

# Pattern classification predicts individuals' responses to affective stimuli

Citation for published version (APA):

Yuen, K. S. L., Johnston, S. J., de Martino, F., Sorger, B., Formisano, E., Linden, D. E. J., & Goebel, R. (2012). Pattern classification predicts individuals' responses to affective stimuli. *Translational Neuroscience*, 3(3), 278-287. <https://doi.org/10.2478/s13380-012-0029-6>

## Document status and date:

Published: 01/01/2012

## DOI:

[10.2478/s13380-012-0029-6](https://doi.org/10.2478/s13380-012-0029-6)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Kenneth S.L. Yuen<sup>1,4</sup>,  
 Stephen J. Johnston<sup>2</sup>,  
 Federico De Martino<sup>3</sup>,  
 Bettina Sorger<sup>3</sup>,  
 Elia Formisano<sup>3</sup>,  
 David E. J. Linden<sup>1,3,5</sup>,  
 Rainer Goebel<sup>3</sup>

<sup>1</sup>Wolfson Centre for Clinical and Cognitive Neuroscience, School of Psychology, Bangor University, Bangor LL57 2AS, UK

<sup>2</sup>Dept. of Psychology, Brunel University, Uxbridge, UB8 3PH, UK

<sup>3</sup>Dept. of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht MD 6200, Netherlands

<sup>4</sup>Institute for Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany

<sup>5</sup>MRC Centre for Neuropsychiatric Genetics & Genomics, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK

Received 13 June 2012  
 accepted 11 July 2012

# PATTERN CLASSIFICATION PREDICTS INDIVIDUALS' RESPONSES TO AFFECTIVE STIMULI

## Abstract

Since the successful demonstration of “brain reading” of fMRI BOLD signals using multivoxel pattern classification (MVPA) techniques, the neuroimaging community has made vigorous attempts to exploit the technique in order to identify the signature patterns of brain activities associated with different cognitive processes or mental states. In the current study, we tested whether the valence and arousal dimensions of the affective information could be used to successfully predict individual's active affective states. Using a whole-brain MVPA approach, together with feature elimination procedures, we are able to discriminate between brain activation patterns associated with the processing of positive or negative valence and cross validate the discriminant function with an independent data set. Arousal information, on the other hand, failed to provide such discriminating power. With an independent sample, we test further whether the MVPA identified brain network could be used for inter-individual classification. Although the inter-subject classification success was only marginal, we found correlations with individual differences in affective processing. We discuss the implications of our findings for future attempts to classify patients based on their responses to affective stimuli.

## Keywords

• Pattern classification • Multivoxel pattern classification (MVPA) • Machine learning • Emotion • Affective processing • fMRI

© Versita Sp. z o.o.

## Introduction

Since the successful demonstration of “brain reading” of fMRI blood-oxygen level dependent (BOLD) signals using multivoxel pattern classification (MVPA) techniques [1-3], the neuroimaging community has made vigorous attempts to exploit the technique in order to identify the signature patterns of brain activities associated with different cognitive processes or mental states. By applying machine learning algorithms to fMRI BOLD signals, MVPA extracts information from patterns of activities within a selected set of brain voxels and classifies these patterns into different categories. Published studies have already demonstrated that MVPA can decipher the information content of brain activity patterns during conscious and sub-conscious visual perception [2,4], object recognition [5], working memory and long-term memory [6], and decision-making processes [7,8]. MVPA is also a promising tool for psychiatric research. Most psychiatric applications have, so far, aimed to classify

individuals into diagnostic groups by their patterns of brain activation or structure [9,10], or to predict treatment response or prognosis [11,12]

In the affective domain, a number of studies have attempted to classify the processing of emotion-related information in the human brain, including near-threshold fear [13], pleasantness of thermal stimuli [14], emotional prosody [15], imagery [16], and cross-modal integration of emotional cues from faces and body signals [17]. These studies show that multivariate statistics can predict the valence of sensory stimuli and even internally generated affective states [16] from the associated brain activation pattern. However, most previous studies did not manipulate the valence and arousal of affective stimuli or states independently, and thus it has remained open, which of these dimensions was driving the classification success. The present study addresses this issue by orthogonal manipulation of valence and arousal in positive and negative visual stimuli.

Another open question addressed in the present paper is whether such classification will work across subjects, which will be important for any diagnostic application of normative data. This is an important area of work because valence-driven classification will open up the possibility to perform intra- and inter-individual classification of abnormal affective states using BOLD fMRI signals.

In the current paper, we demonstrate that multivoxel pattern classifications can decode brain patterns associated with the appraisal of affective information. Furthermore, we show that the brain regions that provide distinct patterns of brain activity have the potential to be utilized as region-of-interest to perform inter-subject prediction of affective states. During fMRI scanning, we showed participants blocks of affective pictures of varying degrees of valence and arousal value and asked them to appraise the affective information conveyed by the pictures. Our stimulus set differed from those used in previous studies in that it was designed along a dimensional model of

\* E-mail: k.yuen@uke.de

emotions and controlled carefully for both valence and arousal. We tested whether the evoked fMRI activation patterns accurately predicted the emotionality of the stimuli (positive vs. negative affect and high vs. low arousal values). We also assessed whether these two dimensions are independent of each other using cross-validation training and classification.

## Materials and Methods

### Participants

Twelve university students (6 female) participated in the study. Their average age was 27.17 (s.d. = 11.57), all right-handed. Among them five participants (3 female) performed two scanning sessions for Study 1; the remaining seven participants only performed one scanning session for Study 2. Subjects reported no history of emotional or neurological impairment. The study was approved by the Research Ethics Committee of the School of Psychology, Bangor University. Written informed consent was obtained from all participants.

### Stimuli and fMRI Paradigm

Two hundred and forty pictures from the International Affective Pictures System (IAPS) were used in the experiment [18]. 192 of the pictures fall into a 2x2 Valence x Arousal design, with 48 pictures in each cell (negative-valence low-arousal, negative-valence high-arousal, positive-valence low-arousal, positive-valence high-arousal). The remaining 48 pictures were neutral filler stimuli which were presented in between the affective blocks as a neutral anchor to wash out the mood induction effect. Hence the neutral picture set was not included in the final analysis (see Figure 1 for the plot of valence and arousal of the selected stimuli).

Previous studies have suggested that basic visual properties, for example spatial frequency, are a crucial factor modulating affective responses in the brain [19]. Therefore we employed the Delplanque et al. [20] method to compare the high and low spatial frequency bands of the three groups of emotional stimuli. An analysis of variance procedure on the low level visual properties of the stimuli (i.e.

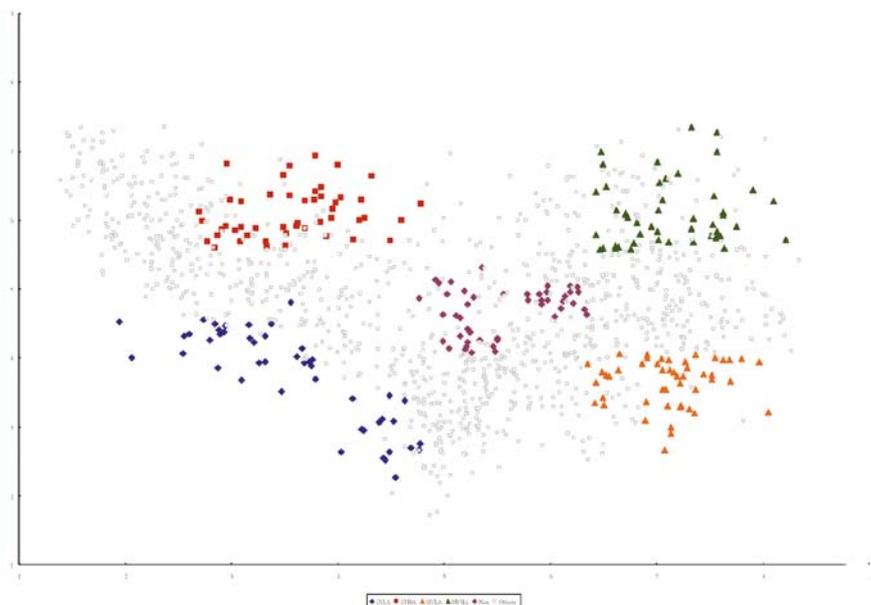
spatial frequency) suggested that the three groups were matched in both low and high spatial frequency bands,  $F(2, 239) = 0.586$ ,  $p > 0.05$  &  $F(2, 239) = 0.037$ ,  $p > 0.05$  respectively. Higher level visual properties, for example the presence of faces, animals, objects and natural scenes, were matched by proportion in each group (Chi-square=0.225,  $p > 0.05$ ). The stimuli were then split into two independent stimulus subsets by randomly dividing a subgroup of the pictures of similar content (e.g. portrait of persons, animals, objects, scenes of nature) into two. One half of the stimuli was presented in the odd runs during fMRI scanning for feature selection and Support Vector Machine (SVM) classifier training (the 'training' set), with the remaining subset presented in even runs and was used for cross-validation of the SVM classifier (the 'test' set). This approach, of separate training and test datasets, is important to avoid the over-inflation of classifier accuracy by 'double-dipping' [21,22].

The experiment used a block design to present the stimuli. Each block consisted of 4 pictures of similar valence value, each presented for 1.5 sec, followed by a 16s fixation period. There were 24 blocks of affective

pictures (6 blocks per condition) and 6 blocks of neutral pictures in each run. Block sequence was pseudo-randomized to control for history effects. Participants were instructed to focus on the emotional content of the pictures and were instructed to feel the emotion conveyed in each picture. This form of instruction has been shown to successfully induce active affective states in previous emotion regulation studies [23]. Study 1 and Study 2 had identical paradigm and scanning protocol and only differed in the number of scanning runs collected for each participant.

### Data Acquisition and Preprocessing

Whole-brain fMRI data were acquired on a 3T Philips Achieva MR Scanner equipped with an 8-channel SENSE head coil. Stimuli were presented using a computer running E-Prime v1.3 (Psychology Software Tools, Philadelphia, USA). An LCD video projector was used to back project the visual stimuli onto a projection screen, situated at the head end of the scanner, which the participant viewed through a mirror mounted above the head coil. We used an echo-planar imaging (EPI) sequence to detect the Blood Oxygen Level



**Figure 1.** Scatterplot of the 240 selected pictures from the whole IAPS stimulus set. Y-axis represents arousal; x-axis represents valence. The five stimulus groups are: negative valence low arousal (blue diamond), negative valence high arousal (purple square), positive valence low arousal (yellow triangle), positive valence high arousal (red triangle), and neutral (green diamond). Data points in grey represent the unselected IAPS pictures.

Depend (BOLD) signal during the experimental task. Each brain volume consists of 28 axial slices covering the whole brain. Imaging parameters were TR=2000ms; TE=35ms; flip angle=90; matrix=96x96; FOV=240x240; slice thickness=3mm, resulting in a voxel size of 2.5x2.5x3mm<sup>3</sup>. In addition, a three-dimensional high resolution T1-weighted anatomical image (isotopic voxel of 1mm<sup>3</sup>) was obtained for functional to anatomical image registration.

BrainVoyager QX (ver. 2.2, Brain Innovation, Maastricht, The Netherlands) was used to preprocess and analyze the functional images. The EPI data of each functional run were motion corrected (sinc interpolated), coregistered with the T1 anatomical image and transformed to volume timecourse (VTC) files in the Talairach stereotaxic space [24] with an isotropic voxel size of 3x3x3mm<sup>3</sup>. VTC files were then detrended to remove signal drift. No spatial or temporal smoothing was performed in order to increase signal sensitivity for MVPA analysis.

### Multivoxel Pattern Classification Study 1: Intra-individual Classification

The general flow of data analysis for Study 1 is presented in Figure 2. Voxel-wise beta estimates for each condition in each run were obtained by a general linear model (GLM). The predictors for GLM were generated by convolving the stimulus presentation, on-off, period with a double-gamma hemodynamic reference function [25]. The GLM predictors were then regressed against the BOLD signal at every voxel. Regressor coefficients (betas) from the set of selected voxels were combined to form the multivoxel patterns for classification. With our 2x2 Valence by Arousal design, the same block of functional scans will be used twice, once with the label positive or negative valence and once with the label high or low arousal.

Following the computation of beta estimates we used the data from the training set to perform feature selection for the training of the SVM classifier. Feature selection was performed in two stages: we first created a whole-brain mask by computing the contrast of all conditions against fixation, which was orthogonal to our later analyses of category-specific response. Since we are interested in the active affective

functioning of the brain, basic visual processing of the stimuli may act as a confounding factor in later MVPA analysis. The visual cortex was thus excluded from the analysis by creating an anatomical mask posterior to the occipital-parietal sulcus, individually defined (Talairach coordinate Y=-54 to -62). Voxels surviving the threshold of our univariate analysis were then further selected using the multivariate Recursive Feature Elimination (RFE) procedure [26]. This method allows us to select the sets of voxels with highest discriminative power in whole-brain SVM classification without *a priori* assumptions on the specific regions of interests. It has advantages over the searchlight approach by considering all voxels *in parallel*, where the searchlight approach only considers

one small region of neighboring voxels at a time to unravel its information content. With RFE it is possible to detect patterns of brain activities distributed over a wide-spread network in the brain. The RFE algorithm has already been successfully applied in the decoding of visual and auditory processes [26-28].

We performed the RFE procedure on the data from the training set twice, once with each block labeled for positive or negative valence and the other time with each block labeled for high or low arousal. For each subject we had a total of 48 trials per condition (i.e. 48 samples for each valence and arousal level). The iterative RFE procedure trained a least-squares Support Vector Machine (ls-SVM) classifier using the top 50% of voxels (5000-8000 voxels) selected

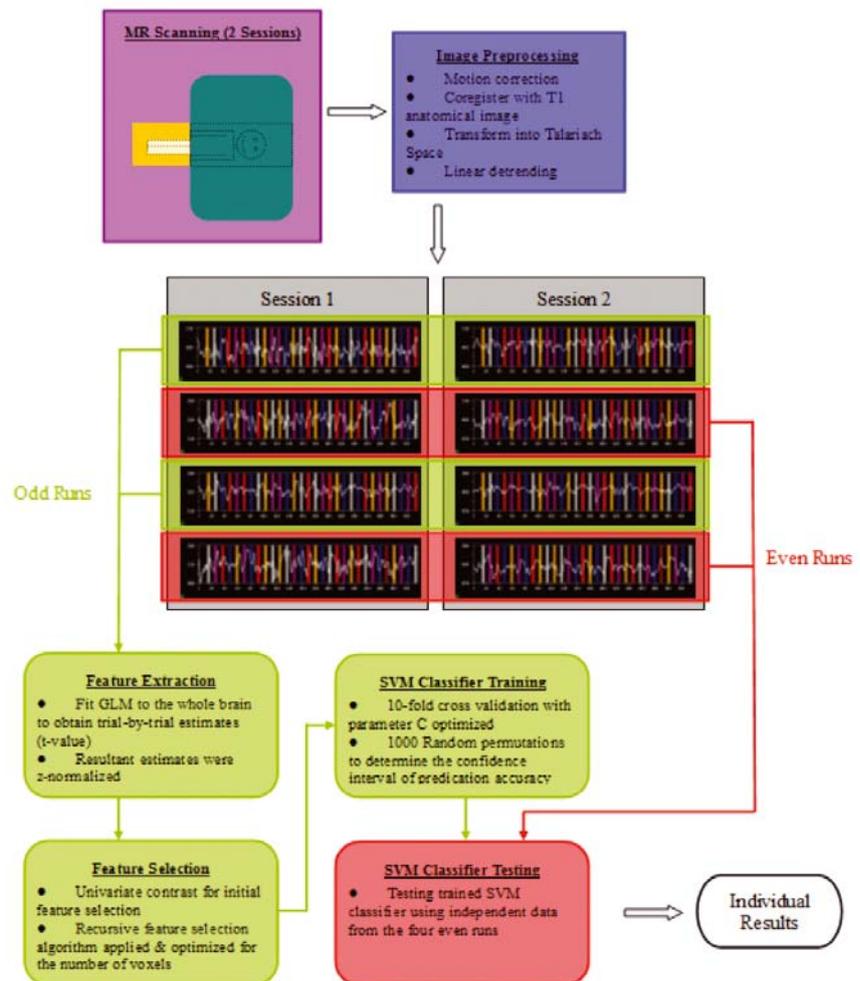


Figure 2. Schematic flow chart of data collection and data analysis. Each subject underwent two scanning sessions. The stimulus protocols are color coded for different conditions: grey = neutral; yellow, red = positive valence with low & high arousal; purple, blue = negative valence with low & high arousal.

based on the preceding univariate analysis. The 96 trials from the two conditions were split into two subsets for the training and classification to obtain the discrimination maps for the two valence/arousal conditions. Voxels with the smallest discriminative ranking were discarded and the remaining voxels were used for training in the next iteration. Ten iterations were performed with 10-fold cross validation on different subsets of data. We tested the optimal RFE parameters by varying the number of voxels eliminated [15] and the best result was obtained when 5% (487-641 voxels) of the most discriminative voxels were used.

We then trained two SVM classifier with these RFE maps, one for valence labeling and one for arousal labeling and tested their predictive ability on the independent 'test' dataset, collected from the even runs. Not only did we use separate runs for the final cross-validation, the stimuli presented in the odd and even runs were also different from each other. This ensures that we avoid inflated prediction accuracy due to non-orthogonal, double-dipping procedures [21,22] or physical stimulus similarities. To determine the statistical significance of the predictive accuracy of the discovered pattern in each individual subject, we ran 1000 permutations of a SVM classification by randomly assigning the stimulus labels to our data to determine the null distribution of predictive accuracy.

### Study 2: Inter-individual Classification

Procedures for image preprocessing and feature extraction were identical to Study 1. With the RFE map obtained from Study 1 used as a mask, we implemented a bootstrapping procedure to obtain the distribution of classification accuracies for each individual participant with data from the other participants as the training set. For an illustration, all four functional runs from Subject 1 are used as the testing data for SVM classification and 16 functional runs were randomly selected from Subject 2 to 7 to train the SVM classifier (Figure 3). This random sampling of the training data set was permuted for 100 times for each subject. We then determined confidence interval for each participant's classification accuracy. To further characterize participants' inter-individual

variability, we also examine the correlation between classification performance and their subjective perception on the differences of valence ratings of presented IAPS pictures.

## Results and Discussion

Using the optimal RFE parameters we trained the valence classifier (positive vs negative valence) and the arousal classifier (high vs low arousal) in Study 1. The valence classifier performed significantly better than chance level for all subjects ( $\alpha < 0.05$  for 4 subjects,  $\alpha < 0.1$  for 1 subject; group average=62.35%,  $t(4)=7.213$ ,  $p=0.002$ ; Kolmogorov-Smirnov test of normality  $Z=0.654$ ,  $p > 0.05$ , normal distribution assumed, Figure 4). Conversely, the arousal classifier only performed at chance level for all subjects ( $\alpha > 0.05$  for all subjects, group average=54.38%,  $t(4)=1.972$ ,  $p=0.120$ ). To further verify whether the RFE selected voxels for the valence classifier contain only valence information, we also use valence-based

RFE features to train an additional arousal classifier with the data from the training set. We then performed cross-validation with the data from the test set for this valence-feature based arousal classifier. This cross-validation yields chance-level predictive accuracy ( $\alpha > 0.05$  for all subjects, group average=53.18%,  $t(4)=1.637$ ,  $p=0.177$ ), suggesting the neural representation of valence dimension as identified by the RFE feature selection procedure does not contain significant levels of arousal information.

Our findings suggested that it is possible to predict the valence of stimuli based on the patterns of neural activation that they evoke. Brain patterns associated with different levels of arousal, on the other hand, failed to provide sufficient information for such predictions. Such differences may suggest that our participants had placed a heavier weight on the valence information during the appraisal processes of the affective stimuli. In fact, it is possible that the appraisal of valence and arousal dimensions

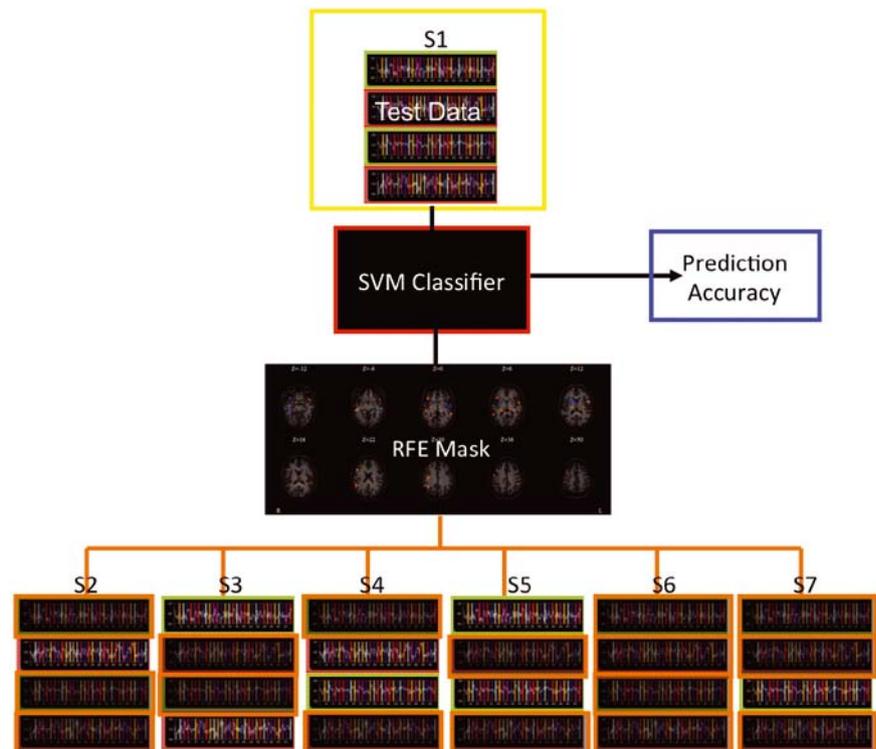


Figure 3. Schematic presentation of the inter-individual permutation classification. For each individual the whole data set (4 fMRI runs) are used as the test data; SVM classifier was trained using a random sampling of 16 fMRI runs from the other subjects. BOLD patterns were limited to a mask obtained from Study 1. This process was repeated for 100 times in each subject to obtain a distribution of classification accuracies.

is simultaneous and correlated with each other. In naturally occurring scenes that trigger an extremely positive or extremely negative valence representation in an organism's brain, for example situations of aggression or sexual stimulation, an extreme state of arousal is often being triggered at the same time. Situations with intermediate levels of arousal, like stimuli being used in the current study, compared with situations with extremely high or extremely low arousal levels, like naturally occurring events, may be interpreted with a different scale and thus result in a reduced sensitivity for our MVPA arousal classifier to pick up the differences.

In order to investigate the common brain regions identified as containing important voxels by the RFE algorithm, we generated group-level activation probability maps from individual feature weight maps for the valence classifier (Figure 5 & Table 1; group probability maps for the arousal classifier were not presented due to its chance-level performance in prediction accuracy.). These maps reflect the spatial overlap of discriminating voxels across participants. Valence information was represented across a distributed brain network that included inferior parietal lobe, posterior cingulate, lateral and mesial temporal cortices, anterior cingulate and medial prefrontal cortex, dorsolateral prefrontal cortex, inferior frontal

gyrus and insula, and thalamus and striatum. This is largely consistent with previous findings in the emotion perception, emotion regulation, and valence learning literature [29,30]. Conversely, our univariate analysis yielded no significant activation under the same threshold of activation probability, demonstrating the

value of this method for use in emotion-based tasks.

The brain regions identified from group probability maps could be considered as the neural signatures for decoding mental states associated with positive and negative valence. While the neural signatures revealed

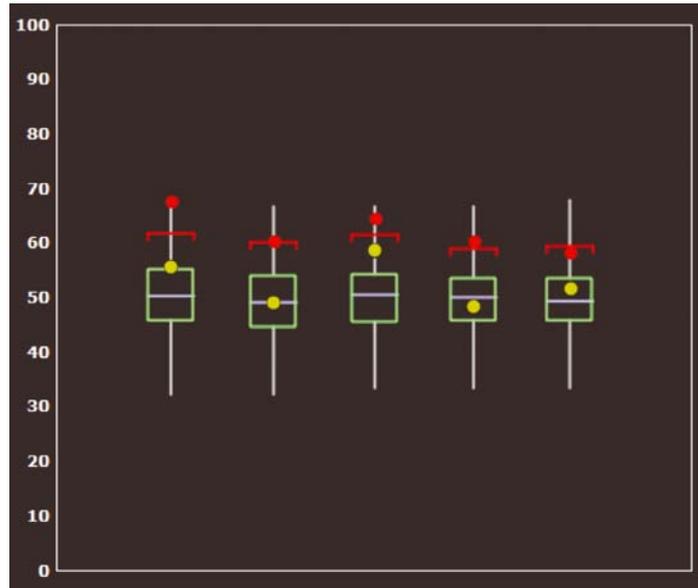


Figure 4. Prediction accuracies across all five subjects. The results suggest that cross-validation accuracies for the valence classifier (red dots) were significantly higher than chance level. However, the arousal classifier only predicts at chance level (yellow dots). Green boxes: 25<sup>th</sup>-75<sup>th</sup> percentile of the prediction accuracy distribution obtained from individual permutation tests; red brackets: 95<sup>th</sup> percentile of the prediction accuracy distribution.

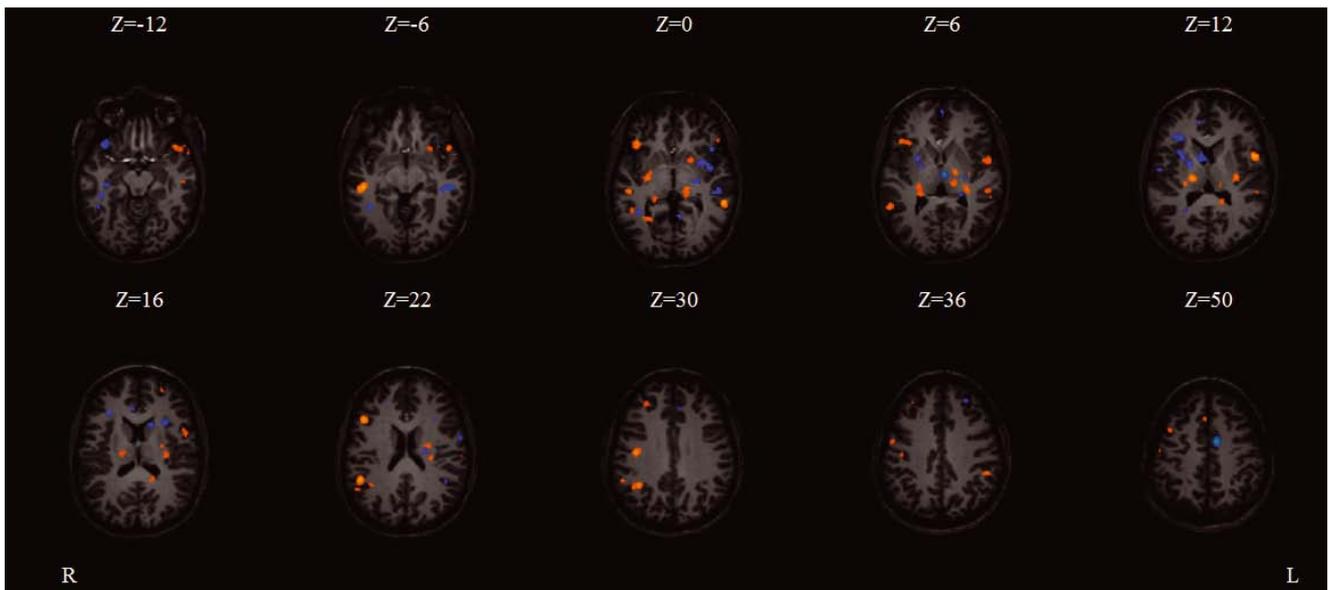


Figure 5. Brain maps of the important features for valence prediction across the five subjects. The probabilistic map is thresholded at 50% of subjects. Voxels coded in hot color indicate that they are discriminative for positive valence stimuli while voxels coded in cold color indicate that they are discriminative for negative valence.

by this analysis are not surprising compared to reported data in the literature, the use of MVPA indeed provided a higher level of sensitivity compared to a conventional univariate analysis as shown above, where no significant group activation was observed at the same threshold of activation probability (see Supplementary Information).

A recent study had shown that inter-subject classifications of pathological motivational states could be successfully performed using regional fMRI BOLD signals [31]. We therefore tested whether our method of extracting the neural signatures associated with different valence representation could be applied to inter-subject classification of affective states. We tested whether the RFE map obtained from Study 1 could be used for inter-individual classification. In Study 2 we recruited an independent sample of seven subjects to test this possibility. With permutation analysis we obtained inter-individual classification accuracies for a given subject based on a random sampling of fMRI data from the remaining sample. Results revealed that in 4 out of 7 subjects the classifier performance is significantly better than chance level (see Figure 6a); the average classification accuracy across the group as a whole is marginal but significantly better than chance (95% C.I. = 50.14 to 51.02). We further explored inter-individual variations in classification by correlating their subjective ratings on the IAPS and classifier performance, and a positive trend was observed,  $r=0.6813$ ,  $p<0.1$  (Figure 6b). Larger samples will be needed to assess the possibility of inter-subject classification conclusively, but the positive correlation trend may explain some of the factors that contribute to the successful classification of individual perceptual states.

A limitation of this study is that we did not assess the emotional states of participants on a block-by-block basis but relied on the ratings of normative samples. Our short (6 second) blocks may not have been sufficient to induce substantial changes in participants' affective states. However, the success of the classification procedure (at least for valence information) suggests that the patterns evoked in the brain by the different classes of stimuli were

Table 1. Discriminative brain regions identified in the group analysis.

Neural Activations	L/R	BA	x	y	z
<b>Positive weight w</b>					
Lingual Gyrus	R	19	29	-57	0
Inferior Parietal Lobe	L	40	-50	-40	37
	R	40	52	-46	24
Posterior Cingulate	L	29	-15	-44	14
Middle Temporal Gyrus	L	22	-56	-40	0
	R	21	56	-48	6
Superior Temporal Gyrus	L	22	-49	-21	-10
	L	41	-53	-31	8
	R	39	39	-52	28
	R	39	54	-56	23
	R	21	49	-27	-4
Parahippocampal Gyrus	R	19	46	-47	-1
Precentral Gyrus	L	44	-52	6	11
	R	6	40	-15	32
	R	6	55	-5	36
	R	4	53	-14	42
Middle Frontal Gyrus	L	10	-25	53	18
	R	6	42	9	46
	R	9	30	37	32
	R	46	47	21	22
Inferior Frontal Gyrus	L	47	-45	16	-10
	L	47	-49	30	-2
	R	47	41	24	3
Medial Frontal Gyrus	R	8	3	22	44
Putamen	L		-26	17	-4
	L		-30	-18	14
	L		-19	8	2
	L		-23	-8	20
	R		29	-11	-1
Thalamus	L		-14	-25	3
	L		-15	-10	7
	L		-28	-28	5
	R		23	-30	6
	R		18	-17	13
<b>Negative weight w</b>					
Superior Temporal Gyrus	L	38	-55	10	-9
Middle Temporal Gyrus	L	21	-47	-27	-4
Hippocampus	R		36	-25	-12
Parahippocampal Gyrus	L	30	-22	-36	6
	R	30	25	-52	10
	R	19	40	-49	-6
	R	36	41	-36	-13
Precentral Gyrus	L	6	-59	2	20
	R	43	54	-9	13
Middle Frontal Gyrus	L	9	-27	41	37
Inferior Frontal Gyrus	L	47	-43	20	1
	R	47	37	21	-12
Anterior Cingulate	L	32	-7	35	27
	L	24	-10	-3	44
	R	24	7	33	15
Medial Frontal Gyrus	L	10	-2	55	6
	R	9	22	42	24
	R	10	10	45	13
Insula	L	13	-44	-48	20
	L	13	-30	19	17
	L	13	-35	3	1
	R	13	48	10	2
	R	13	33	28	13
Caudate	L		-14	15	17
	R		7	5	12
Putamen	R		25	3	10
	L		-21	-14	23
Globus Pallidus	L		-26	-16	3
Thalamus	L		-4	-13	5

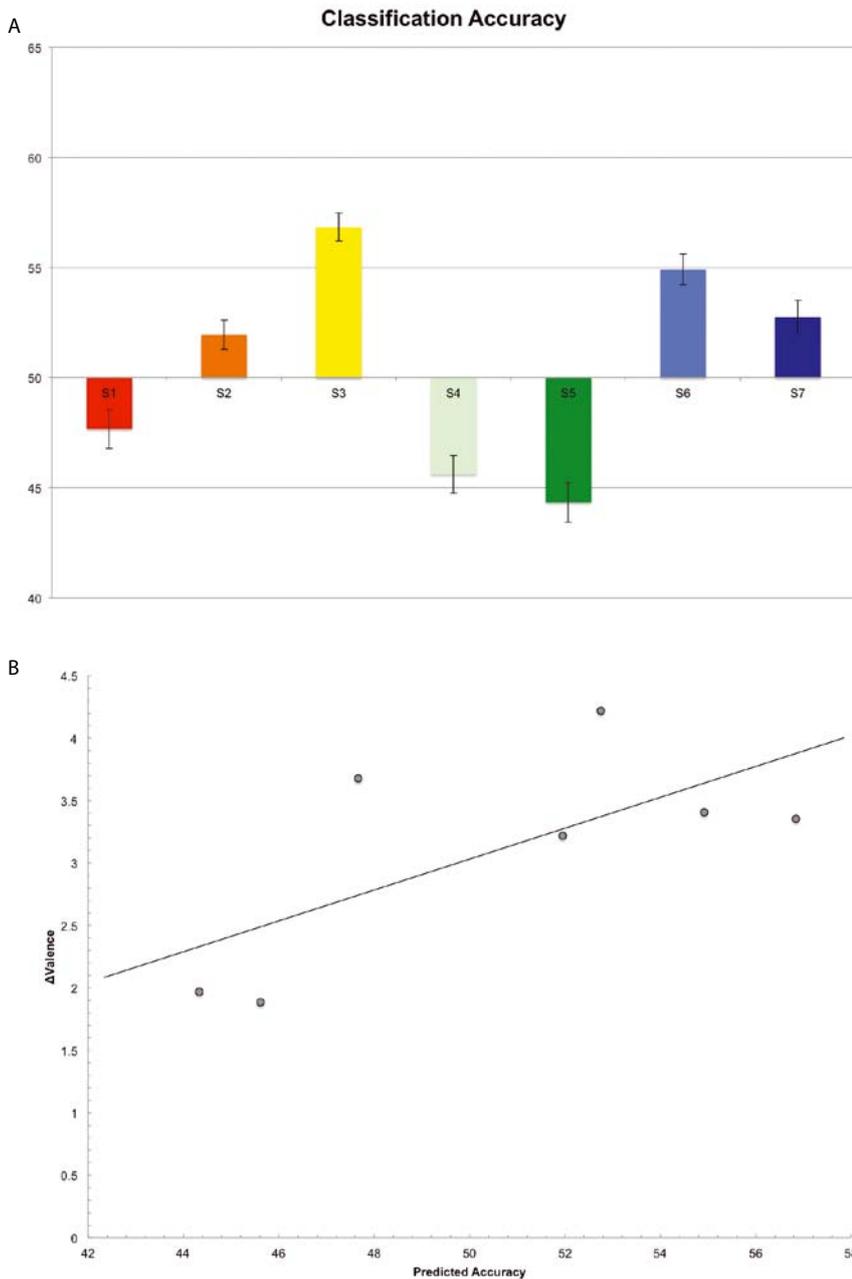


Figure 6. Inter-individual classification accuracies (a) and their correlation with individual's subjective ratings on the IAPS. We computed the averaged difference ( $\Delta$ Valence) of valence ratings between the positive and negative valence pictures for the correlation.

still distinguishable. However, longer block durations may be required in order to bring out the differences between high and low arousing stimuli.

Another limitation that should be noted in this study is that during stimulus preparation we

have excluded pictures of extreme valence and arousal to achieve a balanced factorial design. In the initial conceptualization such design should give us an advantage of comparing valence and arousal with a similar range of ratings, however it may also limit the classifier's

ability to use arousal information to do classification because stimuli with intermediate level of arousal may be appraised in a different manner compared to the extremely arousing stimuli. We have yet to use stimuli with more extreme arousal level to see whether arousal level will add information for the decoding of individual's active affective state.

### Conclusion

We aimed to decode affective representations from brain activation patterns using MVPA. We used a paradigm known to trigger active affective appraisal using pictorial stimuli. Our findings suggest that it is possible to predict the category of induced affective representations in an individual, using the brain activity patterns associated with different valence. The valence dimension was decoded better than the arousal dimension, but further validation with stimuli attracting extreme valence/arousal ratings is needed. Although our classification results were around 65% and thus lower than levels of accuracy reported for classifier performance in visual perception tasks [4], our results are comparable to studies on higher level cognitive processes, e.g. decision making [7,8]. In view of the abstraction of emotional information from the visual input, the use of independent stimulus set for training and classification, and added effects of inter-individual variability in associating the visual stimuli with subjective emotional experience, our classifier performed reasonably well in decoding individuals' emotional perception in the valence dimension.

Most MVPA studies with the aim of performing inter-subject classification of normal versus abnormal states have used imaging modalities other than functional activations (e.g. regional volumetric information and resting-state connectivity between different brain regions) as the input pattern. Conversely MVPA studies using fMRI signals mainly classify different mental states in a within-subject manner. The highly dynamic nature of fMRI activations between different individuals poses a challenge for inter-subject classifications. Nonetheless, recent studies suggested that with the

significant neural signature identified, it is possible to perform group classification with fMRI BOLD signals as well [12,31,32]. With our current method of whole-brain MVPA classification, which has better sensitivity compared to univariate analysis, it is possible to implement a two-stage classification approach: we can first use the RFE approach on whole-brain signals to identify the most important neural signatures corresponding to a cognitive/affective state that is linked to the abnormality, and use BOLD signals from such brain regions to perform second level inter-subject classifications. The possibility to classify inter-subject brain patterns will enable clinicians to perform differential diagnosis on

conditions like dementia and depression. Since we do not normally expect individuals to control their own patterns of brain activity without training, it could also be applied in a forensic setting to determine the mental status of a patient more objectively.

In conclusion, we have demonstrated that affective representations with positive and negative valence can be predicted using patterns of fMRI BOLD signals. With this sensitive methodology as a building block, we propose a future two-stage approach to distinguish individuals with abnormal affective states from healthy individuals, using fMRI patterns. Such work could have great potential in assisting clinicians in diagnosis and prognosis.

## Acknowledgements

The study was supported by Wales Institute of Cognitive Neuroscience internal project funds. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We would also like to thank Dr. Raffael Kalisch for his comments on earlier version of the manuscript.

## Conflict of Interest

The authors declare no conflict of interest.

## References

- [1] Cox D. D., Savoy R. L., Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex, *Neuroimage*, 2003, 19, 261-270
- [2] Haynes J. D., Rees G., Decoding mental states from brain activity in humans, *Nat. Rev. Neurosci.*, 2006, 7, 523-534
- [3] Haxby J. V., Gobbini M. I., Furey M. L., Ishai A., Schouten J. L., Pietrini P., Distributed and overlapping representations of faces and objects in ventral temporal cortex, *Science*, 2001, 293, 2425-2430
- [4] Kamitani Y., Tong F., Decoding the visual and subjective contents of the human brain, *Nat. Neurosci.*, 2005, 8, 679-685
- [5] Shinkareva S. V., Mason R. A., Malave V. L., Wang W., Mitchell T. M., Just M. A., Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings, *PLoS One*, 2008, 3: e1394
- [6] Lewis-Peacock J. A., Postle B. R., Temporary activation of long-term memory supports short-term memory, *J. Neurosci.*, 2008, 28, 8765-8771
- [7] Hampton A. N., O'Doherty J. P., Decoding the neural substrates of reward-related decision making with functional MRI, *Proc. Natl. Acad. Sci. USA*, 2007, 104, 1377-1382
- [8] Soon C. S., Brass M., Heinze H. J., Haynes J. D., Unconscious determinants of free decisions in the human brain, *Nat. Neurosci.*, 2008, 11, 543-545
- [9] Fung G., Stoeckel J., SVM feature selection for classification of SPECT of Alzheimer's disease using spatial information, *Knowl. Inf. Syst.*, 2007, 11, 243-258
- [10] Gerardin E., Chételat G., Chupin M., Cuingnet R., Desgranges B., Kim H. S., et al., Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging, *Neuroimage*, 2009, 47, 1476-1486
- [11] Koutsouleris N., Meisenzahl E. M., Davatyikos C., Bottlender R., Frodl T., Scheuerecker J., et al., Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition, *Arch. Gen. Psychiatry*, 2009, 66, 700-712
- [12] Hahn T., Marquand A. F., Ehlis A., Dresler T., Kittel-Schneider S., Jarczok T. A., et al., Integrating neurobiological markers of depression, *Arch. Gen. Psychiatry*, 2010, 68, 361-368
- [13] Pessoa L., Padmala S., Decoding near-threshold perception of fear from distributed single-trial brain activation, *Cereb. Cortex*, 2007, 17, 691-701
- [14] Rolls E. T., Grabenhorst F., Franco L., Prediction of subjective affective state from brain activations, *J. Neurophysiol.*, 2009, 101, 1294-1308
- [15] Ethofer T., De Ville D. V., Scherer K., Vuilleumier P., Decoding of emotional information in voice-sensitive cortices, *Curr. Biol.*, 2009, 19, 1028-1033
- [16] Sitaram R., Lee S., Ruiz S., Rana M., Veit R., Birbaumer N., Real-time support vector classification and feedback of multiple emotional brain states, *Neuroimage*, 2010, 56, 753-765
- [17] Peelen M. V., Atkinson A. P., Vuilleumier P., Supramodal representations of perceived emotions in the human brain, *J. Neurosci.*, 2010, 30, 10127-10134
- [18] Lang P. J., Bradley M. M., Cuthbert B. N., International affective picture system (IAPS): Affective ratings of pictures and instruction manual, Technical Report A-8, University of Florida, Gainesville, FL, 2008
- [19] Alorda C., Serrano-Pedraza I., Campos-Bueno J. J., Sierra-Vazquez V., Montoya P., Low spatial frequency filtering modulates early brain processing of affective complex pictures, *Neuropsychologia*, 2007, 45, 3223-3233
- [20] Delplanque S., N'diaye K., Scherer K., Grandjean D., Spatial frequencies or emotional effects? A systematic measure of spatial frequencies for IAPS pictures by a discrete wavelet analysis, *J. Neurosci. Methods*, 2007, 165, 144-150

- [21] Kriegeskorte N., Simmons W., Bellgowan P., Baker C., Circular analysis in systems neuroscience: the dangers of double dipping, *Nat. Neurosci.*, 2009, 12, 535-540
- [22] Vul E., Harris C., Winkelman P., Pashler H., Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition, *Perspect. Psychol. Sci.*, 2009, 4, 274-291
- [23] Mak A. K. Y., Hu Z. G., Zhang J. X. X., Xiao Z., Lee T. M. C., Sex-related differences in neural activity during emotion regulation, *Neuropsychologia*, 2009, 47, 2900-2908
- [24] Talairach J., Tournoux P., Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system : an approach to cerebral imaging, Stuttgart, Thieme Medical Publishers, 1988
- [25] Boynton G. M., Engel S. A., Glover G. H., Heeger D. J., Linear systems analysis of functional magnetic resonance imaging in human V1, *J. Neurosci.*, 1996, 16, 4207-4221
- [26] De Martino F., Valente G., Staeren N., Ashburner J., Goebel R., Formisano E., Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns, *Neuroimage*, 2008, 43, 44-58
- [27] Formisano E., De Martino F., Bonte M., Goebel R., "Who" is saying "What"? Brain-based decoding of human voice and speech, *Science*, 2008, 322, 970-973
- [28] Staeren N., Renvall H., De Martino F., Goebel R., Formisano E., Sound categories are represented as distributed patterns in the human auditory cortex, *Curr. Biol.*, 2009, 19, 498-502
- [29] Adolphs R., Neural systems for recognizing emotion, *Curr. Opin. Neurobiol.*, 2002, 12, 169-177
- [30] Phillips M. L., Drevets W. C., Rauch S. L., Land R., Neurobiology of emotion perception I: the neural basis of normal emotion perception, *Biol. Psychiatry*, 2003, 54, 504-514
- [31] Ihssen N., Cox W. M., Wiggett A., Fadardi J. S., Linden D. E., Differentiating heavy from light drinkers by neural responses to visual alcohol cues and other motivational stimuli, *Cereb. Cortex*, 2011, 21, 1408-1415
- [32] Mourão-Miranda J., Oliveira L., Ladouceur C. D., Marquand A., Brammer M., Birmaher B., et al., Pattern recognition and functional neuroimaging help to discriminate healthy adolescents at risk for mood disorders from low risk adolescents, *PLoS One* 2012, 7: e29482

## Supplementary Information

### Comparing sensitivity between Univariate Analysis versus Multivoxel Pattern Analysis

In order to compare the sensitivity between univariate analysis and MVPA, we had adopted similar imaging processing parameters in creating the group maps for both analysis. The raw EPI images were smoothed with a 4mm

FWHM Gaussian kernel, comparable with the spatial smoothing we had performed on the RFE maps. Both analysis utilized activation probability mapping to observe the group effect. When we thresholded the univariate analysis at 60% chance of overlap, no significant group activation was observed. Nevertheless, significant clusters of discriminative voxels were still observable on the

RFE group map (Figure S1). More importantly, the clusters of discriminative voxels observed on the RFE group map is consistent with previous reports of neural circuitry responsible for emotional perception [1,2]. This observation suggested that the use of whole-brain RFE procedure provides a higher level of sensitivity compared to univariate analysis.

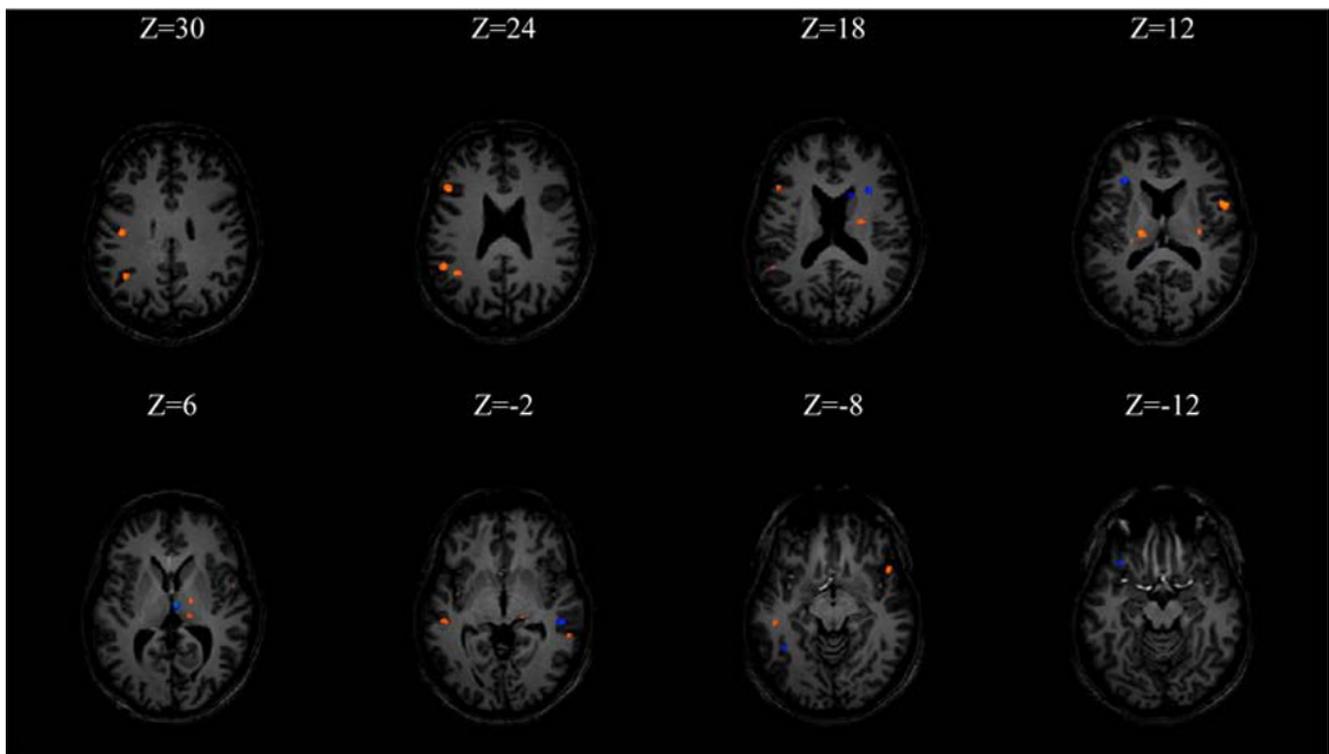


Figure S1. Group activation probability map showing valence-sensitive brain regions as revealed by the MVPA analysis thresholded at 60%. With the same threshold, the univariate group probability map showed no activations.

## References

- [1] Adolphs R., Neural systems for recognizing emotion, *Curr. Opin. Neurobiol.*, 12, 169-177
- [2] Phillips M. L., Drevets W. C., Rauch S. L., Land R., Neurobiology of emotion perception I: the neural basis of normal emotion perception, *Biol. Psychiat*, 54, 504-514