

Abstract Representations of Associated Emotions in the Human Brain

Junsuk Kim,¹ Johannes Schultz,^{2,3} Tim Rohe,² Christian Wallraven,¹ Seong-Whan Lee,¹ and Heinrich H. Bühlhoff^{1,2}

¹Department of Brain and Cognitive Engineering, Korea University, 136-713 Seoul, Republic of Korea, ²Department of Human Perception, Cognition and Action, Max-Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany, and ³Department of Psychology, Durham University, DH1 3LE Durham, United Kingdom

Emotions can be aroused by various kinds of stimulus modalities. Recent neuroimaging studies indicate that several brain regions represent emotions at an abstract level, i.e., independently from the sensory cues from which they are perceived (e.g., face, body, or voice stimuli). If emotions are indeed represented at such an abstract level, then these abstract representations should also be activated by the memory of an emotional event. We tested this hypothesis by asking human participants to learn associations between emotional stimuli (videos of faces or bodies) and non-emotional stimuli (fractals). After successful learning, fMRI signals were recorded during the presentations of emotional stimuli and emotion-associated fractals. We tested whether emotions could be decoded from fMRI signals evoked by the fractal stimuli using a classifier trained on the responses to the emotional stimuli (and vice versa). This was implemented as a whole-brain searchlight, multivoxel activation pattern analysis, which revealed successful emotion decoding in four brain regions: posterior cingulate cortex (PCC), precuneus, MPFC, and angular gyrus. The same analysis run only on responses to emotional stimuli revealed clusters in PCC, precuneus, and MPFC. Multidimensional scaling analysis of the activation patterns revealed clear clustering of responses by emotion across stimulus types. Our results suggest that PCC, precuneus, and MPFC contain representations of emotions that can be evoked by stimuli that carry emotional information themselves or by stimuli that evoke memories of emotional stimuli, while angular gyrus is more likely to take part in emotional memory retrieval.

Key words: abstract representation; associated emotion; fMRI; MVPA

Introduction

Emotional states in humans can be elicited in various ways. Stimuli in different sensory modalities such as an emotional face or voice can arouse emotions, but emotions can also be triggered by the memory of a past emotional event, for example, a song reminiscent of a romantic episode. Different routes to elicit very similar emotions suggest that the brain hosts abstract or “supramodal” representations of emotions, and considerable effort has been devoted to identify representations that are invariant to the modality by which the emotions are conveyed (Peelen et al., 2010; Aubé et al., 2015). In particular, the fMRI study by Peelen et al. (2010) found strong evidence of supramodal emotional processing using a searchlight multivoxel pattern analysis (MVPA). They compared neural activity patterns across three different perceptual modalities: body movements, facial expressions, and emotional

voices. Their results indicate that MPFC and left posterior STS carried supramodal representations of perceived emotions in the human brain, because these regions showed modality-independent but emotion category-specific activity patterns.

The work cited above is compatible with the notion that emotions can be evoked by the brain from sensory inputs that are inherently emotional. However, the brain can also create emotions on the basis of mental representation of bodily changes (described as “Core affect”) and representations of prior experiences (termed “Conceptualization”); for a review, see Lindquist et al., 2012). A core affect is elicited by emotional information from the sensory inputs, such as somatic, visceral, vascular, and motor cues. On the other hand, conceptualization refers to a process by which sensations from the body or external world are made meaningful in a given context using representations of prior experience. For instance, emotional words require conceptualization because the symbolic words need a learned link to emotional categories to evoke their affective content. Similar to symbolic words, a neutral, nonemotional stimulus (an image of a fractal, for example) previously associated with an emotion could activate the representation of the emotion associated with it. From the perspective of the brain, the neutral image would activate emotional representations even though the image itself does not contain emotion information.

Although numerous studies have demonstrated that emotions are coded at a supramodal level regardless of the specific

Received Sept. 23, 2014; revised Jan. 28, 2015; accepted Feb. 25, 2015.

Author contributions: J.K., J.S., C.W., and H.H.B. designed research; J.K. performed research; S.-W.L. and H.H.B. contributed unpublished reagents/analytic tools; J.S., T.R., C.W., and H.H.B. analyzed data; J.K., J.S., and T.R. wrote the paper.

This research was supported by the Brain Korea 21 Plus Program through the National Research Foundation of Korea funded by the Ministry of Education.

The authors declare no competing financial interests.

Correspondence should be addressed to Heinrich H. Bühlhoff, Department of Human Perception, Cognition and Action, Max-Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany. E-mail: heinrich.buehlhoff@tuebingen.mpg.de.

DOI:10.1523/JNEUROSCI.4059-14.2015

Copyright © 2015 the authors 0270-6474/15/355655-09\$15.00/0

sensory cues conveying the emotion, these studies have focused only on the neural representation elicited from stimuli inherently conveying emotional information. Hence, it is unknown where and how the brain forms abstract emotional representations from arbitrary stimuli. In this study, we set out to identify abstract emotion representations evoked both by emotional stimuli and by neutral images (fractals) that have been associated with emotions. If there are brain regions encoding emotions by abstracting from specific stimulus characteristics, these regions should be even activated by the memory of an emotional stimulus triggered by an associated, but inherently nonemotional, abstract stimulus. Using searchlight MVPA on fMRI data, we therefore searched for abstract representations of emotions that extend beyond emotional stimuli to nonemotional stimuli associated with emotional stimuli.

Materials and Methods

Participants and ethics approval. Sixteen healthy volunteers (12 males, Koreans, 26.8 ± 2.5 years old) with no history of neurological disorders participated in the study after having given written informed consent. Experimental procedures were approved by the Korea University Institutional Review Board (KU-IRB-12-16-A-1) and the study was conducted in accordance with the Declaration of Helsinki.

Stimuli. Five emotions (anger, disgust, fear, happiness, and sadness) displayed in video clips of two different forms of human movements (dynamic facial expressions and emotional body movements) were used in the experiment. To minimize other-race effects (Walker and Tanaka, 2003), we used facial expressions performed by Korean actors, and emotional body movements in which the face was covered with gray-colored headwear. Video clips of facial expressions and emotional body movements were taken from the Korea University facial expression database (Lee et al., 2012) and the stimulus set created previously (Atkinson et al., 2004, 2007), respectively. In advance of the current study, a pilot experiment had been conducted to select stimuli delivering emotional contents most effectively. In this pilot experiment, eight volunteers (four males, Koreans, 25.8 ± 2.3 years old) without any known difficulties in emotional processing participated. A total of 40 video clips (5 emotions \times 2 stimulus types \times 4 videos per condition) were shown to the participants. Of the four videos per condition, two depicted male actors and two depicted female actors. After the stimulus presentation, participants were asked to rate the perceived emotion intensity displayed, on a scale of 1 (very weak) to 7 (very strong). For each condition, one video clip with the highest intensity rating was chosen for the experiment; mean intensity ratings of selected videos for facial expressions and emotional body movements were 7 and 6.8 ± 0.45 , respectively. In these selected video clips, emotions were expressed by three female and two male actors for each stimulus type (facial expression: disgust/happiness/sadness by female, anger/fear by male; body movement: anger/disgust/fear by female, happiness/sadness by male). Each video clip lasted 3 s and did not contain auditory information.

Nonemotional visual symbols (fractals) for association with emotions were randomly created by Apophysis software (<http://apophysis.org>). Their mean luminance and contrast were matched using the SHINE Toolbox (Willenbockel et al., 2010) for MATLAB (The MathWorks) to minimize the effects of differences in low-level image properties.

Design and procedure. For each form of emotional stimuli, the experiment consisted of three sessions: learning, test, and fMRI session (Fig. 1). Learning and test sessions were conducted outside of the MRI scanner. The fMRI design was 5×4 factorial and comprised of the factors emotions (anger, disgust, fear, happiness, and sadness) and stimulus types (face, body, fractal associated with face stimuli, and fractal associated with body stimuli). The fMRI session consisted of four runs. All emotion types were manipulated within a run, and stimulus type was fixed within a run but varied across the four runs (for more details about the fMRI experiment design, see below).

During the learning session, participants performed 10 blocks of five trials and the order of presentation was counterbalanced across blocks.

Each trial was initiated by the participant via a button press, and the sequence of events in each trial was as follows: blank screen (1 s), fixation cross (2 s), fractal (1.5 s), emotional stimulus (3 s), and fractal again (1.5 s). The same fractal was shown twice in a trial, once before the emotional stimulus and once afterward, to increase the likelihood of an association being formed between the emotional and nonemotional stimuli. A randomly chosen fractal was associated in this manner with each emotional stimulus. Participants were instructed to associate as strongly as possible the fractal with the emotion in the presented video clip and were informed that they would be tested on the association.

The association between fractals and emotional stimuli was tested after the learning session. Participants were presented with the fractal stimuli and tested on how quickly and accurately they could identify the emotion associated with the fractals. This was tested with 10 fractals, 5 associated with emotional stimuli in the learning session and 5 novel fractals. Trials were organized as follows: first, the following instruction text was presented for 5 s: "Select the emotion associated with the following fractal." This was followed by a fixation cross presented for 3 s, followed by a fractal presented for 0.3 s. Participants then had to choose between six options (five emotions, plus "none") using a button press; their response time and correctness were recorded. The percentage of correct answers was shown to the participant at the end of the test session. Participants needed to reach an accuracy of 90% to move on to the fMRI session. If they failed to reach this performance level, they repeated two learning blocks and were tested again until they reached the 90% accuracy threshold. This process guaranteed that every participant had an adequate level of association between emotional and nonemotional stimuli before taking part in the fMRI study.

During the functional image acquisition, participants lay supine in the MR scanner and passively viewed the stimuli through MR-compatible head-mounted display goggles (resolution: 800×600 pixels, Nordic ICE; NordicNeuroLab). Participants were asked to recall the associated emotion as vividly as possible when the fractals were presented. Functional MRI data were acquired in four runs, with two runs for faces and bodies, respectively: one in which the associated fractals were presented and another in which the emotional stimuli were presented. Each run started and ended with a 9 s fixation baseline period. Within each run, 75 trials were presented in 15 blocks of five trials (the five associated fractals or the five emotional stimuli). Trials in each block were presented in pseudorandomized order and each block ended with a fixation cross presented for 6 s. In each trial, a fixation cross was followed by a stimulus, both presented for 3 s. The duration of each block was thus 36 s and each run lasted 9 min 18 s.

MR data acquisition and preprocessing. Magnetic resonance images of the brain were obtained using a 3 T MRI system (Magnetom TrioTim; Siemens Medical Systems) using a standard 16-channel head coil. Anatomical images were obtained using a T1-weighted 3D MPRAGE sequence: TR = 2200 ms, TE = 2.26 ms, flip angle = 9° , FOV = 256 mm, and spatial resolution = $1 \times 1 \times 1$ mm³. Functional images, covering the whole cerebrum, were obtained using a T2*-weighted gradient EPI sequence with 30 slices, TR = 3000 ms, TE = 30 ms, flip angle = 60° , FOV = 250 mm, slice thickness = 4 mm (no gap), and in-plane resolution = 2×2 mm². Preprocessing of the fMRI data was performed using SPM8 (Wellcome Department of Imaging Neuroscience, University College of London, London, UK). The EPI data were corrected for slice-timing difference, realigned for motion correction, coregistered to the individual T1-weighted images, and convolved with a 4 mm FWHM Gaussian kernel to spatially smooth the data. A high-pass filter of 128 s was used to reduce low-frequency noise. Data were analyzed in native subject space; only the performance maps resulting from the individual searchlight analyses were normalized to the MNI space.

Data analysis. The objective of our analysis was to identify brain regions representing emotions regardless of the kind of stimuli conveying the emotion: emotional stimuli or fractals associated with an emotional stimulus. To this end, we searched for information in local neural activation patterns that would allow differentiating between our five emotions. Such MVPAs (Kriegeskorte et al., 2008) have been successfully used to search the whole brain for subtle responses to emotional stimuli (Baucom et al., 2012).

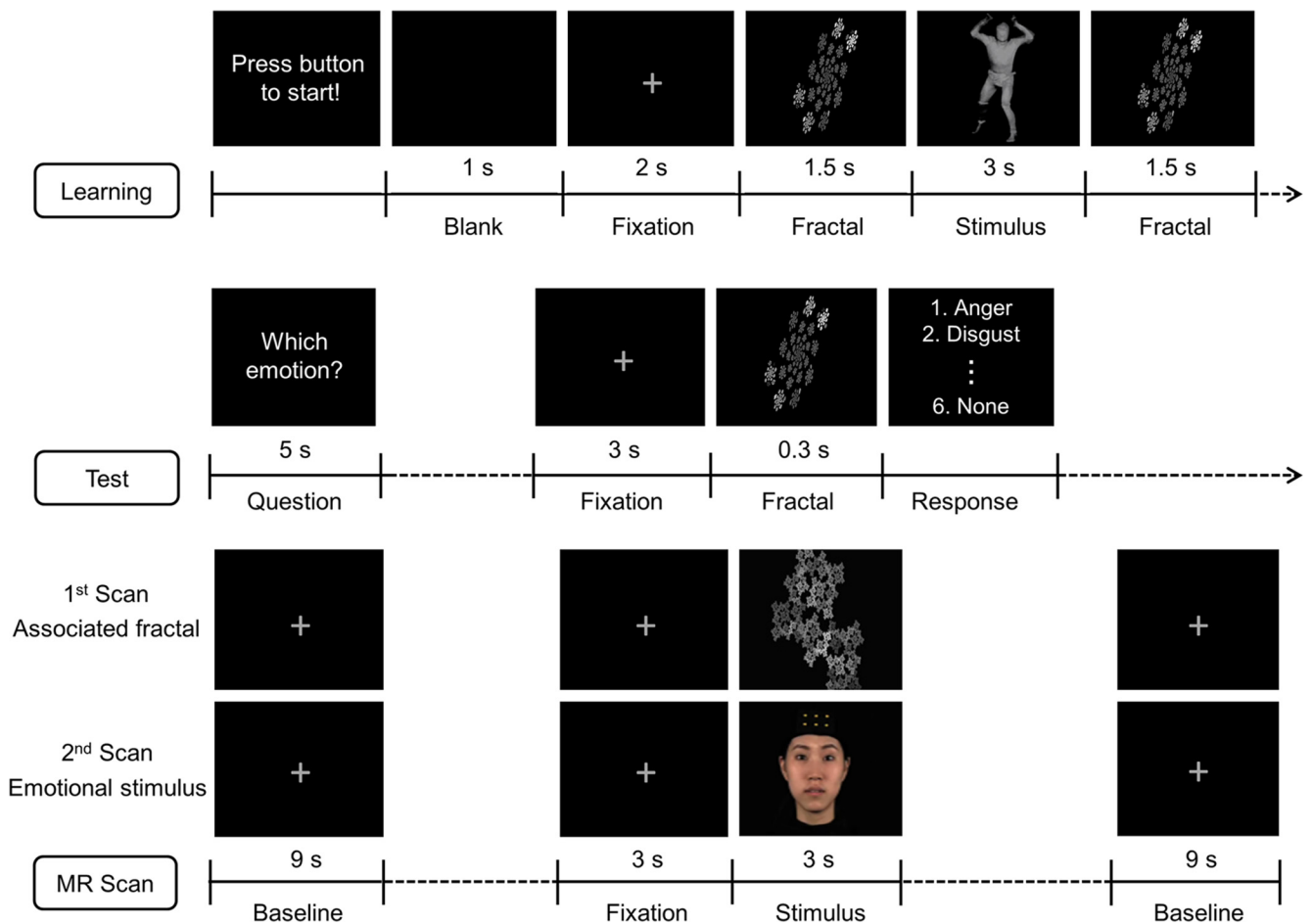


Figure 1. Schematic overview of the experimental design. Participants took part in learning, test, and four fMRI scanning runs in one session. For each stimulus type (body movements and facial expressions), there was one run in which the fractals associated with the emotional stimuli were presented, and another in which the emotional stimuli were presented (total = $2 \times 2 = 4$ runs). Before functional image acquisition, participants were required to associate emotional stimuli with the corresponding symbols (i.e., fractals) and their level of association was tested. Participants repeated the learning session until they reached an accuracy of 90%. Each fMRI run consisted of 15 blocks of five trials and started and ended with a 9 s fixation baseline period. Each block was separated by 6 s of fixation period and trials were presented in pseudorandomized order. In each trial, fixation cross was presented for 3 s, followed by a 3 s stimulus.

Before multivariate analysis, fMRI data were fitted with a GLM implemented in SPM8 to obtain parameter estimates of the response to each stimulus type in each brain voxel (i.e., voxel activation patterns). To increase the number of activation patterns [i.e., exemplars for training the Gaussian Naïve Bayes (GNB) classifier], we divided each run into three arbitrary sub-runs. Specifically, each regressor modeled trials from five blocks, so three regressors per emotion per run were obtained. This provided a good trade-off between the number of trials used to estimate a parameter (more trials = better estimation) and the number of exemplars used for the classifier (more patterns = more exemplars for classifier training). Overall, a total of 60 event-related regressors (5 emotions \times 4 stimulus types \times 3 regressors per run) were used to model the data of each participant. The stimulus onsets were convolved with the canonical HRF of SPM8 and its time derivative in all models. Parameter estimates captured how much a stimulus contributed to the variation of the neuronal signal in a voxel and have been used as input features to the searchlight MVPA in a previous study (Peelen et al., 2010). These parameter estimates were used as input features for an MVPA implemented in the SearchMight Toolbox (Pereira and Botvinick, 2011).

For each voxel v , we selected a searchlight of a $7 \times 7 \times 7$ cube centered on the voxel v (the maximum searchlight size was thus 343 adjacent voxels). Such searchlights were run through the whole brain. In each searchlight, a GNB five-class classifier decoded the five emotion categories from the activation patterns using a twofold cross-validation procedure. For this procedure, data elicited by the emotional stimuli (i.e., video clips of emotional body movements and facial expressions) were

labeled as Group 1 and data elicited by nonemotional symbols (i.e., fractals) were labeled as Group 2. In the first cross-validation step, the GNB classifier was trained on Group 1 data and tested on Group 2 data, and in the second cross-validation step, Group 2 data were used for training and Group 1 data for testing. This cross-validation scheme ensured that only if emotional stimuli as well as emotion-associated fractals arouse similar activity patterns, the GNB classifier could correctly predict the emotion category from the test data. The resulting classification performances from both cross-validation steps were averaged and assigned to the center voxel of the searchlight. Chance-level accuracy (0.2 in this case) was subtracted from the accuracy value stored in each voxel to yield deviations from chance. This yielded individual accuracy maps that were subsequently spatially normalized to the MNI space. A random-effects whole-brain group analysis was performed on the single-subject accuracy maps to identify commonalities among individual decoding accuracies representing the amount of emotion information in the local voxel response patterns. This test was implemented as a one-sample t test against 0 to identify above-chance decoding accuracy in the MVPA. To find an empirical cluster size threshold for a group of participants with searchlight accuracy maps, we applied the following previously described randomization procedure (Oosterhof et al., 2010). We compared the size of the clusters obtained in the group analysis to a reference distribution of clusters obtained by chance. Under the null hypothesis (= no effect), searchlight accuracy values are equally likely to be higher or lower than chance performance. To identify how large clusters would be in this null-hypothesis case, we sampled 16 individual searchlight result maps

and randomly flipped the sign of the maps of a random number of participants (recall that we had subtracted chance performance from these individual maps, such that they contain deviations from chance performance). These maps were then considered one group sample from the null-effect case, and a random-effect analysis on these maps was calculated and the size of the biggest cluster was collected. This procedure was repeated 1000 times, yielding a distribution of cluster sizes under the null hypothesis. We considered as significant clusters obtained of our real analysis if they were in the 5% of upper tail of the null-hypothesis distribution, which therefore amounts to applying the threshold of $p < 0.05$ corrected for multiple comparisons via cluster size.

The significant clusters identified from the searchlight analysis were investigated in more detail in two follow-up analyses: mean activation differences analysis and multidimensional scaling (MDS) analysis.

For the univariate mean activation differences analysis, the β -values for each condition were averaged across voxels of each cluster and then averaged across participants according to the computation procedure described by Peelen et al. (2010), as follows. This procedure assessed whether the mean response in these clusters was more similar within than between emotion categories. To this end, the mean response to each stimulus type (bodies, faces, fractals associated with body movements, and fractals associated with facial expressions) was subtracted from the data. Then, differences between different emotions within the same modality (e.g., fearful faces vs happy faces) and differences between different emotions across different modalities (e.g., fearful faces vs happy bodies) were compared. This was done separately for fractal and video stimuli. The resulting values from each cluster were input into a repeated-measures ANOVA: 5 emotions \times 4 stimulus types. This analysis tested the possibility that the results of the multivariate decoding analysis were caused by mean activation level differences between the conditions.

MDS was used to visualize the abstract emotion representations obtained from the voxel response patterns in each cluster (Edelman et al., 1998; Kriegeskorte et al., 2008). This was computed separately for each cluster, as follows. First, the mean response to each of the 20 stimuli (5 emotions \times 4 stimulus types) was subtracted from the data, as in the univariate mean activation differences analysis above. Then, a matrix of pairwise correlations between the patterns of voxel responses evoked by each of the 20 stimuli was computed. These matrices were Fisher z -transformed and averaged across participants. After inverse z -transform, we applied MDS on this averaged correlation matrix. This yielded a spatial depiction of the representation of emotions for each stimulus type, in each cluster. This analysis allowed us to test the following hypothesis: If the neural structures within an activation cluster indeed represented an emotion independently from the physical stimuli conveying it, one would expect that a given emotion would always be located in the same part of the MDS map, regardless of the stimulus type; for example, all stimuli displaying or associated with “happy” would cluster closely together, and different emotions would be located in different locations and thus remain rather separate in the MDS map.

To test differences between decoding accuracy based on emotional video decoding and associated fractal decoding, we performed two additional searchlight analyses: (1) analysis on data evoked by the emotional video stimuli (i.e., trained on video face and tested on video body, and the reverse) and (2) analysis on data evoked by the associated fractal stimuli (i.e., trained on fractal face and tested on fractal body, and the reverse). Other analysis steps were identical to the aforementioned searchlight analysis. We then assessed the differences between the results of these two analyses (video only and fractal only) using a paired t test random-effects model in SPM8.

Results

Behavioral results

The primary aim of the learning sessions was to form a strong association between the emotional stimuli and the fractals. To ensure the participants obtained a sufficient level of association, their accuracy and response time were measured during the test session (Fig. 2). Two of 16 participants could not reach the 90% accuracy threshold in the first test session, but did reach the cri-

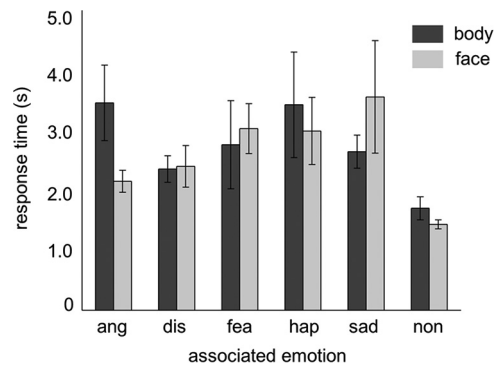


Figure 2. Behavioral results of the test session. Response times for recognition of the emotional conditions associated from presented fractal stimuli, calculated after excluding error trials ($<5\%$ of trials were errors). Mean and SEM are shown. Unassociated fractals were recognized significantly faster than emotion-associated fractals. Ang, anger; dis, disgust; fea, fear; hap, happiness; sad, sadness; non, unassociated.

terion after repeating two learning blocks. Participants' mean accuracies for the body-fractal and face-fractal association were 95.6 ± 5.1 and $98.8 \pm 3.4\%$, respectively. Two-way ANOVAs with the factors stimulus type and emotion type revealed no significant main effects on response time (stimulus types: $F_{(1,150)} = 0.05$, $p = 0.83$; emotion types: $F_{(4,150)} = 0.57$, $p = 0.67$) or accuracy (stimulus types: $F_{(1,150)} = 2.95$, $p = 0.09$; emotion types: $F_{(4,150)} = 0.71$, $p = 0.59$). Moreover, there was no significant interaction between stimulus type and emotion (response time: $F_{(4,150)} = 0.97$, $p = 0.43$; mean accuracy: $F_{(4,150)} = 1.19$, $p = 0.32$). Thus, behavioral results showed no signs of difference in the degree of association across conditions.

fMRI results

Searchlight MVPA identified regions with abstract representations of emotions

The main analysis aimed to identify regions in which the local pattern of activation allowed the discrimination of the five emotions regardless of the stimuli mediating the emotion. To this end, a random-effects group analysis ($N = 16$) was performed on the individual accuracy maps obtained in the searchlight analysis. Each accuracy map was the result of the average decoding performance of a classifier trained on BOLD activation patterns evoked by emotional stimuli and tested on activation patterns evoked by fractal stimuli associated with the emotional stimuli, and the reverse (classifier trained on fractal activation patterns and tested on emotional activation patterns). The analysis revealed four clusters: the posterior portion of the cingulate cortex (PCC), the precuneus, the MPFC, and the angular gyrus in the right hemisphere ($p < 0.001$ uncorrected, cluster size >50 ; Fig. 3, left column, Table 1). These clusters were unlikely to have been found by chance as a result of performing a large number of tests: a randomization procedure (Oosterhof et al., 2010) revealed that the probabilities of obtaining clusters as large as ours given our data are <0.05 . Our clusters were significant after correcting for multiple comparisons (Nichols and Hayasaka, 2003; Oosterhof et al., 2010). To ensure that the clusters identified in this searchlight analysis contain information on emotion type, we performed cluster-based MVPA as suggested as a “Confirmatory test” by Etzel et al. (2013). We retrained a classifier on voxel values in the identified clusters using the cross-training MVPA approach by Polyn et al. (2005). Results showed that decoding accuracies were significantly higher than chance level (PCC: $t_{(15)} = 7.8$, $p < 0.05$; precuneus: $t_{(15)} = 12.2$, $p < 0.05$; MPFC: $t_{(15)} = 13.1$, $p < 0.05$; angular gyrus: $t_{(15)} = 9.6$, $p < 0.05$; Fig. 4). Therefore, this anal-

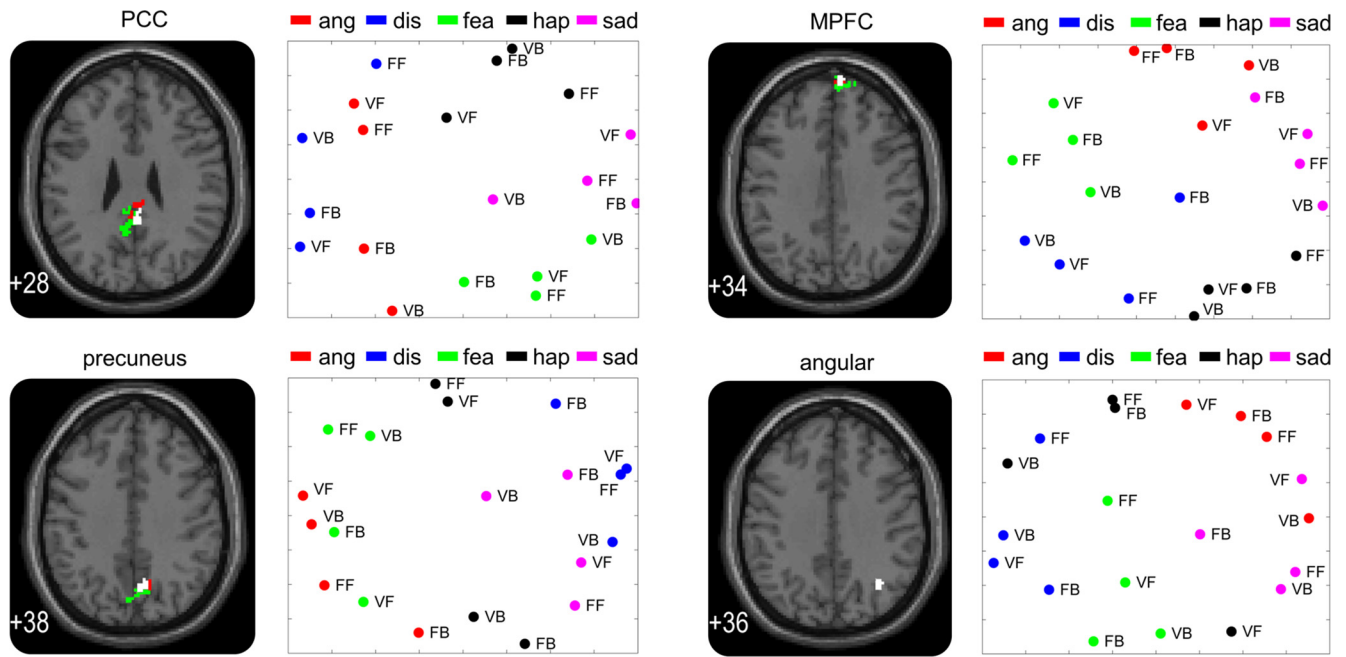


Figure 3. Results of the whole-brain searchlight and the multidimensional scaling analyses. Four brain regions showed emotion-specific multivoxel patterns regardless of the stimulus types. For each region (PCC, precuneus, MPFC, and angular gyrus), the left side shows a sagittal slice of the brain (z-coordinate of slice indicated in bottom left corner) with significant clusters identified by the searchlight analysis run on data elicited by both the emotional video clips and the associated fractals (colored in white), solely by the emotional video clips (colored in green), and solely by the associated fractals (colored in red). The right side shows multidimensional scaling maps of the emotions based on voxel response patterns in each searchlight cluster exhibiting emotion-specific information. Each color represents one of the five emotional conditions and each text represents one of the four stimulus types; i.e., VB, video clips of body movements; VF, video clips of facial expressions; FB, fractals of body movements; FF, fractals of facial expressions; ang, anger; dis, disgust; fea, fear; hap, happiness; sad, sadness.

Table 1. Significant clusters involved in modality-independent emotional processing ($p < 0.001$ uncorrected, cluster size > 50)

Anatomical region	Hemisphere	MNI coordinates			Cluster size	<i>T</i>	<i>Z</i>
		<i>x</i>	<i>y</i>	<i>z</i>			
PCC	R	4	−44	28	164	8.3	5.01
Precuneus	R	14	−62	38	145	7.65	4.81
MPFC	R	8	62	34	71	6.62	4.46
Angular gyrus	R	38	−62	36	107	6.42	4.39

R, right; L, left; cluster size indicates *N* voxels; *T* indicates peak *t* values; *Z* indicates peak *z* values.

ysis confirmed that the activation patterns in the clusters identified in our whole-brain searchlight contain enough information to allow discrimination between the five emotional conditions, even though the brain signals were elicited not only by the emotional stimuli but also by the emotion-associated fractals. In other words, these clusters contained stimulus-independent information on specific emotions encoded in multivoxel patterns of neural activity.

To test whether the decoding results in the four significant clusters arose from mean activation differences between conditions, we averaged the stimulus-related activation (i.e., parameter estimates from the GLMs) across the clusters' voxels (Fig. 5). As in the study by Peelen et al. (2010), activations in all conditions were below baseline, which is commonly observed in these regions in response to emotional stimuli (Mitchell et al., 2002), due to high activation in these regions during resting states (Gusnard and Raichle, 2001). A repeated-measures ANOVA with the factors stimulus type and emotion did not find any effects of stimulus or emotion or interaction between these factors on the mean activation (parameter estimates) in the four clusters (all $F_{(4,60)}s < 1$, $p > 0.42$). This confirms that multivoxel activation patterns and not mean activation differences underlay the decoding results.

Cluster-based MVPA confirmed representations of emotions based solely on nonemotional, associated stimuli

To examine whether emotions could be decoded just from the associated fractal data in the identified clusters, we trained a GNB classifier on fractal-face data and tested on fractal-body data, and the reverse. Then, we compared the decoding accuracy against chance. While decoding using data from the angular gyrus was not higher than chance ($t_{(15)} = 0.52$), decoding based on data from the PCC, the precuneus, and the MPFC was significantly higher than chance (all $ps < 0.05$; PCC: $t_{(15)} = 3.64$, precuneus: $t_{(15)} = 5.71$, MPFC: $t_{(15)} = 6.08$). These results suggest that the activation patterns in PCC, precuneus, and MPFC indeed represented emotional content abstractly, i.e., without reference to the type of emotional stimulus, which had originally conveyed the emotional information.

Cluster-based MVPA comparing video-fractal association strength

Furthermore, for each of the four clusters identified in the main searchlight analysis, we systematically investigated whether emotions conveyed by facial expressions or by body movements provided stronger neural evidence of associations to fractals. For this purpose, we compared the classification performances across two association conditions: (1) the classifier was trained on video-face data and tested on fractal-face data and (2) the classifier was trained on video-body data and tested on fractal-body data. Paired *t* tests revealed no significant difference between these association conditions in any searchlight cluster (all $t_{(15)}s < 2.07$, $p > 0.06$). Thus, neural signals of associations between fractals and emotional stimuli were about equally strong for both types of emotional stimuli (body and face videos).

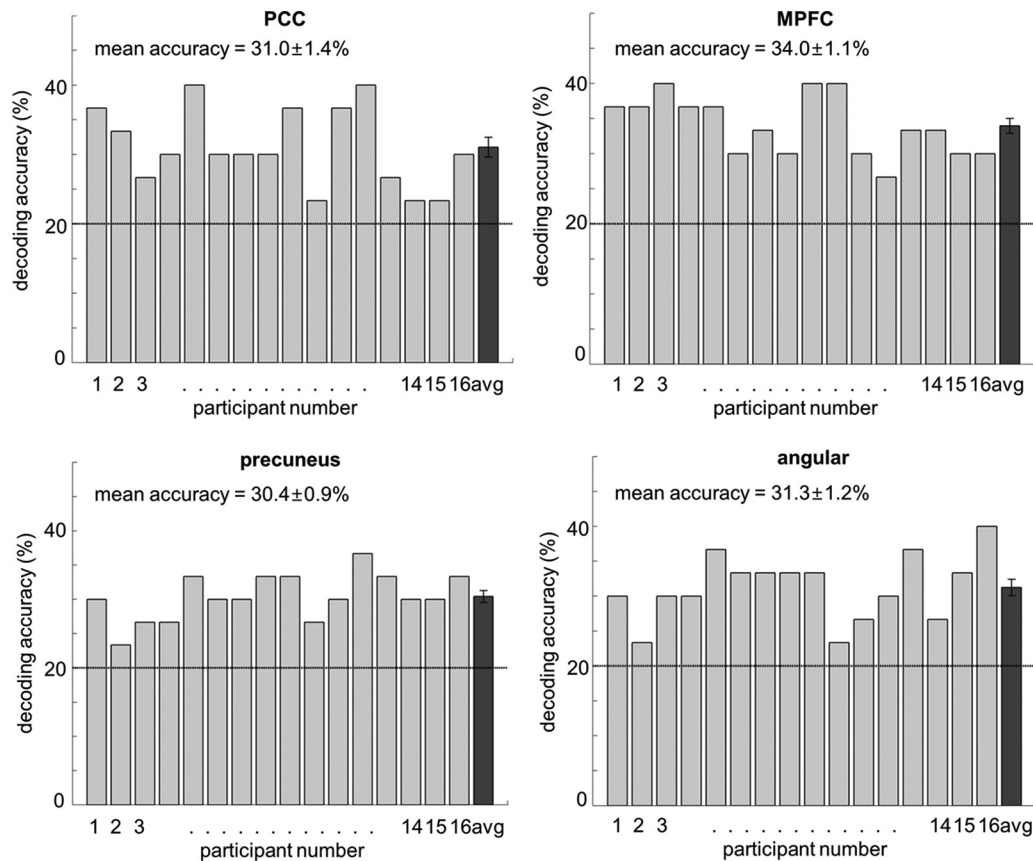


Figure 4. Decoding accuracies of a “confirmatory test” for each of the 16 participants. In this test, decoding was performed only on the voxels in the cluster, not on the searchlights surrounding each of the voxels in the cluster. The rightmost bar shows the average and SEM accuracy across participants. Decoding accuracies were significantly higher than chance (20%, indicated by the dashed line) in every cluster, which confirms that clusters contain emotion-specific information.

Searchlight MVPA comparing representations of emotions evoked only by emotional or nonemotional, associated stimuli

To compare representations of emotions based on stimuli directly carrying emotional information (video stimuli) and those based on stimuli associated with emotions (fractals), we ran two additional searchlight analyses. The first analysis was run only on the data elicited by the emotional stimuli and identified four significant clusters. These clusters were located in MPFC, in cuneus extending into precuneus, PCC, and superior frontal gyrus ($p < 0.001$ uncorrected, cluster size > 50 ; Table 2). The second analysis was run only on the data elicited by the associated fractal stimuli and identified clusters in MPFC, PCC, and precuneus ($p < 0.001$ uncorrected, cluster size > 50 ; Table 3). A randomization procedure (Oosterhof et al., 2010) confirmed that those clusters were significant at $p < 0.05$ after correcting for multiple comparisons. These results further confirmed that PCC, precuneus, and MPFC contained abstract representations of emotions. We then assessed the differences between the results of these two analyses (video only and fractal only) using a paired t test, random-effects model. The results show that video stimuli led to significantly higher decoding performance than associated fractal stimuli in cuneus and the superior frontal gyrus, but no cluster showing higher decoding performance based on the associated fractal stimuli was identified ($p < 0.001$ uncorrected, cluster size > 50).

Visual depiction of emotion representation based on multidimensional scaling analysis

To illustrate the representation of emotions in the clusters we identified in the searchlight analysis, we investigated the similarities between voxel response patterns evoked by the 20 stimuli (5 emotions \times 4 stimulus types) using an MDS (see Materials and Methods; Edelman et al., 1998; Kriegeskorte et al., 2008). The MDS map (Fig. 3, right column) shows that voxel response patterns of specific emotions grouped together regardless of the specific stimulus by which an emotion was aroused. Thus, independent from the physical stimulus (i.e., an emotional body movement or facial expression, or a fractal associated with an emotion), specific emotions were represented in similar voxel response patterns in the four clusters. To estimate how closely the same emotions clustered together in these MDS maps, we calculated the Euclidean distances between all possible pairs of coordinates representing the same emotion expressed by different stimuli (within-emotion distances), and the distances between all possible pairs of coordinates representing different emotions expressed by different stimuli (between-emotion distances). Then, we compared these distances using a three-way ANOVA, with the factors (1) within versus between emotions, (2) emotions, and (3) clusters. A significant main effect of the factor within/between emotions was found ($F_{(1,1360)} = 184.63$, $p < 0.05$), but no main effect of emotions or clusters. There was a significant interaction between within/between emotions and clusters ($F_{(3,1360)} = 3.3$, $p < 0.05$), indicating that the grouping of emotions was not

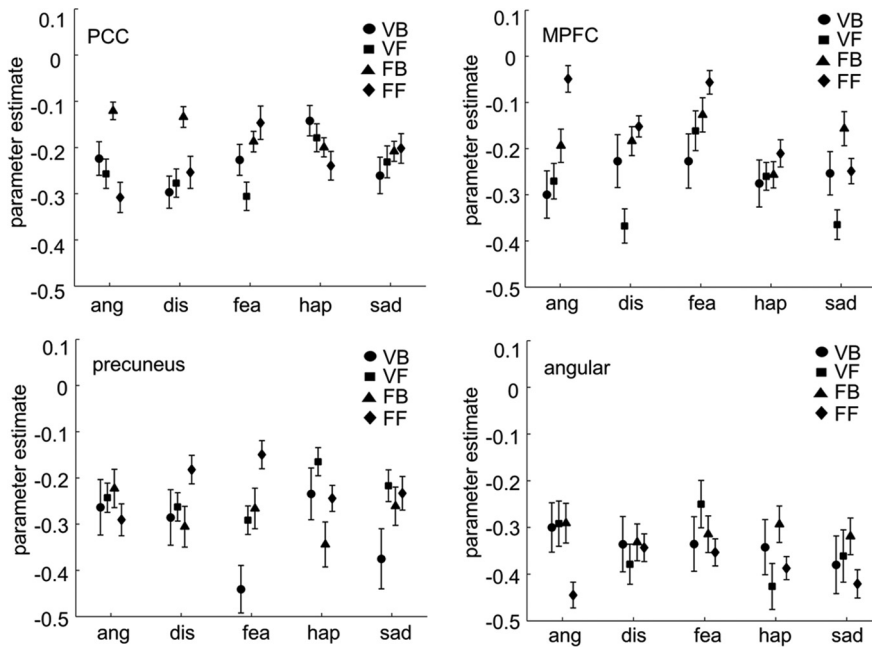


Figure 5. Univariate mean activation (parameter estimates) differences for each experimental condition in the four clusters allowing significant emotion decoding. Error bars show SEM. VB, video clips of body movements; VF, video clips of facial expressions; FB, fractals of body movements; FF, fractals of facial expressions; ang, anger; dis, disgust; fea, fear; hap, happiness; sad, sadness.

Table 2. Significant clusters from a searchlight analysis on the fMRI data elicited by the emotional video stimuli ($p < 0.001$ uncorrected, cluster size > 50)

Anatomical region	Hemisphere	MNI coordinates			Cluster size	T	Z
		x	y	z			
MPFC	R	8	62	30	189	6.01	4.22
Cuneus/precuneus	R	10	-84	12	538	5.21	3.88
PCC	L	0	-44	28	157	5.04	3.78
Superior frontal gyrus	L	-20	42	48	103	4.99	3.77

R, right; L, left; cluster size indicates N voxels; T indicates peak t values; Z indicates peak z values.

Table 3. Significant clusters from a searchlight analysis on the fMRI data elicited by the associated fractal stimuli ($p < 0.001$ uncorrected, cluster size > 50)

Anatomical region	Hemisphere	MNI coordinates			Cluster size	T	Z
		x	y	z			
MPFC	R	12	60	22	84	5.61	4.06
PCC	R	0	-44	24	82	5.61	4.06
Precuneus	R	6	-62	54	62	5.47	3.99

R, right; L, left; cluster size indicates N voxels; T indicates peak t values; Z indicates peak z values.

equally clear in all clusters. A Tukey *post hoc* test revealed a marginally larger within/between difference (i.e., a more distinct representation) in MPFC than in angular gyrus ($p = 0.09$).

Discussion

In this study, we investigated which brain regions represent emotions at an abstract level. We searched for neural representations independent from the stimulus mediating the emotion, and independent of whether the stimulus itself carried an emotional meaning or only evoked a memory of an emotional stimulus. Using searchlight MVPA, we found that neural activity patterns in PCC, precuneus, MPFC, and angular gyrus of the right hemisphere contained emotion-specific information regardless of the stimulus type arousing the emotion. This emotion-specific information could even be decoded just from the data evoked by frac-

tals that had been associated with emotions, i.e., stimuli that carry no intrinsic emotional information. Mean activation difference in these clusters did not vary with stimulus type or emotion, which confirms that the decoding results were based on fine-grained activation patterns. By using a cross-training scheme in which emotion classifiers were trained either on voxel response patterns evoked by emotional stimuli or by fractals, our searchlight results support the hypothesis that these brain regions encode perceived emotions at an abstract level. In line with this observation, a multidimensional scaling analysis revealed that activation patterns of these regions clustered for specific emotions, independent from the stimulus input, which conveyed the emotion.

Previous studies reported that PCC and precuneus play a similar role in emotional information processing because of their structural and functional similarities (Ochsner et al., 2004; Cavanna and Trimble, 2006). In addition, these regions are reciprocally connected to areas involved in emotional processing such as anterior cingulate and orbital frontal cortices (Baleydier and Mauguier, 1980). Neuroimaging studies revealed that PCC and precuneus were activated during the retrieval of emotional memories (Andreasen et al., 1995), the evaluation of emotional words (Maddock and Buonocore, 1997; Maddock et al., 2003), the processing of self-relevant affect (Vogt et al., 2006), and the retrieval of autobiographical memory (Maddock et al., 2001). Moreover, activations of these regions are consistently evoked by visual or auditory emotional stimuli, or autobiographical recall of emotional events (Phan et al., 2002). Given these findings it is plausible that PCC and precuneus carry emotion-specific information from different stimulus types. Providing more support to this hypothesis, a recent fMRI study revealed that PCC exhibits greater BOLD responses to congruent than to incongruent auditory-visual stimuli of different emotional categories, suggesting an important role in representing emotional stimuli (Klasen et al., 2011).

There is considerable evidence for MPFC's contribution to affective processing. A lesion study revealed that damage to the ventromedial prefrontal cortex impaired both emotional responses and emotion regulation (Koenigs et al., 2007). Human neuroimaging studies showed regional specific activations related to distinct emotional conditions and suggested that MPFC modulates brain activity for affective processes (Phan et al., 2002; Murphy et al., 2003; Etkin et al., 2011). Previous literature therefore suggests that the MPFC plays a general role in emotional processing. Recently, it was proposed that the MPFC represented emotions supramodally across face, body, and voice stimuli (Peelen et al., 2010). Our present findings provide further and more elaborate evidence for the role of MPFC in abstract emotional processing by showing that emotion-specific representations can even be evoked by memory. The key finding was that similar emotion-specific voxel response patterns were elicited by emotional stimuli as well as by emotional memories cued by abstract symbols previously associated with emotional stimuli. The current results, together with those of Peelen et al. (2010),

suggest that MPFC encodes emotional information independently of the physical characteristics of stimuli.

Interestingly, all of the resulting brain regions are well known as substantial parts of the brain networks for “theory-of-mind” (Premack and Woodruff, 1978). Theory-of-mind, also known as mentalizing, refers to a cognitive ability to attribute mental states to oneself and others, and understanding others’ emotional states. A significant amount of neuroscientific research investigated the neural bases of this ability (for review, see Frith and Frith, 1999; Mar, 2011). A quantitative meta-analysis of neuroimaging studies on emotion by Kober et al. (2008) identified six groups of brain regions playing a role in emotional processing. Among these functional groups, the brain areas “medial posterior group” and “medial prefrontal cortex group” are considered as parts of the network that helps to constitute mentalizing or conceptualization. Our searchlight analysis results are largely consistent with this previous meta-analysis. Combining evidence from the present and the previous findings, we suggest that these regions carry abstract representations of emotions that could contribute to understanding and evaluating others’ emotional states.

It is worth noting the difference between the results from three searchlight analyses; the searchlight analysis on the fMRI data elicited (1) by both the emotional video clips and the associated fractals, (2) solely by the emotional video clips, and (3) solely by the associated fractals. The PCC, the precuneus, and the MPFC regions were found as significant in all three analyses. However, the angular gyrus appeared only in the analysis (1) and the cuneus and the superior frontal gyrus were observed only in the analysis (2). A possible explanation for the absence of angular gyrus activation in analysis (2) is that it may be involved in emotional memory retrieval rather than the abstract emotional processing. Previous neuroimaging studies showed that angular gyrus played a role in the retrieving information held in the memory system of the brain, including emotional stimulus retrieval (Fischer et al., 1996; McDermott et al., 1999). For example, activation of the angular gyrus was found during the retrieval of affective memories in an event-related fMRI study (McDermott et al., 1999) and a PET study (Fischer et al., 1996). Our MDS results, which showed the sparsest clustering pattern in angular gyrus, provide additional supports for this speculation. Furthermore, we observed cuneus and superior frontal gyrus activations only in analysis (2). This might be due to aspects of the characteristics of visual stimuli, i.e., some motion characteristics common to the stimuli evoking a given emotion (e.g., video clips of body movements and facial expressions of fear contain common low-level visual features). This could be a possible explanation for low-level visual responses in the cuneus. Another explanation would be an emotion-specific top-down modulation of attention, which could explain both superior frontal gyrus and cuneus involvements. Attempting to identify these common features is beyond the scope of the current study; the aim of analysis (2) was to replicate the study of Peelen et al. (2010).

To our knowledge, our study is the first to show that emotion-associated fractals and inherently emotional stimuli share abstract emotional representations: MVPA revealed that emotions that were internally generated by associated fractals activated neural response patterns similar to external, inherently emotional stimuli. In line with our observation, previous neuroimaging studies reported that internal stimuli that are “mentally imagined” (conceptualized) recruit similar brain regions and elicit similar brain activation or activation patterns as the actual external stimuli: similar brain regions are activated by emotions that are aroused by emotional input such as emotional films and

the memory thereof (Reiman et al., 1997; Damasio et al., 2000; Phan et al., 2002). Visual stimuli and mental images of the stimuli both activate sensory visual regions and are represented as similar retinotopic activation patterns (Kosslyn et al., 1999; Klein et al., 2004; Thirion et al., 2006). Methodologically similar to our study, previous fMRI studies showed that abstract shapes associated with reward objects or potential outcomes of a decision activate similar neural representation as the reward objects or decision outcomes (Barron et al., 2013; Klein-Flügge et al., 2013). Thus, our findings support the general notion that external stimuli and internal, mentally imagined events recruit similar neural mechanisms.

Our findings may be limited by the fact that participants’ emotions were evoked from displays evoking emotional states of others, but not self-generated emotions. Neural emotion representations elicited by perceiving others’ emotional states are likely to be different from self-generated emotions. In this respect, our results can be confined to the understanding of other’s emotional states. Another limitation is that the emotional states were conveyed only via the visual modality. Thus, our results may not generalize to other modalities as, for example, auditory emotional stimuli. Last, the generality of our findings could be compromised by the fact that the sample size of this study was rather small ($N = 16$) for an fMRI study on healthy participants, and the fact that the gender distribution was unequal. Even with these potential limitations, our study introduced a fruitful new approach to access abstract representations of emotions using emotion-associated symbols together with multivariate decoding methods.

In conclusion, we found abstract representations of emotions using both emotional stimuli and emotion-associated fractals in four brain regions: PCC, precuneus, MPFC, and angular gyrus in the right hemisphere. Among these four candidates, the first three probably contain emotion-specific but modality-independent representations of perceived emotions in accordance with previous studies (Peelen et al., 2010; Klaser et al., 2011), while angular gyrus is more likely to contribute to the retrieval of emotional memory. Future studies are needed to verify the detailed functional roles of these areas, for example, a functional connectivity analysis to examine inter-region interactions.

References

- Andreasen NC, O’Leary DS, Cizadlo T, Arndt S, Rezai K, Watkins GL, Ponto LL, Hichwa RD (1995) Remembering the past: two facets of episodic memory explored with positron emission tomography. *Am J Psychiatry* 152:1576–1585. [CrossRef Medline](#)
- Atkinson AP, Dittrich WH, Gemmell AJ, Young AW (2004) Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception* 33:717–746. [CrossRef Medline](#)
- Atkinson AP, Tunstall ML, Dittrich WH (2007) Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition* 104:59–72. [CrossRef Medline](#)
- Aubé W, Angulo-Perkins A, Peretz I, Concha L, Armony JL (2015) Fear across the senses: brain responses to music, vocalizations and facial expressions. *Soc Cogn Affect Neurosci* 10:399–407. [CrossRef Medline](#)
- Baleydier C, Mauguier F (1980) The duality of the cingulate gyrus in monkey -Neuroanatomical study and functional hypothesis. *Brain* 103:525–554. [CrossRef Medline](#)
- Barron HC, Dolan RJ, Behrens TE (2013) Online evaluation of novel choices by simultaneous representation of multiple memories. *Nat Neurosci* 16:1492–1498. [CrossRef Medline](#)
- Baucom LB, Wedell DH, Wang J, Blitzer DN, Shinkareva SV (2012) Decoding the neural representation of affective states. *Neuroimage* 59:718–727. [CrossRef Medline](#)
- Cavanna AE, Trimble MR (2006) The precuneus: a review of its functional

- anatomy and behavioural correlates. *Brain* 129:564–583. [CrossRef Medline](#)
- Damasio AR, Grabowski TJ, Bechara A, Damasio H, Ponto LL, Parvizi J, Hichwa RD (2000) Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat Neurosci* 3:1049–1056. [CrossRef Medline](#)
- Edelman S, Grill-Spector K, Kushnir T, Malach R (1998) Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology* 26:309–321.
- Etkin A, Egner T, Kalisch R (2011) Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn Sci* 15:85–93. [CrossRef Medline](#)
- Etzel JA, Zacks JM, Braver TS (2013) Searchlight analysis: promise, pitfalls, and potential. *Neuroimage* 78:261–269. [CrossRef Medline](#)
- Fischer H, Wik G, Fredrikson M (1996) Functional neuroanatomy of robbery re-experience: affective memories studied with PET. *Neuroreport* 7:2081–2086. [CrossRef Medline](#)
- Frith CD, Frith U (1999) Cognitive psychology—interacting minds—a biological basis. *Science* 286:1692–1695. [CrossRef Medline](#)
- Gusnard DA, Raichle ME (2001) Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci* 2:685–694. [CrossRef Medline](#)
- Klassen M, Kenworthy CA, Mathiak KA, Kircher TT, Mathiak K (2011) Supramodal representation of emotions. *J Neurosci* 31:13635–13643. [CrossRef Medline](#)
- Klein I, Dubois J, Mangin JF, Kherif F, Flandin G, Poline JB, Denis M, Kosslyn SM, Le Bihan D (2004) Retinotopic organization of visual mental images as revealed by functional magnetic resonance imaging. *Brain Res Cogn Brain Res* 22:26–31. [CrossRef Medline](#)
- Klein-Flügge MC, Barron HC, Brodersen KH, Dolan RJ, Behrens TE (2013) Segregated encoding of reward-identity and stimulus-reward associations in human orbitofrontal cortex. *J Neurosci* 33:3202–3211. [CrossRef Medline](#)
- Kober H, Barrett LF, Joseph J, Bliss-Moreau E, Lindquist K, Wager TD (2008) Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *Neuroimage* 42:998–1031. [CrossRef Medline](#)
- Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, Damasio A (2007) Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* 446:908–911. [CrossRef Medline](#)
- Kosslyn SM, Pascual-Leone A, Felician O, Camposano S, Keenan JP, Thompson WL, Ganis G, Sukel KE, Alpert NM (1999) The role of Area 17 in visual imagery: convergent evidence from PET and rTMS. *Science* 284:167–170. [CrossRef Medline](#)
- Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis—connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4. [CrossRef Medline](#)
- Lee H, Shin A, Kim B, Wallraven C (2012) P2–35: the KU facial expression database: a validated database of emotional and conversational expressions. *i-Perception* 3:694.
- Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, Barrett LF (2012) The brain basis of emotion: a meta-analytic review. *Behav Brain Sci* 35:121–143. [CrossRef Medline](#)
- Maddock RJ, Buonocore MH (1997) Activation of left posterior cingulate gyrus by the auditory presentation of threat-related words: an fMRI study. *Psychiatry Res* 75:1–14. [CrossRef Medline](#)
- Maddock RJ, Garrett AS, Buonocore MH (2001) Remembering familiar people: the posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience* 104:667–676. [CrossRef Medline](#)
- Maddock RJ, Garrett AS, Buonocore MH (2003) Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Hum Brain Mapp* 18:30–41. [CrossRef Medline](#)
- Mar RA (2011) The neural bases of social cognition and story comprehension. *Annu Rev Psychol* 62:103–134. [CrossRef Medline](#)
- McDermott KB, Ojemann JG, Petersen SE, Ollinger JM, Snyder AZ, Akbudak E, Conturo TE, Raichle ME (1999) Direct comparison of episodic encoding and retrieval of words: an event-related fMRI study. *Memory* 7:661–678. [CrossRef Medline](#)
- Mitchell JP, Heatherton TF, Macrae CN (2002) Distinct neural systems subserve person and object knowledge. *Proc Natl Acad Sci U S A* 99:15238–15243. [CrossRef Medline](#)
- Murphy FC, Nimmo-Smith I, Lawrence AD (2003) Functional neuroanatomy of emotions: a meta-analysis. *Cogn Affect Behav Neurosci* 3:207–233. [CrossRef Medline](#)
- Nichols T, Hayasaka S (2003) Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat Methods Med Res* 12:419–446. [CrossRef Medline](#)
- Ochsner KN, Knierim K, Ludlow DH, Hanelin J, Ramachandran T, Glover G, Mackey SC (2004) Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other. *J Cogn Neurosci* 16:1746–1772. [CrossRef Medline](#)
- Oosterhof NN, Wiggett AJ, Diedrichsen J, Tipper SP, Downing PE (2010) Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex. *J Neurophysiol* 104:1077–1089. [CrossRef Medline](#)
- Peelen MV, Atkinson AP, Vuilleumier P (2010) Supramodal representations of perceived emotions in the human brain. *J Neurosci* 30:10127–10134. [CrossRef Medline](#)
- Pereira F, Botvinick M (2011) Information mapping with pattern classifiers: a comparative study. *Neuroimage* 56:476–496. [CrossRef Medline](#)
- Phan KL, Wager T, Taylor SF, Liberzon I (2002) Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage* 16:331–348. [CrossRef Medline](#)
- Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-specific cortical activity precedes retrieval during memory search. *Science* 310:1963–1966. [CrossRef Medline](#)
- Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind. *Behav Brain Sci* 1:515–526. [CrossRef](#)
- Reiman EM, Lane RD, Ahern GL, Schwartz GE, Davidson RJ, Friston KJ, Yun LS, Chen K (1997) Neuroanatomical correlates of externally and internally generated human emotion. *Am J Psychiatry* 154:918–925. [CrossRef Medline](#)
- Thirion B, Duchesnay E, Hubbard E, Dubois J, Poline JB, LeBihan D, Dehaene S (2006) Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage* 33:1104–1116. [CrossRef Medline](#)
- Vogt BA, Vogt L, Laureys S (2006) Cytology and functionally correlated circuits of human posterior cingulate areas. *Neuroimage* 29:452–466. [CrossRef Medline](#)
- Walker PM, Tanaka JW (2003) An encoding advantage for own-race versus other-race faces. *Perception* 32:1117–1125. [CrossRef Medline](#)
- Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW (2010) Controlling low-level image properties: the SHINE toolbox. *Behav Res Methods* 42:671–684. [CrossRef Medline](#)