainstem and thalamus (Adolphs, 2010; Damasio and Carvalho, 2013; Kragel and LaBar, 2014). These regions serve as candidate areas containing distinct neural signatures for different basic emotions (Peelen et al., 2010; Saarimäki et al., 2016). Yet, it is currently unclear whether these regions also code for other, non-basic emotions. Prior studies using univariate analyses have quantified neural responses to emotions such as regret (Coricelli et al., 2005; Eryilmaz et al., 2011), guilt (Wagner et al., 2011), pride (Takahashi et al., 2008; Zahn et al., 2009; Simon-Thomas et al., 2012), rejoice (Chandrasekhar et al., 2008) and maternal love (Bartels and Zeki, 2004), as well as aesthetic feelings such as wonder or nostalgia (Vuilleumier and Trost, 2015). Yet, these studies (except Kassam et al., 2013; Kragel and LaBar 2014) have usually compared brain activation differences between two emotions in a univariate fashion. Thus, it remains

unclear whether similar, distinct circuits as previously observed for basic emotions would also support these types of emotions.

Here, we investigated the neural underpinnings of multiple basic and non-basic emotion categories. We induced 14 emotions in participants by guided mental imagery while their brain activity was measured with fMRI. First, to examine whether different emotions have distinguishable brain bases, we employed pattern classification. Second, using hierarchical clustering, we examined the similarity of neural substrates of different emotions and tested how this similarity was related to how similarly these emotions are experienced. Third, we mapped representation of different emotions in the core emotion processing regions of the brain, by characterizing the emotion-dependent neural response profiles using univariate analyses and cumulative activation mapping.

Materials and methods

Participants

Twenty-five female volunteers (aged 19–38, mean age 23.6 years) participated in the experiment. All were right-handed, neurologically healthy and with normal or corrected-to-normal vision, and gave written informed consent according to the Declaration of Helsinki. The Institutional Review Board of Aalto University approved the experimental protocol. Female participants were chosen to maximize the power of the experiment, as when compared to males, females typically experience and portray more intensive emotional reactions (see review in Fischer and LaFrance, 2015), and show greater brain responses during emotions (e.g., Hofer et al., 2006) and stronger facial mimicry as indexed by EMG (Grossman and Wood, 1993).

Stimuli

The stimuli for the guided affective imagery were sixty 5–20-s long narratives describing an antecedent event triggering prominently one emotional state. Narrative-based guided imagery is known to be an efficient emotion induction technique that engages reliably the brain's emotion circuits (Costa et al., 2010; Nummenmaa et al., 2014b) and results in strong subjective emotional feelings. Each narrative elicited primarily one out of possible 14 emotions, or a neutral emotional state. Targeted emotions included six basic or primary emotions (anger, fear, disgust, happiness, sadness and surprise) and eight social or secondary emotions (shame, pride, longing, guilt, love, contempt, gratitude and despair). The narratives included a short description of a situation that participants were instructed to imagine would happen to them, for instance, 'It is a late night on a dimly-lit parking lot. Your car stands alone in a dark corner. Suddenly you hear a gun shot behind you.' (fear), or 'Your lover is lying next to you on a bed. You look into his eyes when he gently touches your hair and bends to kiss your lips.' (love). Based on an online pilot experiment (see [Supplementary Material](#)), we selected four narratives per category (total of 60 narratives; see [Supplementary Table S1](#) for all narrative stimuli) to be included in the fMRI experiment.

The selected narratives were spoken by a female speaker using neutral prosody without cues for the affective content of the narrative. The background noise in the recording room was recorded and equalized (bandpass filter 50–10 000 Hz) with Apple Logic Pro 9 (Apple Inc.), and gate and compressor were used to attenuate the background noise during moments of silence and slightly compress the voice dynamic to limit the

variance of the sound power. The loudness of each narrative was normalized according to the peak value.

The recorded narratives were divided into 4 runs of 15 narratives, with one narrative per category in each run. The runs contained the same narratives for all participants but the presentation order was randomized within each run and for each participant. During fMRI, the four sets of recorded narratives were all presented twice thus resulting in altogether eight runs. Each run lasted 7–8 min and consisted of 15 trials. A trial started with a fixation cross shown for 0.5 s, followed by a 2-s presentation of a word describing the target emotion (anger, fear, disgust and so forth) to prepare participants for the forthcoming emotional imagery task and thus to make the induction more powerful. Next, the narrative was spoken out, followed by a 10-s imagery period. The trial ended with a 10-s wash-out period to counter for possible carryover effects. Participants were instructed to try to get involved in the narratives by imagining the described events as happening to themselves and to experience the corresponding feeling as vividly as possible.

Auditory stimuli were delivered through Sensimetrics S14 insert earphones (Sensimetrics Corporation, Malden, MA, USA). Sound was adjusted for each participant to be loud enough to be heard over the scanner noise. The visual stimuli were delivered using Presentation software (Neurobehavioral Systems Inc., Albany, CA, USA), and 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, participants were presented with the narratives again. For each narrative, the participants rated the felt intensity of each of the 14 emotions (plus neutral state) using a continuous scale arbitrarily ranging from 0 to 1. (for details, see [Supplementary Material](#)). We ran k-means clustering on the resulting intensity profiles (i.e. intensity ratings per category for each narrative) to test whether the 15 target categories could be identified from the emotion-wise intensity ratings, thus revealing whether or not the narratives elicited distinct categorical emotions. Also, to test the similarity of narratives belonging to the same category, we calculated the Euclidean distance between the intensity profiles of each pair of narratives.

Finally, participants rated the similarity of all possible pairs of the emotions induced during the experiment (a total of 105 pairs) using a direct comparison method. The participants were shown one pair of emotion words at the time and asked to rate how subjectively similar they experience the emotions (ranging from no similarity [0] to full similarity [1]). Based on the ratings, we extracted the recollected experiential similarity matrix for each participant individually and averaged these across all participants.

MRI data acquisition and preprocessing

MRI data were collected on a 3-T Siemens Magnetom Skyra scanner at the Advanced Magnetic Imaging center, Aalto University, using a 20-channel Siemens volume coil. Whole-brain functional scans were collected using a whole brain T2*-weighted EPI sequence with the following parameters: 33 axial slices, TR = 1.7 s, TE = 24 ms, flip angle = 70°, voxel size = 3.1 × 3.1 × 4.0 mm³, matrix size = 64 × 64 × 33, FOV = 256 × 256 mm. A custom-modified bipolar water excitation radio frequency (RF) pulse was used to avoid signal from fat. High-resolution anatomical images with isotropic 1 × 1 × 1 mm³ voxel size were collected using a T1-weighted MP-RAGE sequence.

Data were preprocessed using FSL 5.0 (Jenkinson et al., 2012). Motion correction was performed using MCFLIRT (Jenkinson

et al., 2002) and non-brain matter was removed using BET (Smith, 2002). High-pass temporal filtering was applied using Gaussian-weighted least-squares straight line fitting with sigma of 100 s. Participant-wise gray matter masks were generated by segmenting the T1-weighted images into gray and white matter and cerebrospinal fluid using the FAST segmentation tool (Zhang et al., 2001) and transforming the masks to the native space using FLIRT (Jenkinson et al., 2002) with nine degrees of freedom. The gray matter maps were subsequently thresholded using intensity threshold > 0.5 to create participant-specific masks. This threshold was chosen to include those voxels with a higher probability of belonging to the gray matter and, subsequently, the masks were visually inspected to make sure most gray matter was included. On average, the gray matter mask included 16 000 voxels.

For univariate general linear model (GLM) analysis, the preprocessed functional data were registered to 2-mm Montreal Neurological Institute (MNI) 152 standard space template using FLIRT (Jenkinson et al. 2002) and 9 degrees of freedom, and smoothed using a Gaussian kernel with FWHM 8.

Multivariate pattern classification within participants

The classification of emotion categories was performed using the Princeton multi-voxel pattern analysis (MVPA) toolbox (<https://pni.princeton.edu/pni-software-tools/mvpa-toolbox>) in Matlab 2012b using each participant's data in native space. A separate classifier was trained for each participant and, after all steps, the classification results were averaged across the participants. After preprocessing, voxels outside gray matter were masked out using the participant-specific gray matter masks and the functional data from each run were standardized to have a mean of zero and variance of one. Next, each participant's data were divided into training ($N-1$ runs) and testing sets (the remaining run). Feature selection was performed using one-way ANOVA (testing for the main effect of emotion category) for the training set to select the voxels with a significant ($P < 0.05$) main effect for emotion, i.e. to select the voxels whose mean activation differed between at least some of the 15 possible emotion conditions. The feature selection preserved on average (across cross-validation folds and participants) 41% of the voxels. Hemodynamic lag was corrected for by convolving the boxcar category regressors with the canonical double gamma HRF function and thresholding the convolved regressors using a sigmoid function to return the regressors to the binary form. The classification was performed on all the standardized, HRF-convolved fMRI volumes from the 10 s imagery period following the narrative (treating all single time points per category as samples for that category; median 6 volumes per one emotion category in one run) to extract only emotion-related brain activity, and to minimize activity related to the acoustic and semantic features of the stimuli. Thus, each of the eight runs included on average 90 TRs (5–6 TRs per category) that were used in the classification. A linear neural network classifier without hidden layers was trained to recognize the correct emotion category out of 15 possible ones (multiclass classification, see Polyn et al. 2005 for details). Naïve chance level, derived as a ratio of 1 over the number of categories, was 6.7%. A leave-one-run-out cross-validation was performed, thus, dividing the data into all possible $N-1$ run combinations and repeating the classification pipeline for each such cross-validation fold, and the participant-wise classification accuracy was calculated as an average percentage of correct guesses across all the cross-validation folds. To test whether

classification accuracy exceeded chance level, we used permutation tests to simulate the probability distribution of the classification by shuffling the category labels of the training set (across training set runs) and re-running the whole classification pipeline, repeated 1000 times for each participant. FDR correction at $P < 0.05$ was used for multiple comparisons.

Hierarchical clustering

We next investigated the similarities between emotions using the category confusions from the whole-brain classification. Clustering was performed for exploratory purposes to characterize the similarities in neural and recollected experiential data, this however does not provide statistical evidence for the similarity structures. Nevertheless, similarity between neural and experiential data would suggest similar representational segregation in the neural code and subjective feelings. From the group-averaged confusion matrix, we calculated a distance matrix by taking the category confusion vectors for each pair of emotions and by calculating the Euclidean distance between these vectors (see Reyes-Vargas et al., 2013). We then employed hierarchical cluster analysis in Matlab to investigate how different emotions cluster together based on their neural similarities. The agglomerative hierarchical cluster tree was calculated on the distance matrix using *linkage* function with 'complete' option (i.e. the furthest distance method). Finally, we constructed the clusters from the cluster tree (*cluster* function) and chose the solution that minimized the number of categories while keeping at least two emotions per category. Note that the clustering was selected solely for data visualization rather than for statistical inference. To visualize the similarities in subjective and neural organization of emotions, we extracted the clusters in both neural and recollected experiential data, and subsequently plotted the cluster solutions using alluvial diagrams (Rosvall and Bergstrom, 2010; www.mapequation.org).

We then investigated to which extent the neural similarities between different emotional states correspond to their recollected experiential (subjectively felt) differences. Experiential similarity matrices were calculated based on pairwise similarity ratings of emotions and averaged over the participants. Subsequently, the mean neural and experiential similarity matrices were correlated using Spearman's rank correlation coefficient. The P level for the Spearman test was obtained with a permutation test by shuffling the neural matrix and recalculating the correlation for 5000 times.

Finally, we tested whether basic and non-basic emotions generally differ in how categorical their experience is recognized. To do so, we calculated prototypical experience scores for each emotion, defined as the sum of off-diagonal elements in the experiential and neural similarity matrices separately. This analysis was based on the assumption that the off-diagonal elements represent confusions across emotions, thus indicating a lack of sharp categorical representation within a group of emotions. In other words, we tested whether emotion confusions were systematically different for basic and non-basic types, in both subjective and neural data. The resulting 'prototypicality' scores (i.e. off-diagonal scores) for basic and non-basic emotions were then compared using Mann-Whitney U test.

Regional effects in GLM

To investigate the overall effect of any emotion on brain activity, we first ran GLM to compare all emotions together against the neutral baseline, and then ran separate GLMs to compare

each of the 14 emotions against the neutral baseline. First level analysis was performed in SPM 12 (www.fil.ion.ucl.ac.uk/spm/) to obtain individual contrast maps and second level analysis was then performed with FSL *randomize* with the threshold free cluster enhancement option as it implements the least biased way to control for multiple comparisons (Eklund et al. 2016; $N = 5000$ permutations). Emotion-wise t maps were then qualitatively summarized across emotions as a cumulative map where each voxel shows the number of statistically significant emotions, at the cluster corrected level of $P < 0.05$.

Visualization of emotion clusters in the brain

Finally, to summarize and visualize where emotion-wise activation patterns were located, we mapped the three principal clusters obtained with hierarchical clustering on the cortical and subcortical maps using R, G, B color space. The last cluster containing surprise and neutral was plotted separately (see [Supplementary Figure S6](#)) given it contained neutral-like states only. For each emotion, we took the unthresholded second level t maps obtained from the GLM analysis, summed them for emotions belonging to the same cluster, and assigned the summed values to the corresponding R, G, B channels. The color channels were subsequently visualized in MNI space. Consequently, the RGB color at each voxel reflects the cluster distribution of that voxel, and can be used for localizing brain regions contributing to different emotions.

Results

Behavioral results

Behavioral ratings showed that the narratives successfully elicited reliable and strong target emotions (intensity profiles in Figure 1A; mean intensities per target category: pride 0.80, longing for 0.77, happiness 0.84, gratitude 0.84, love 0.90, surprise 0.76, neutral 0.88, disgust 0.85, sadness 0.96, fear 0.89, shame 0.75, anger 0.68, guilt 0.91, contempt 0.66, despair 0.87). In k -means clustering, the accuracy to assign a narrative to the correct target category based on its intensity profile was 97% (against the chance level 6.7%). Also, the narratives within each category had highly similar intensity profiles (Figure 1B); i.e. narratives belonging to the same category elicited similar emotions.

Classification of basic and non-basic emotions

Mean classification accuracy across the 14 emotions and the neutral state was 17% (against naïve chance level of 6.7%; 95th percentile of the permuted classification accuracy distribution was 8.4%). After correcting for multiple comparisons, the classification performance was above a permutation-based significance level for all emotions except shame and longing ($P < 0.05$, Figure 2; see [Supplementary Table S2](#) for effect sizes and [Supplementary Figure S1](#) for a confusion matrix). On average, basic emotions (anger, disgust, fear, happiness, sadness, surprise) could be classified more accurately than the non-basic emotions (26% vs 15%, respectively, $t(24) = 7.39$, $P < 0.0001$, Cohen's $h = 0.16$), while we found no significant differences in classifier accuracies between positive and negative emotions (17% vs 14%, respectively, $t(24) = 1.52$, $P = 0.075$). We also trained separate classifiers for each a priori selected region-of-interest (see [Supplementary Methods](#)) which showed that classification accuracies were above chance level in frontal ROIs, especially in frontal pole, and in somatomotor ROIs, especially for pre- and post-central gyri, yet did not exceed that of the whole-brain classification ([Supplementary Figure S2](#)).

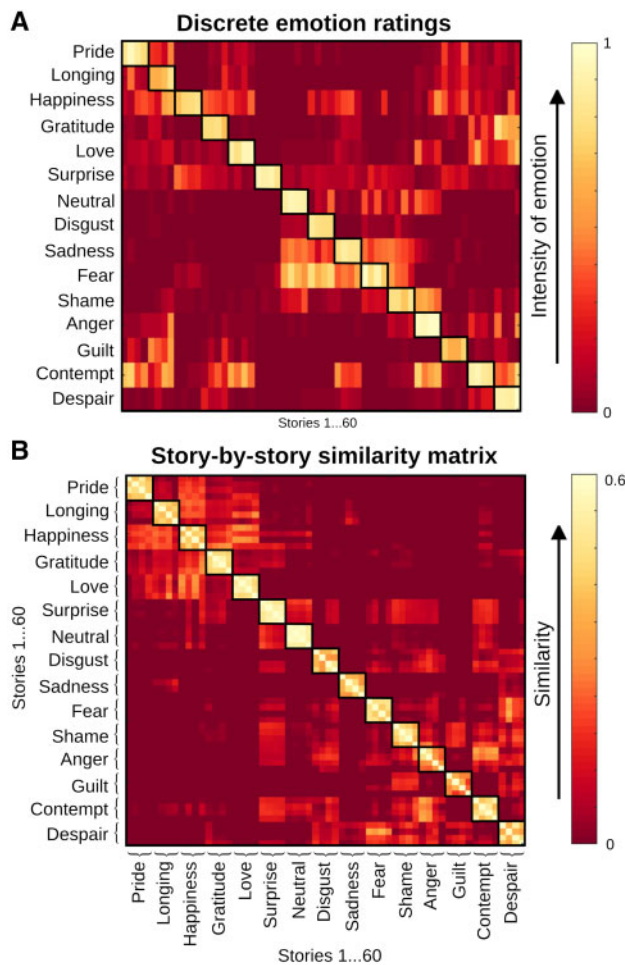


Fig. 1. The stimuli consisted of 60 brief (5–20 s) narratives that induced 14 emotional states and a neutral state. (A) Participants rated on a scale from 0 to 1 how strongly each emotion was elicited by the narrative. The coloring indexes the mean intensity for experiencing each emotion for each narrative. (B) Based on the emotion intensity ratings, we calculated the similarity of emotional experience between narratives by using Euclidean distances.

We also ran separate classifiers across positively and negatively valenced emotion categories to test whether classification *within* a similarly valenced superordinate category is comparable to that of classification with all categories. Average classifier accuracy for positive emotions was 30.5% (naïve chance level 20%) and all emotions could be classified with above-chance level accuracy: love (36.9%), happiness (31.5%), gratitude (30.8%), pride (28.5%), and longing (24.9%). Average classifier accuracy for negative emotions was 22.4% (naïve chance level 12.5%) and above chance level accuracy was observed for each emotion: sadness (28.8%), disgust (27.8%), despair (25.7%), fear (25.1%), anger (20.5%), contempt (19.0%), and shame (15.9%).

Similarity in neural basis corresponds to experienced similarity

Organization of recollected experiential similarity matrices derived from behavioral ratings was significantly associated with the neural similarity matrices derived from confusion matrices from whole-brain classification ($r=0.37$, $P=0.0048$; Figure 3A). Clustering of confusion matrices divided the emotions into four clusters (Figure 3A): (1) happiness, pride, gratitude, love, and longing; (2) surprise and neutral; (3) disgust, sadness, fear, and shame; and (4) anger, contempt, guilt, and despair, which mostly corresponded to the four main clusters of experiential similarities (Figure 3B): (1) happiness, pride, gratitude, and love; (2) surprise and neutral; (3) longing and sadness; and (4) disgust, anger, contempt, shame, guilt, fear, and despair.

To test whether basic and non-basic emotions generally differ in how prototypical their experience is, we calculated a prototypical experience score for each emotion (i.e. rate of confusions with other emotions) and then compared the average prototypicality scores for basic and non-basic emotions. There were no differences between basic and non-basic in either the neural or experiential data.

Affective space in the brain

To investigate the brain regions generally activated and deactivated by our emotion stimuli, we contrasted the brain activity

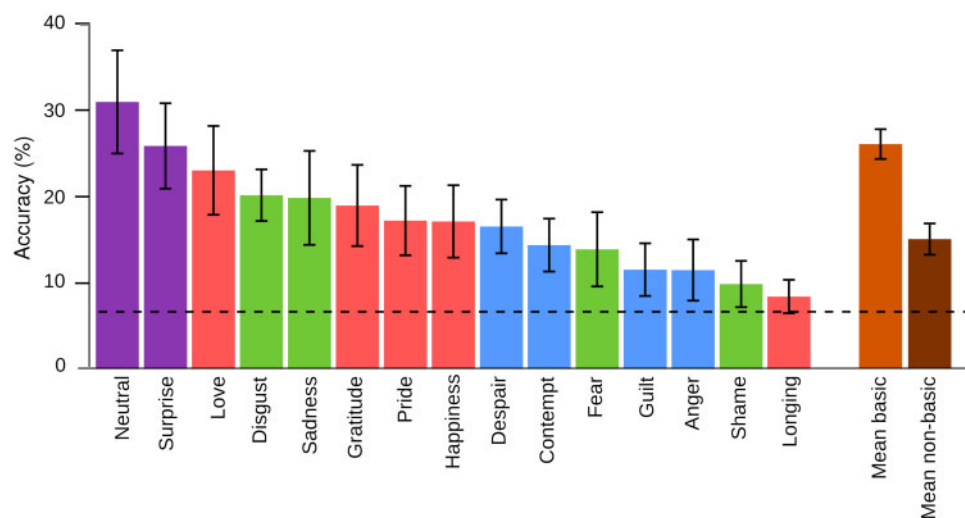


Fig. 2. Means and standard errors for emotion-wise whole-brain classification accuracy. Dashed line represents chance level (6.7%). Colors reflect the clusters formed on the basis of experienced similarity of emotions (see Figure 3).

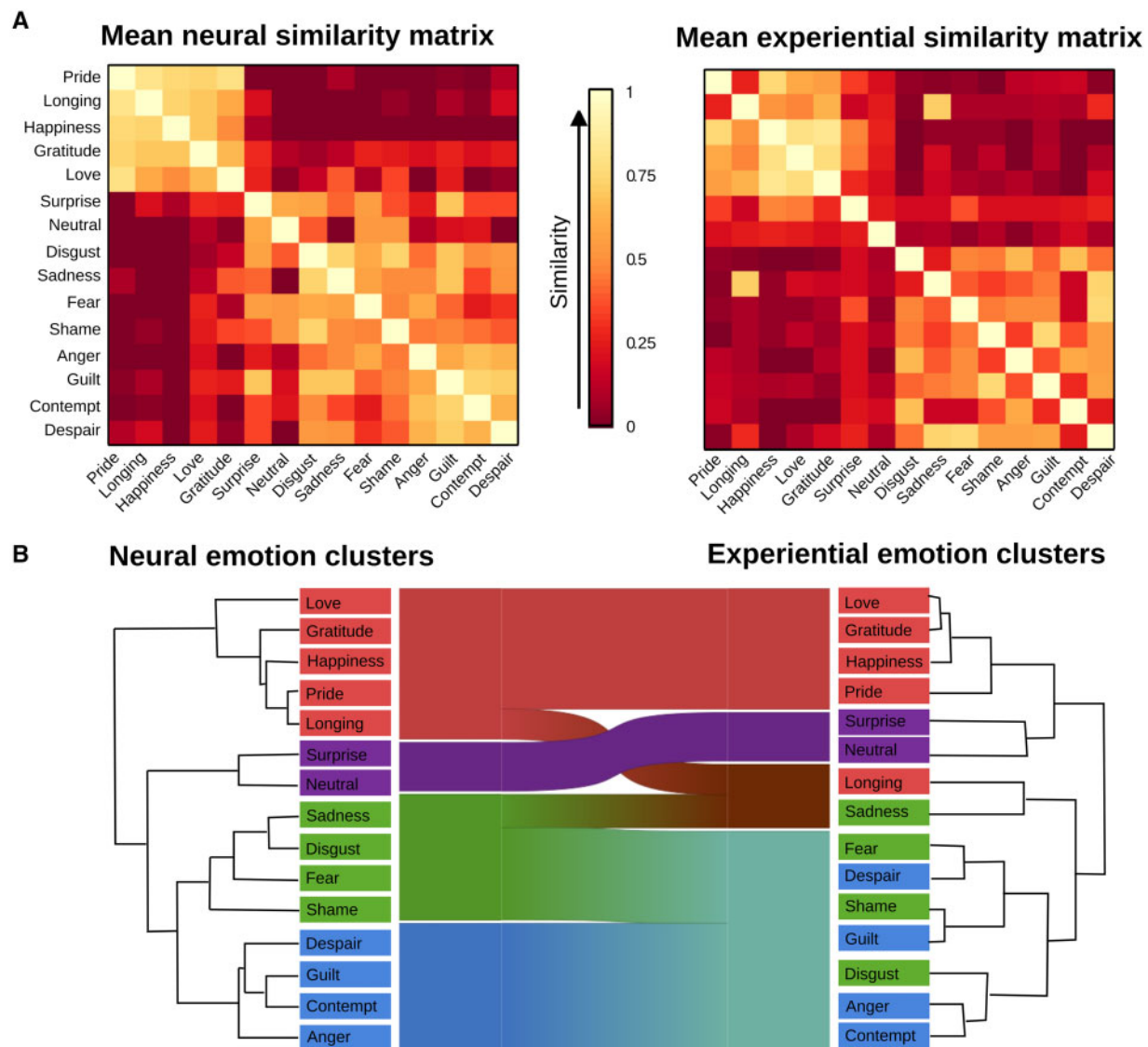


Fig. 3. (A) Left: Neural similarity matrix extracted from the classifier confusion matrix. The similarity matrix was created by calculating the Euclidean distance between each pair of emotions based on their category confusion vectors. Right: Experiential similarity matrix based on pairwise similarity ratings for emotions elicited by the narratives. (B) Alluvial diagram showing the similarity of hierarchical cluster structure of the experiential and neural similarities. Coloring of the emotion categories is based on the clusters in the neural similarity matrix.

related to all emotions with the neutral condition (Figure 4 outlines; see also [Supplementary Figure S4](#)). The areas activated by emotion in general included pre-motor cortex, thalamus, insula and putamen. Deactivations were observed in visual and auditory cortices, precuneus, PCC, right anterior PFC and right lateral parietal areas.

To reveal the brain regions contributing most consistently to different emotions, we constructed cumulative activation and deactivation maps of emotion-driven hemodynamic responses (Figure 4; corresponding effect size maps in [Supplementary Figure S4](#); the emotion-wise statistical *t* maps are available in <http://neurovault.org/collections/TWZSVODU>). These maps reveal how many of the 14 possible emotional states activated each voxel. The cumulative activation map (Figure 4A) showed that most emotions involved activation of midline regions including anterior cingulate cortex (ACC) and precuneus, as well as subcortical regions including brain stem and hippocampus, motor areas including cerebellum, and visual cortex. The

cumulative deactivation map (Figure 4B) showed that most emotions involved deactivation of auditory cortex, frontal areas including superior frontal gyri, right middle frontal gyrus, and left inferior frontal gyrus and parietal areas including supramarginal gyrus.

Finally, to visualize where specific emotions are encoded in the brain, we mapped the clusters resulting from the hierarchical clustering on cortical and subcortical surfaces (Figure 5). All emotions activated areas in the visual cortex, ACC, right temporal pole, supplementary motor area and subcortical regions. In addition, positive emotions belonging to Cluster 1 (happiness, pride, gratitude, love, longing) were more prominent in anterior frontal areas including vmPFC. Negative basic emotions from Cluster 2 (disgust, sadness, fear, shame) activated especially insula, supplementary motor area and specific parts of subcortical structures. Negative social emotions belonging to Cluster 3 (anger, contempt, guilt, despair; see [Supplementary Table S1](#) for the story stimuli targeting different

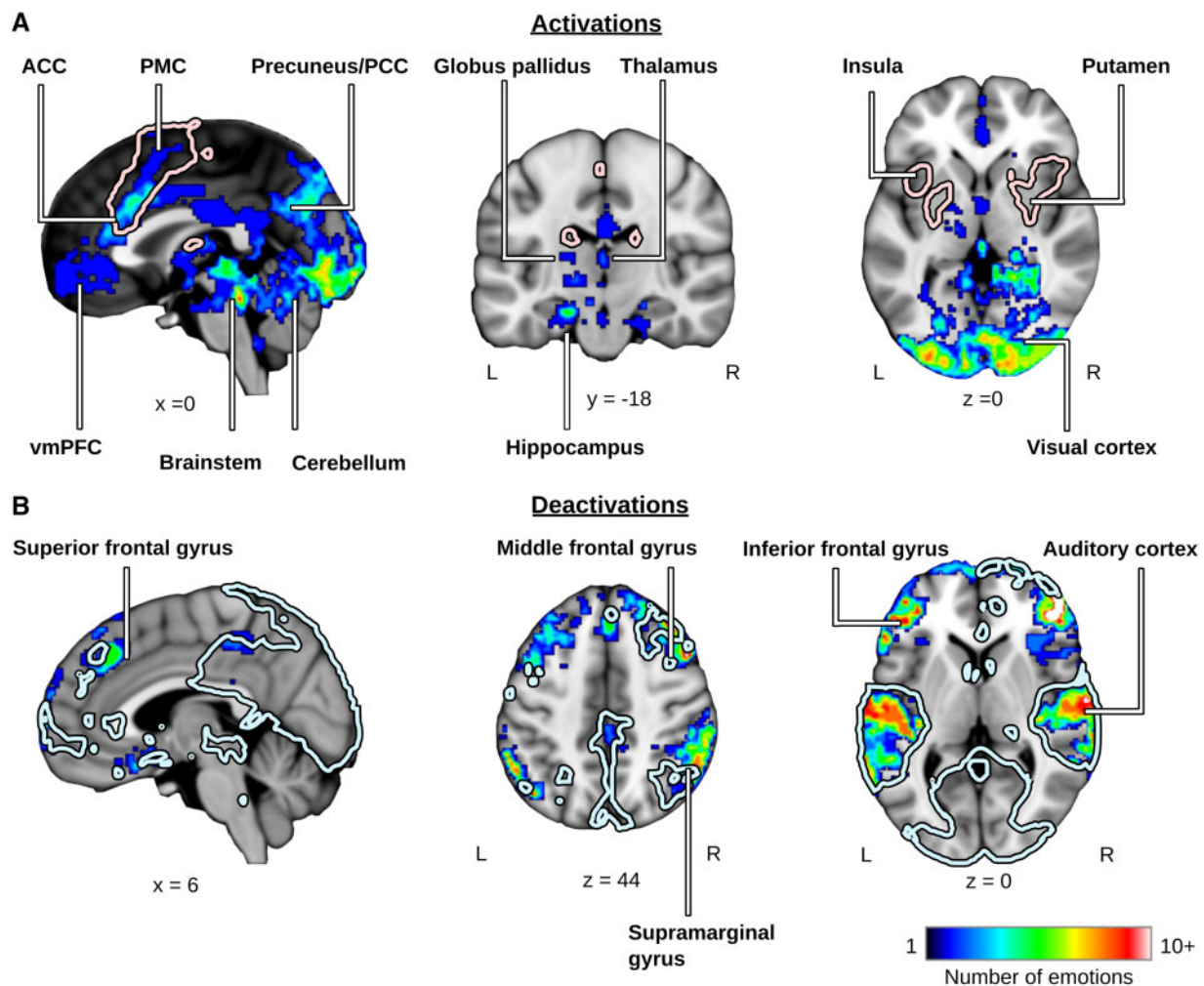


Fig. 4. (A) Cumulative activation map showing the cumulative sum of binarized t maps ($P < 0.05$, cluster-corrected) across each emotion vs neutral condition. Outline shows the GLM results for all emotions contrasted against the neutral condition ($P < 0.05$, cluster-corrected). (B) Cumulative deactivation map showing the cumulative sum of binarized t maps ($P < 0.05$, cluster-corrected) across neutral vs each emotion. Outline shows the GLM results for the neutral condition contrasted against all emotions ($P < 0.05$, cluster-corrected).

emotions) were most prominent in left insula and the adjacent frontal areas. Finally, surprise (Cluster 4, [Supplementary Figure S6](#)) activated especially parts of auditory cortex, supplementary motor areas, and left insula.

Discussion

Our results reveal that multiple emotion states have distinct and distributed neural bases, as evidenced by the above chance-level classifier performance for all emotions, except longing and shame. Together with the spatial location of emotion-dependent brain activation, this suggests that a multitude of different emotions are represented in the brain in a distinguishable manner, yet in partly overlapping regions: each emotion state likely modulates different functional systems of the brain differently, as shown by distinct patterns measured with BOLD-fMRI, and the overall configuration of the regional activation patterns defines the resulting emotion. While a set of 'core' emotion processing areas in cortical midline regions, motor areas, sensory areas, and subcortical regions are engaged during practically all emotions, the relative engagement of these areas varies between emotions. This unique neural signature of each emotion might relate to the corresponding subjective feeling, as

evidenced by the correspondence between neural and recollected experiential similarity between emotions.

Different emotions are characterized by distinct neural signatures

Altogether 12 emotions (excluding longing for and shame) out of the 14 included in the study could be reliably classified from the fMRI signals. Our results extend previous studies, which have shown classification of specific emotional states usually focusing on the basic emotions only or a subset of these (Ethofer et al., 2009; Peelen et al., 2010; Said et al., 2010; Kotz et al., 2013; for a review, see Kragel and LaBar, 2014). While the 'classic' basic emotions have attracted most attention in psychological and neurophysiological studies, they constitute only a small portion of the emotions humans universally experience (Edelstein and Shaver, 2007). Furthermore, accumulating behavioral evidence suggests that other, non-basic emotions are also characterized by distinctive features in facial expressions (Baron-Cohen et al., 2001; Shaw et al., 2005), bodily changes (Nummenmaa et al., 2014a), and physiology (Kreibitz, 2010; Kragel and LaBar, 2013). The present data corroborate these findings by showing that also emotions not considered as 'basic'

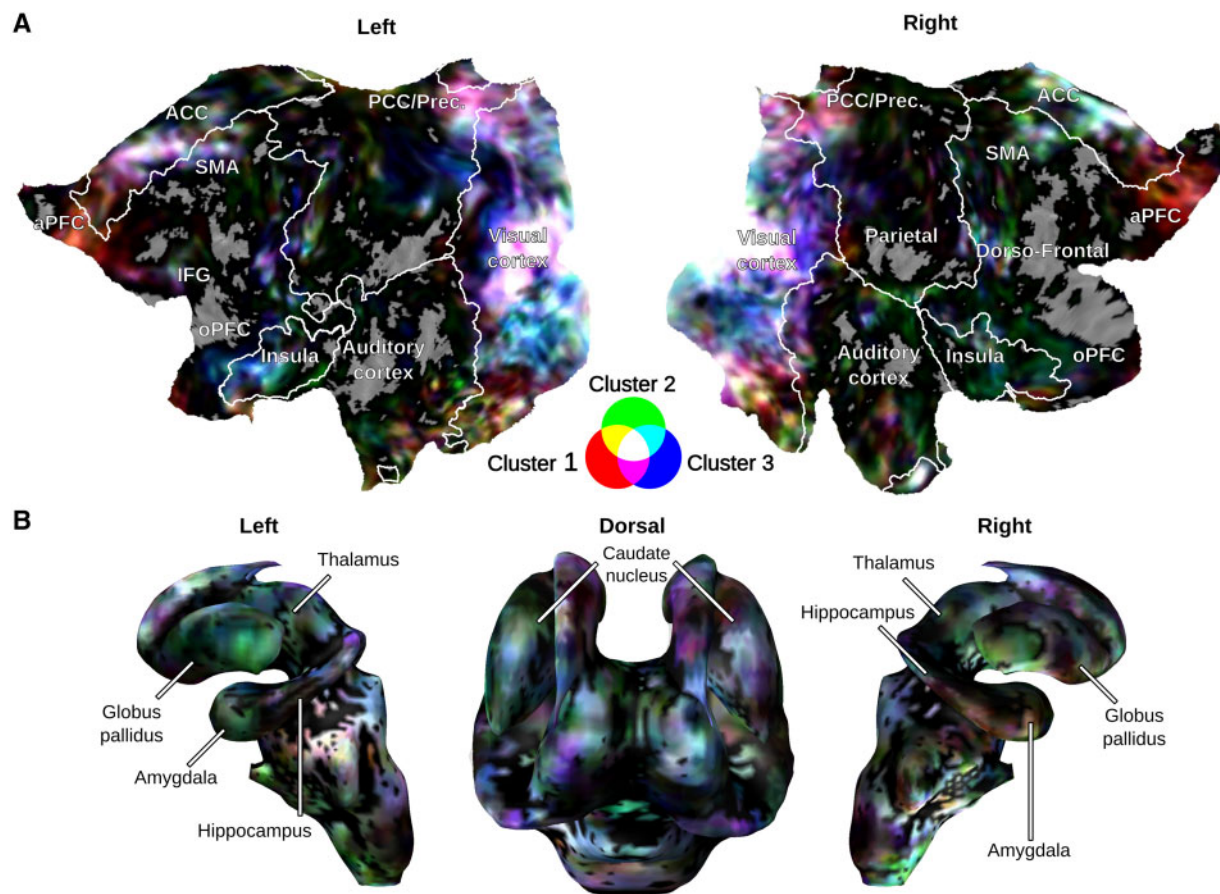


Fig. 5. Activation maps showing the summed uncorrected *t* maps for each cluster obtained from the hierarchical clustering analysis in (A) cortical regions and (B) subcortical regions. Colors represent the three clusters: positive (red), negative basic (green), and negative social (blue) emotions.

may each have distinctive brain activation patterns. The local brain activity patterns underlying different emotions are most probably to some extent variable across participants and reflect individual responses, as brains are always intrinsically shaped by individual development and experiences.

If we consider discrete emotion systems as wide-spread, distinct neural activation distinct to each emotion state, successful pattern classification of brain states across emotions would provide support for separate neural systems for each emotion. In turn, constructivist emotion theories suggest that all emotions are generated by a shared set of fundamental functional systems which are however not specific to emotional processing *per se* (Kober *et al.*, 2008; Lindquist *et al.*, 2012). The present data show that different emotions are associated with granular activation changes across multiple functional systems, and their spatially distributed configuration ultimately defines the specific emotion at both psychological and behavioral levels (Meaux and Vuilleumier, 2015). For instance, two emotions might share their somatosensory representations, but underlying interoceptive representations could be different. Thus, the general configuration of the central and peripheral nervous system leads to distinct emotion states.

However, we stress that the current data cannot resolve whether these functional signatures of distinct emotions would be necessary for each emotion, or whether the presently observed structure of the emotions is optimal, as the classification solution is contingent on the employed *a priori* category labels. Further, it must be noted classification analysis does not

readily reveal the actual neural organization of each emotion system, as the pattern classification only tells us that, on average and at the level measurable with BOLD-fMRI, the cortical/neuronal pattern underlying each category differ enough to be separated, whereas localizing the actual source of differences is more difficult. Therefore, we have complemented the pattern classification analysis with visualization of different emotion categories using GLM and clustering. Furthermore, the pattern recognition techniques employed in this study cannot provide causal evidence for the existence of basic emotion systems: even if we find emotion-specific activation patterns for specific emotions, it does not prove that these patterns are strictly necessary for the corresponding emotional state (Nummenmaa and Saarimäki, 2017). The causality issues can be resolved, for instance, in lesion studies or with brain stimulation techniques. Moreover, the current analyses do not directly answer whether the data are better described by discrete versus dimensional models of emotion.

If basic emotions were somehow 'special' or different from non-basic emotions at the neural level, we should observe (1) distinct neural activation patterns for basic emotions but not for non-basic emotions, or (2) different (or perhaps additional) neural systems underlying basic and non-basic emotions. Our classification results and cumulative maps show that both basic and non-basic emotions could be classified accurately, and they elicited activation in largely overlapping brain areas. On average, classification accuracies were higher for basic emotions than for non-basic emotions. This suggests that while there

exists partially exclusive neural codes for a multitude of emotion states, as those investigated here, the canonical basic emotions are more discrete than the complex social emotions included in this study, suggesting that there are both universal and experience-dependent components to emotions. Also, classification accuracies within single regions showed that only basic emotions could be distinguished in some areas, including somatomotor regions (insula and supplementary motor area), midline areas (PCC and ACC), and inferior frontal gyrus (Supplementary Figure S3). Another potential explanation for the differences in classification accuracies between basic and non-basic emotions is that maybe there is a clearer cultural understanding or prototypical experience related with basic emotions, manifested as clearer patterns underlying these emotions. To test this hypothesis, we calculated prototypical experience scores for each emotion, defined as the sum of off-diagonal elements in recollected experiential and neural similarity matrices separately, and compared the average prototypicality scores for basic and non-basic emotions. This allowed us to quantify the specificity vs confusion of recognition for different emotion categories. There were no differences between basic and non-basic in either the neural or recollected experiential data, suggesting that both basic and non-basic emotions have equally distinct neural and experiential underpinnings.

Correspondence between neural and phenomenological organization of emotions

An explorative hierarchical clustering of emotion-specific neural patterns identified four clusters in the neural data that correspond to positive emotions (pride, longing, happiness, gratitude, love), negative basic emotions (disgust, sadness, fear, shame) and negative social emotions (anger, guilt, contempt, despair) and surprise (Figure 3). Clustering is an exploratory and descriptive technique that does not allow strong inferences about underlying causal structure. Rather, clustering of neural similarities was used to reduce the dimensionality of the data for visualization purposes. Comparison of subjectively experienced similarity of emotions and the similarity of the neural patterns suggested a direct link between the whole-brain neural signatures of emotions and the corresponding subjective feelings: the more similar neural signatures two emotions had, the more similar they were experienced. This accords with prior work suggesting that emotion-specific neural activation patterns might explain why each emotion feels subjectively different (Damasio et al., 2000; Saarimäki et al., 2016). Emotions might constitute distinct activity patterns in regions processing different emotion-related information, such as somatosensory (bodily sensations), motor (actions), as well as brainstem and thalamocortical loops (physiological arousal). Activation from these areas is then integrated in the cortical midline, such integration then giving rise to the interpretation of the subjective feeling (Northoff and Bermpohl, 2004; Northoff et al., 2006). Thus, a subjective feeling of a specific emotion stems from the net activation of different sub-processes, rather than solely on the basis of any single component of emotion.

Organization of the affective space in the brain

Our hierarchical clustering analysis and cumulative mapping reveal how different patterns of activity may give rise to different emotions. First, midline regions including ACC, PCC, and precuneus were activated during most emotions. These regions might code emotional valence (Colibazzi et al., 2010; Chikazoe

et al., 2014), participate in self-relevant and introspective processing (Northoff and Bermpohl, 2004; Northoff et al., 2006), and integrate information of internal, mental, and bodily states (Northoff et al., 2006; Buckner and Carroll, 2007). We also found consistent emotion-dependent activity in the brainstem, including periaqueductal grey, pons and medulla, for almost all emotions. This activation might reflect the control of autonomic nervous system's reactions to different emotions (Critchley et al., 2005; Linnman et al., 2012) and/or covert activation of particular motor programs (Blakemore et al., 2016). Also, subcortical regions including amygdala and thalamus showed distinct activity patterns that differed between clusters. Both of these regions are related to salience processing and emotional arousal (Anders et al., 2004; Adolphs, 2010; Damasio and Carvalho, 2013; Kragel and LaBar, 2014) and show specific activation patterns for basic emotions (Wang et al., 2014), findings that we now extend also to non-basic emotions.

Second, somatomotor areas including premotor cortex, cerebellum (including vermis and the anterior lobe), globus pallidus, caudate nucleus, and posterior insula were activated during most emotions, but according to the cluster visualizations especially during the processing of emotions that have a strong impact on action tendencies and avoidance-oriented behaviors (fear, disgust, sadness, shame, surprise; Frijda et al. 1989). These areas are engaged during emotion perception (Nummenmaa et al., 2008, 2012; Pichon et al., 2008), emotion regulation (Schutter and van Honk, 2009), and somatomotor processing and action preparation related to emotional processing (Kohler et al., 2002; Wicker et al., 2003; Mazzola et al., 2013).

Third, anterior prefrontal cortex was activated especially during positive emotions (happiness, love, pride, gratitude, longing) according with previous research linking anterior prefrontal cortex with positive affect (Bartels and Zeki, 2004; Zahn et al., 2009; Vytal and Hamann, 2010). Fourth, negative emotions such as guilt, contempt, anger, and despair clustered together, potentially reflecting their social dimension and interpersonal aspects or their self-conscious nature. Especially, left hemisphere activation in orbitofrontal cortex connected to rewards and punishments (Kringelbach and Rolls, 2004), as well as in inferior frontal cortex and dorsolateral prefrontal cortex, which subserve language and executive control (Poldrack et al., 1999; Kane and Engle, 2002), and in anterior insula linked to processing of social emotions (Lamm and Singer, 2010) was activated during these emotions. Fifth, surprise did not resemble any of the other emotions included in this study, but was instead closest to the neutral state. This is in line with previous research showing that surprise tends to separate from other emotions in subjective ratings (Toivonen et al., 2012).

Finally, we also found decreased activation in auditory areas and increased activation in visual areas during the imagery of all emotion categories, likely reflecting the offset of the auditory stimulation followed by mental imagery of the emotion-evoking situation (Ganis et al., 2004) and emotion-related modulation of this activity (Holmes and Mathews, 2005; Nummenmaa et al., 2012; Kassam et al., 2013).

Limitations

Despite the careful selection of the emotional stimuli and k-means clustering suggesting clear categorical structure in the evoked affect (see Figure 1), it is possible that the narratives did not fully capture the target emotion only, and might have elicited also a mixture of emotions. Yet these may (1) arise in different time points during the narratives and (2) be not as strong

as the main target emotions (see Figure 1), thus average trial-wise activations most likely pertain to the target emotion. Despite this, the observed MVPA pattern may reflect whether each narrative is dominated by one emotion, or at least show the weighted influence of each emotion on a voxel activity. Therefore, the successful classification *per se* shows that at least the target emotions were successfully elicited, yet, better classification accuracies could potentially be reached if stimuli could target more selectively one category at the time.

Emotions were elicited using narrative-guided emotional imagery. The narrative stimuli across different categories of emotion differed along dimensions that are not of interest (for instance, different words or different kinds of imagery prompted), thus creating variation both within and between emotion categories, though the latter is unlikely to systematically correlate with particular emotions. Moreover, any within-category variation would probably work against the classifier by lowering the classification accuracy. Despite this, we observed above chance level accuracy for all emotion categories except longing and shame. However, it is possible that there were differences between categories in, for instance, evoked emotional imagery, that vary across emotion categories in some uncontrolled manner, thus potentially in part affecting the classification accuracy. This may constitute a limitation of our study but is also inherently related to variations in scenarios and appraisal dimensions that constitute distinct emotion types.

Conclusions

Our results characterize the distinct and distributed neural signatures of multiple emotional states. Different emotions result from differential activation patterns within a shared neural circuitry, mostly consisting of midline regions, motor areas and subcortical regions. The more similar the neural underpinnings of these emotions, the more similarly they also are experienced. We suggest that the relative engagement of different parts of this system defines the current emotional state.

Supplementary data

Supplementary data are available at SCAN online.

Acknowledgements

We acknowledge the computational resources provided by the Aalto Science-IT project. We thank Marita Kattelus for her help in fMRI data collection and Matthew Hudson for a language check of the manuscript.

Funding

This study was supported by the aivoAALTO project of the Aalto University, Academy of Finland (#265917 and #294897 to L.N., and #138145 and #276643 to I.P.J.), ERC Starting Grant (#313000 to L.N.); Finnish Cultural Foundation (#00140220 to H.S.); and the Swiss National Science Foundation National Center of Competence in Research for Affective Sciences (#51NF40-104897 to P.V.).

Conflict of interest. None declared.

References

- Adolphs, R. (2002a). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, 1, 21–62.
- Adolphs, R. (2002b). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, 12(2), 169–77.
- Adolphs, R. (2010). What does the amygdala contribute to social cognition?. *Annals of the New York Academy of Sciences*, 1191, 42–61.
- Adolphs, R. (2017). How should neuroscience study emotions? by distinguishing emotion states, concepts, and experiences. *Social Cognitive and Affective Neuroscience*, 12, 24–31.
- Adolphs, R., Damasio, H., Tranel, D., Cooper, G., Damasio, A.R. (2000). A role for somato-sensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *Journal of Neuroscience*, 20, 2683–90.
- Anders, S., Lotze, M., Erb, M., Grodd, W., Birbaumer, N. (2004). Brain activity underlying emotional valence and arousal: a response-related fMRI study. *Human Brain Mapping*, 23(4), 200–9.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I. (2001). The “Reading the Mind in the Eyes” test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42(2), 241–51.
- Barrett, L.F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12(11), 1833–23.
- Bartels, A., Zeki, S. (2004). The neural correlates of maternal and romantic love. *Neuroimage*, 21(3), 1155–66.
- Blakemore, R.L., Rieger, S.W., Vuilleumier, P. (2016). Negative emotions facilitate isometric force through activation of prefrontal cortex and periaqueductal gray. *Neuroimage*, 124(Pt A), 627–40.
- Buckner, R.L., Carroll, D.C. (2007). Self-projection and the brain. *Trends in Cognitive Science*, 11(2), 49–57.
- Chandrasekhar, P.V., Capra, C.M., Moore, S., Noussair, C., Berns, G.S. (2008). Neurobiological regret and rejoice functions for aversive outcomes. *Neuroimage*, 39(3), 1472–84.
- 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.
- Colibazzi, T., Posner, J., Wang, Z., et al. (2010). Neural systems subserving valence and arousal during the experience of induced emotions. *Emotion*, 10(3), 377–89.
- Coricelli, G., Critchley, H.D., Joffily, M., O’Doherty, J.P., Sirigu, A., Dolan, R.J. (2005). Regret and its avoidance: a neuroimaging study of choice behavior. *Nature Neuroscience*, 8(9), 1255–62.
- Costa, V.D., Lang, P.J., Sabatinelli, D., Versace, F., Bradley, M.M. (2010). Emotional imagery: assessing pleasure and arousal in the brain’s reward circuitry. *Human Brain Mapping*, 31(9), 1446–57.
- Critchley, H.D., Rotshtein, P., Nagai, Y., O’Doherty, J., Mathias, C.J., Dolan, R.J. (2005). Activity in the human brain predicting differential heart rate responses to emotional facial expressions. *Neuroimage*, 24(3), 751–62.
- Cunningham, W.A. (2013). Introduction to special section: psychological constructivism. *Emotion Review*, 5(4), 333–4.
- Damasio, A.R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. San Diego: Houghton Mifflin Harcourt.
- Damasio, A.R., Grabowski, T.J., Bechara, A., et al. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience*, 3(10), 1049–56.

- Damasio, A.R., Carvalho, G.B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nature Reviews Neuroscience*, *14*(2), 143–52.
- 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. *Proceedings of the National Academy of Sciences of the USA*, *101*(47), 16701–6.
- Edelstein, R.S., Shaver, P.R. (2007). A cross-cultural examination of lexical studies of self-conscious emotions. In: Tracy, J.L., Robins, R.W., Tangney, J.P., editors. *The Self-Conscious Emotions: Theory and Research*. New York: Guilford Press. 194–208.
- Eklund, A., Nichols, T.E., Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the USA*, *113*(28), 7900–5.
- Ekman, P. (1992). An argument for basic emotions. *Cognition Emotion*, *6*(3–4), 169–200.
- Ekman, P. (1999). Facial expressions. In: Dalgleish, T., Power, M., editors. *Handbook of Cognition and Emotion*. New York: John Wiley & Sons Ltd. 301–20.
- Ekman, P., Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364–70.
- 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*(3), 2481–91.
- Ethofer, T., Van De Ville, D., Scherer, K., Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology*, *19*(12), 1028–33.
- Fischer, A., LaFrance, M. (2015). What drives the smile and the tear: why women are more emotionally expressive than men. *Emotion Review*, *7*(1), 22–9.
- Frijda, N.H., Kuipers, P., Ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, *57*(2), 212–28.
- Ganis, G., Thompson, W.L., Kosslyn, S.M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, *20*(2), 226–41.
- Grossman, M., Wood, W. (1993). Sex differences in intensity of emotional experience: a social role interpretation. *Journal of Personality and Social Psychology*, *65*(5), 1010–22.
- Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends in Cognitive Science*, *16*(9), 458–66.
- Hofer, A., Siedentopf, C.M., Ischebeck, A., et al. (2006). Gender differences in regional cerebral activity during the perception of emotion: a functional MRI study. *Neuroimage*, *32*(2), 854–62.
- Holmes, E., Mathews, A. (2005). Mental imagery and emotion: a special relationship? *Emotion*, *5*(4), 489–97.
- Jenkinson, M., Bannister, P.R., Brady, J.M., Smith, S.M. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825–41.
- Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M. (2012). FSL. *Neuroimage*, *62*(2), 782–90.
- Kane, M.J., Engle, R.W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychon B Review*, *9*(4), 637–71.
- 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.
- Kober, H., Barrett, L.F., Joseph, J., Bliss-Moreau, E., Lindquist, K., Wager, T.D. (2008). Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *Neuroimage*, *42*(2), 998–1031.
- Kohler, E., Keysers, C., Umiltà, M.A., Fogassi, L., Gallese, V., Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, *297*(5582), 846–8.
- Kotz, S.A., Kalberlah, C., Bahlmann, J., Friederici, A.D., Haynes, J.D. (2013). Predicting vocal emotion expressions from the human brain. *Human Brain Mapping*, *34*(8), 1971–81.
- 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), 160–74.
- Kragel, P.A., LaBar, K.S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience*, *10*(11), 1437–48.
- Kragel, P.A., LaBar, K.S. (2016). Decoding the nature of emotion in the brain. *Trends in Cognitive Science*, *20*(6), 444–55.
- Kreibig, S.D. (2010). Autonomic nervous system activity in emotion: a review. *Biological Psychology*, *84*(3), 394–421.
- Kringelbach, M.L., Rolls, E.T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Progress in Neurobiology*, *72*(5), 341–72.
- Lamm, C., Singer, T. (2010). The role of anterior insular cortex in social emotions. *Brain Structure and Function*, *214*(5–6), 579–91.
- Lewis, M.D., Liu, Z.X. (2011). Three time scales of neural self-organization underlying basic and nonbasic emotions. *Emotion Review*, *3*(4), 416–23.
- 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 Science*, *35*(3), 121–43.
- Linnman, C., Moulton, E.A., Barmettler, G., Becerra, L., Borsook, D. (2012). Neuroimaging of the periaqueductal gray: state of the field. *Neuroimage*, *60*(1), 505–22.
- Mazzola, V., Vuilleumier, P., Latorre, V., et al. (2013). Effects of emotional contexts on cerebello-thalamo-cortical activity during action observation. *PLoS ONE*, *8*(9), e75912.
- Meaux, E., Vuilleumier, P. 2015. Emotion perception and elicitation. In: Toga, A.W., editor. *Brain Mapping: An Encyclopedic Reference*. Oxford (UK), Elsevier.
- Northoff, G., Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Science*, *8*(3), 102–7.
- Northoff, G., Heinzel, 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*(1), 440–57.
- Nummenmaa, L., Hirvonen, J., Parkkola, R., Hietanen, J.K. (2008). Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *Neuroimage*, *43*(3), 571–80.
- Nummenmaa, L., Glerean, E., Viinikainen, M., Jaaskelainen, I.P., Hari, R., Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences of USA*, *109*(24), 9599–604.
- Nummenmaa, L., Glerean, E., Hari, R., Hietanen, J.K. (2014a). Bodily maps of emotions. *Proceedings of the National Academy of Sciences of the USA*, *111*(2), 646–51.
- Nummenmaa, L., Saarimäki, H., Glerean, E., Gotsopoulos, A., Hari, R., Sams, M. (2014b). Emotional speech synchronizes

- brains across listeners and engages large-scale dynamic brain networks. *Neuroimage*, **102**, 498–509.
- Nummenmaa, L., Saarimäki, H. (2017). Emotions as discrete patterns of systemic activity. *Neuroscience Letters*, doi: 10.1016/j.neulet.2017.07.012.
- Panksepp, J. (1982). Toward a general psychobiological theory of emotions. *Behavioral and Brain Science*, **5**(03), 407–22.
- Panksepp, J., Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion Review*, **3**(4), 387–96.
- Peelen, M.V., Atkinson, A.P., Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience*, **30**(30), 10127–34.
- Pichon, S., de Gelder, B., Grezes, J. (2008). Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Society for Neuroscience*, **3**(3–4), 199–212.
- Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., Gabrieli, J.D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage*, **10**(1), 15–35.
- Polyn, S.M., Natu, V.S., Cohen, J.D., Norman, K.A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, **310**(5756), 1963–6.
- Reyes-Vargas, M., Sánchez-Gutiérrez, M., Fuginer, L., et al. (2013). Hierarchical clustering and classification of emotions in human speech using confusion matrices. In: International Conference on Speech and Computer. Springer International Publishing. 162–9.
- Rosvall, M., Bergstrom, C.T. (2010). Mapping change in large networks. *PLoS One*, **5**(1), e8694.
- Russell, J.A. (2003). Core affect and the psychological construction of emotions. *Psychological Review*, **110**(1), 145–72.
- Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I.P., et al. (2016). Discrete neural signatures of basic emotions. *Cerebral Cortex*, **26**(6), 2563–73.
- 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*, **1**, 11.
- Schutter, D.J.L.G., van Honk, J. (2009). The cerebellum in emotion regulation: a repetitive transcranial magnetic stimulation study. *Cerebellum*, **8**(1), 28–34.
- Shaw, P., Bramham, J., Lawrence, E.J., Morris, R., Baron-Cohen, S., David, S. (2005). Differential effects of lesions of the amygdala and prefrontal cortex on recognizing facial expressions of complex emotions. *Journal of Cognitive Neuroscience*, **17**(9), 1410–9.
- Simon-Thomas, E.R., Godzik, J., Castle, E., et al. (2012). An fMRI study of caring vs self-focus during induced compassion and pride. *Social Cognitive and Affective Neuroscience*, **7**(6), 635–48.
- Smith, S.M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, **17**(3), 143–55.
- Takahashi, H., Matsuura, M., Koeda, M., et al. (2008). Brain activations during judgments of positive self-conscious emotion and positive basic emotion: pride and joy. *Cerebral Cortex*, **18**(4), 898–903.
- Toivonen, R., Kivelä, M., Saramäki, J., Viinikainen, M., Vanhatalo, M., Sams, M. (2012). Networks of emotion concepts. *PLoS One*, **7**(1), e28883.
- Trost, W., Ethofer, T., Zentner, M., Vuilleumier, P. (2012). Mapping aesthetic musical emotions in the brain. *Cerebral Cortex*, **22**(12), 2769–83.
- Vuilleumier, P., Trost, W. (2015). Music and emotions: from enchantment to entrainment. *Annals of the New York Academy of Sciences*, **1337**, 212–22.
- 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.
- Wagner, U., N'Diaye, K., Ethofer, T., Vuilleumier, P. (2011). Guilt-specific processing in the prefrontal cortex. *Cerebral Cortex*, **21**(11), 2461–70.
- Wang, S., Tudusciuc, O., Mamelak, A.N., Ross, I.B., Adolphs, R., Rutishauser, U. (2014). Neurons in the human amygdala selective for perceived emotion. *Proceedings of the National Academy of Sciences of the USA*, **111**(30), E3110–9.
- Wicker, B., Keysers, C., Plailly, J., Royet, J.P., Gallese, V., Rizzolatti, G. (2003). Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron*, **40**(3), 655–64.
- Zahn, R., Moll, J., Paiva, M., et al. (2009). The neural basis of human social values: evidence from functional MRI. *Cerebral Cortex*, **19**(2), 276–83.
- 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 Transactions of the Medical Imaging*, **20**(1), 45–57.