Classification images reveal the information sensitivity of brain voxels in fMRI

Fraser W. Smith, Lars Muckli, David Brennan, Cyril Pernet, Marie L. Smith, Pascal Belin, Frederic Gosselin, Donald M. Hadley, Jonathan Cavanagh, Philippe G. Schyns

PII: S1053-8119(08)00039-6
DOI: doi: 10.1016/j.neuroimage.2008.01.029
Reference: YNIMG 5186

To appear in: *NeuroImage*

Received date: 12 July 2007
Revised date: 2 November 2007
Accepted date: 5 January 2008

Please cite this article as: Smith, Fraser W., Muckli, Lars, Brennan, David, Pernet, Cyril, Smith, Marie L., Belin, Pascal, Gosselin, Frederic, Hadley, Donald M., Cavanagh, Jonathan, Schyns, Philippe G., Classification images reveal the information sensitivity of brain voxels in fMRI, *NeuroImage* (2008), doi: 10.1016/j.neuroimage.2008.01.029

Classification Images Reveal the Information Sensitivity of Brain Voxels in fMRI

Fraser W. Smith*, Lars Muckli*, David Brennan~, Cyril Pernet*, Marie L. Smith*, Pascal Belin*, Frederic Gosselin^, Donald M. Hadley~, Jonathan Cavanagh', & Philippe G. Schyns*

*Centre for Cognitive Neuroimaging, Department of Psychology, University of Glasgow, Glasgow, UK

~Institute of Neurological Sciences, Southern General Hospital, Glasgow, UK

'Sackler Institute of Psychobiological Research, Division of Community Based Sciences, University of Glasgow, Southern General Hospital, Glasgow, UK

^Departement de psychologie, Universite de Montreal, Montreal, Canada

Correspondence should be addressed to FWS (e-mail: fraser@psy.gla.ac.uk).

Department of Psychology, 58 Hillhead Street, University of Glasgow, Glasgow, UK. G12 8QB.

Phone: +44 (0)141 330 5003; Fax: +44 (0)141 330 4606.

ABSTRACT

Reverse correlation methods have been widely used in neuroscience for many years and have recently been applied to study the sensitivity of human brain signals (EEG, MEG) to complex visual stimuli. Here we employ one such method, Bubbles (Gosselin & Schyns, 2001), in conjunction with fMRI in the context of a 3AFC facial expression categorization task. We highlight the regions of the brain showing significant sensitivity with respect to the critical visual information required to perform the categorization judgments. Moreover, we reveal the actual subset of visual information which modulates BOLD sensitivity within each such brain region. Finally, we show the potential which lies within analyzing brain function in terms of the information states of different brain regions. Thus, we can now analyse human brain function in terms of the specific visual information different brain regions process.

The use of reverse correlation methods in neuroscience, and in neurophysiology in particular, has a long history (see Ringach & Shapley, 2004). Recently similar methods have been applied by researchers studying high-level vision in humans and non-human primates (Adolphs et al., 2005; Gosselin & Schyns, 2001; Sekuler et al, 2004; Smith et al., 2004; Neilsen et al., 2006; 2007). The great asset of these methods is that they provide a fine grained representation of the stimulus information which is optimal for some response function: for example, for a behavioural decision or that which tunes a given population of neurons. On each trial of an experiment observers are presented with randomly distorted information while the experimenter measures a given dependent variable such as the correct performance of an observer, or the amplitude of a brain signal. The given dependent variable is then reverse correlated with the stimulus information presented on each trial. Reverse correlation methods can depict the information subspace that was most effective for a given brain region (or a certain behaviour). This contrasts with the more common forward correlation methods that detect brain regions (or behaviour) correlating with significant variation of the inputs (e.g. different object categories).

Recently, a particular instantiation of reverse correlation methods (i.e. Bubbles) has been applied to the analysis of human brain signals such as EEG (Smith et al., 2004; 2006; 2007a; Schyns et al., 2003, 2007) and MEG (Smith et al., 2007b). Bubbles is an information sampling technique that presents sparse samples of stimulus information to an observer--for example, the image is randomly sampled through Gaussian apertures either in the 2D image plane or in a 3D space encompassing both the image plane plus spatial frequency bands (see Figure 1, and Gosselin & Schyns, 2001). By reverse correlating observers' performance with the information samples Bubbles determines the information subspace that is diagnostic for a particular behavioural decision, or that which is correlated with the modulations of a particular brain signal. For example, Smith et al. (2004) have shown, that it is possible to characterise the evolution of brain sensitivity to face information over the time course of the N170 (eye sensitive stage) and P300 (task sensitive stage) ERP components and further have found that different diagnostic portions of the spatial frequency spectrum correlate with different EEG temporal frequency bands

in stimulus perception (Smith et al., 2006). The method has also been used to demonstrate that the N170 integrates visual information from the eyes of a face downwards, terminating when the diagnostic feature relevant to the current judgement is reached (Schyns et al., 2007).

Building on these lines of research, we sought to apply Bubbles to the analysis of fMRI data. Standard methods of fMRI data analysis involve the contrast of activation patterns obtained from a small number of experimental conditions of interest, relying essentially on a subtractive method. Such methods do not have much power to resolve questions regarding the fine-grained response properties of given voxels in the human brain, beyond a basic correlation of a given voxel with a given experimental condition. More recent analysis methods (e.g. Kamitani & Tong, 2005; Haynes & Rees., 2005; Haxby et al., 2001; Kriegeskorte et al., 2006) have shown that it is possible to find reliable brain sensitivity to specific types of information (e.g. visual category, orientation) that is distributed weakly across many voxels, when the approach is multivariate. Hence this kind of sensitivity is not evident from standard, univariate, methods of analysis.

The Bubbles method, like the newer multivariate techniques, goes beyond standard methods of fMRI data analysis. The power of Bubbles, however, is to provide a fine-grained description (2D image) of the response properties of each individual voxel (these can be summed across a collection of voxels to represent a region of the brain) with respect to the visual information contained in complex stimuli (hence it is *univariate* in the present implementation). That is, in terms of describing what features of some (reasonably complex) input stimulus (such as a face) correlate with modulations of signal amplitudes at each specific voxel (i.e. to find the "optimal" stimulus for each voxel relative to the given task). Thus, for instance, we might expect a set of brain regions to highly correlate with the presentation of the eyes when participants make fearful judgements to faces (i.e. the diagnostic information), whereas to the mouth when participants make happy judgements (e.g. Smith et al., 2005; Adolphs et al., 2005; Schyns et al., 2007).

It is unclear from work using standard subtractive methods what the regions usually activated in such tasks actually do, in terms of the face information they are sensitive to. Hence

the potential of Bubbles is to provide such a characterisation, an important step to depict the brain as an information processing system (Smith et al., 2007).

Our observers performed a 3AFC expressions categorization task where they had to decide whether each sparsely sampled face (see Figure 1 and Methods) was a happy, fearful or neutral face. We concurrently measured the fMRI BOLD signal elicited. We reverse correlated BOLD amplitudes to information samples, after appropriate preprocessing, to reveal the 'information states' of each voxel in the brain. That is, to reveal the facial information modulating each voxel across different emotional expressions.

Methods

Stimuli

Original face stimuli were gray-scale images of five females and five males taken under standardized illumination, each displaying three facial expressions (happy, fearful, neutral). All 30 stimuli (normalized for the location of the nose and mouth) complied with the Facial Action Coding System (FACS; Ekman & Friesen, 1978), and form part of the California Facial Expressions (CAFE) database (Dailey, Cottrell & Reilley, 2001).

Participants

Two adult subjects (ETS & EGA) with normal (or corrected to normal) vision participated in the study. Both gave informed consent prior to taking part in the experiment. The procedure of the experiment was approved by the local ethics committee in Frankfurt.

Imaging Methods

Participants performed multiple runs of the present experiment (ETS – 20; EGA – 12), with four or more runs collected in each scanning session. Different scanning sessions were performed on different days. During each functional run, we acquired 572 EPI image volumes (17 slices, 3 Tesla Siemens Allegra and Trio, TR=1000ms, TE=30ms, FA=62, 3.1 x 3.1 x 4mm, PACE motion correction, PSF distortion correction) resulting in 11,440 EPI image volumes for ETS and 6,864 for EGA. In addition we acquired a high resolution 3D anatomical reference scan (magnetization-prepared rapid acquisition gradient echo MPRAGE sequence, TR, 2000 ms; TE, 4.38 ms; FA, 15°; FOV, 240; voxel size, $1x1x1mm^3$) for the first scanning session and lower resolution reference scans (3D MPRAGE, TR, 1240 ms; TE, 2.6 ms; voxel size, $1x1x2\ mm^3$) in some of the subsequent scanning sessions, which were later aligned to the first high resolution reference scan. MR Imaging was performed at the Brain Imaging Center in Frankfurt (BIC-Frankfurt).

Experimental Paradigm

Prior to the experimental runs in the scanner, participants had to achieve a criterion level (95%) of performance in expression classification of the original face images used in the experiment. Participants performed multiple runs of the main experiment, where each run consisted of 138 trials. We discarded the first three trials from each run in order to compensate for signal stabilisation. We further controlled the assignment of expression to trials for a history of two previous trials. Within each run each expression was repeated 45 times (excluding the first three trials) across 10 identities, with the repetition of each identity for each expression counterbalanced across couplets of runs. On each trial participants were presented with a sparse version of one face (see Figure 1, visual angle of approximately 7 by 4.5 degrees), which we generated by randomly sampling the 2D image space with Gaussian apertures (sigma=0.28 degrees of visual angle), using the Bubbles technique (Gosselin & Schyns, 2001). A different but constant number of apertures were used for each expression based on pilot work, with different participants, estimating the number required to keep participants at 75% correct for each expression (17, 29 and 34, for happy, fear and neutral respectively). Note that we only include the neutral condition in order to obtain a reasonable level of task complexity: if we only include happy and fear we risk observers being able to determine the expression category solely on the basis of one feature (such as the presence or absence of the wide open mouth) and hence not process important information for the other category (i.e. the eyes in fear). Including the neutral condition lessens the chance of observers adopting such a strategy. As such, the fMRI data of the neutral condition is not analysed in the present paper.

INSERT FIGURE 1 ABOUT HERE

Participants had to judge the expression of the sparse stimulus (happy, fearful or neutral) by pressing the appropriate response key. A fixation marker (a small black and white checkerboard, 0.1X0.1 degrees) remained on the centre of the screen throughout all trials in each run. After the first 400 ms of each trial, a sparse stimulus appeared for 400 ms. There was then a response interval of 3200 ms. Participants were required to maintain fixation on the checkerboard throughout each run in order to decrease the chance of any eye movements affecting the data.

Classification Image Analysis (1) – behaviour

On each trial of the expression categorization task, the randomly located Gaussian apertures make up a two-dimensional mask (the bubble mask) that reveals a sparse face. Observers will tend to be correct when this sampled information is diagnostic for the categorization of the considered expression. To identify the image features used for each facial expression categorization, the probability of being correct was computed per pixel. We compute this by summing together all the bubble masks leading to correct categorizations, for a given expression, and dividing the result by the sum of all bubble masks shown (for correct and incorrect categorizations) for that expression. This is analogous to performing a least-square multiple regression. These probabilities were then transformed into Z-scores and thresholded to locate the statistically significant regions ($p < .05$, Cluster Test, Chauvin et al., 2005) corresponding to the features used to accurately perform the categorization of each expression. If multiplied with the original face image the thresholded Z-score maps reveal the essential information necessary for performing the categorization correctly: we refer to such information as the *diagnostic* information.

Classification Image Analysis (2) – brain

The following steps were carried out independently for each observer. Functional data for each run of the experiment were slice time corrected, corrected for 3D motion, temporally filtered (high pass filtered at 0.01hz, and linearly detrended), and finally spatially normalised into the Talaraich space with Brain Voyager QX (Brain Innovation, Masstricht, The Netherlands). Further analysis was carried out in Matlab (Mathworks, Massachusetts, US). For every trial, for each voxel and run, we select out the BOLD amplitude value at a relatively early time point where we expected the haemodynamic response function to capture valuable information, which here we took at 4 s post-stimulus onset. We chose this delay after preliminary analyses, similar to those which we report below, demonstrated that good visual information sensitivity was found here. We then z-scored the selected BOLD data independently for each voxel within each run.

INSERT FIGURE 2 ABOUT HERE

For each voxel, independently for each expression, we define the classification image as the sum of all the bubble masks leading to greater than that voxel's median BOLD amplitude, minus the sum of all the bubble masks leading to less than that voxel's median BOLD amplitude (see Figure 2). Thus we have one classification image per voxel, per expression and per observer. The classification image describes what visual information modulates BOLD activity for a given voxel, expression and observer.

Diagnostic Information Sensitivity

We measure the diagnostic information sensitivity of each voxel, for a given expression and observer, by pearson-correlating the raw classification image (i.e. unthresholded) obtained for each voxel with a diagnostic template (thresholded, liberally at $z > 1.96$) obtained from behaviour (see Figures 1 & 2). This allows us to produce an r-map of the brain where high values indicate regions which are highly sensitive (i.e. have higher BOLD signal) to the information required to perform the task. Note that we correlate both a mouth (diagnostic template for happy) and an eyes (diagnostic template for fear) template with the voxel-based classification images for each expression independently. This allows us to separate out brain regions which are sensitive to diagnostic information from those sensitive to specific visual features across the expressions. Thus we obtain four r-maps for each observer, consisting of both a mouth (diagnostic for happy; non-diagnostic for fear) and an eyes (diagnostic for fear; non-diagnostic for happy) sensitivity map for each expression (happy and fear) analysed.

Significance of r-maps

We need to devise a method which will allow us to infer which voxels, in a given r-map, display a non-chance relationship (r-value) with the relevant template. In order to assess the significance of the r values in our r-maps, we perform a randomisation test where we create a series of null distributions, one per voxel, independently for each expression, observer and template (note we have an independent set of classification images for each expression and observer). To create one such series, we randomly permute the mapping of BOLD amplitudes to bubble masks 999 times, while each time using the given mapping to create a classification image for each voxel (this preserves within each random mapping the inter-correlational structure of the real BOLD data). We correlate, on each mapping, each voxel's classification image with the relevant template to obtain the set of r-values for that mapping. The null distribution for any voxel is simply the distribution of r-values which we obtain for that voxel across the 999 random mappings. The

(one-sided) p value for a given voxel is simply the probability of observing the actual r value (or greater) in the null distribution (we compute it as the number of times a value equal to, or greater than the actual r occurs in the null distribution, as our hypothesis is one sided). Thus we obtain a p-map for each expression, template and observer. Finally, in order to correct for multiple comparisons, we set a cluster level threshold (see Rainer et al., 2006; Forman et al., 1995) for each p-map independently, based on keeping the probability of observing a false positive cluster at .05 (voxel-wise p values are first thresholded at p <= .05 in this procedure).

Information States of Brain Regions

In addition to discovering where significant information sensitivity is present across the brain, we also want to be able to describe the specific visual information each sensitive region is maximally modulated by. To describe the information state of a whole brain region, i.e. a cluster of voxels displaying significant information sensitivity (such as left Anterior Cingulate or right Fusiform Gyrus) for a particular observer, expression and template, we sum together all the voxel-based classification images (across voxels) within that region for the given observer and expression, and threshold the resulting image at z>=1.96 to reveal the visual information which modulates activity within this region.

Reverse Analysis

In order to corroborate our analysis we performed a reverse analysis: independently for each observer, we ran a GLM with four regressors. Each regressor represented, for a given expression, the correlation between each bubble mask shown, and a given feature template (i.e. mouth or

eyes; two expressions X two features = 4 regressors). This allows us to search for voxels displaying a significant relation between BOLD amplitude and the amount of visual feature (mouth or eye) information revealed. We performed this analysis first of all using a HRF sample point of 4s post-stimulus (for comparability to our forward analysis). Due to the relative ease with which such an analysis can be run in comparison with our forward analysis, we also ran the analysis sampling the HRF every second between 2 and 8 s post stimulus for a richer representation of the underlying effects.

INSERT FIGURE 3 ABOUT HERE

Results & Discussion

We show, in Figure 1, the visual information which each participant required to correctly classify the sparse faces for each expression (note we do not analyse the neutral condition in what follows). Replicating previous work, we find that the eyes are especially important for correct classification of fear whereas the mouth is important for correct classification of happy faces (e.g. Smith et al., 2005; Adolphs et al., 2005; Schyns et al., 2007). Independently for each observer and expression, we pearson-correlated their voxel-based classification images with both an eye (diagnostic for fear) and a mouth (diagnostic for happy) feature template (see Methods, Diagnostic Information Sensitivity). Significant regions (voxel wise p <=.05, cluster level p<=.05, cluster size of 300 voxels) of information sensitivity for each combination of expression, feature template, and observer are reported in Table 1.

INSERT TABLE 1 ABOUT HERE

Regions of Brain Sensitivity to Diagnostic Information

Figure 3a shows the brain regions, for observer ETS on happy trials, responding significantly to diagnostic information (the mouth) alongside the face information each region is sensitive to. We find significant sensitivity in the anterior cingulate bilaterally and in right posterior cingulate. These regions are known to be important in the processing of facial emotion (e.g. Bush et al., 2000; Britton et al., 2006 – anterior; Winston et al., 2003a - anterior and posterior). In addition, we find significant sensitivity in both the right middle temporal gyrus and in left inferior occipital gyrus, both of which are areas important in face perception (Haxby et al., 2000) and have been found active in facial expression tasks, albeit for different expressions than happy (e.g. Fitzgerald et al., 2006). Figure 3a also shows the face information each of these regions is maximally sensitive to: thus we can ascribe specific information content to the processing of each of these regions. We are the first to demonstrate that this putative network of regions important in emotion recognition is highly sensitive to the face information needed for correct behavioural performance (note that no significant sensitivity to the eyes is found anywhere in the brain on happy trials).

Turning now to the same condition for observer EGA (see Figure 3b), we find two different regions showing significant sensitivity: right insula and right parietal cortex (precuneus). The insula is another structure that has been found to be important in expression tasks (e.g. Winston et al., 2003a; Britton et al., 2006; Adolphs, 2002) while precuneus activation has also been reported (e.g. Wang et al., 2004; Habel et al., 2005, with respect to induction of sad emotions). We defer a comparison of the brain regions showing sensitivity across our two observers to a later section (see Reverse Analysis).

INSERT FIGURE 4 ABOUT HERE

Figure 4 shows the significantly sensitive (voxel-wise p<=.05; cluster threshold p<=.05) regions to both diagnostic (eyes) and non-diagnostic (mouth) information for observer ETS on fearful trials (nothing significant at these thresholds for EGA, most likely due to the smaller number of trials collected for this participant). The only region sensitive to the diagnostic information is the superior frontal gyrus whereas we find that the lingual gyrus, cuneus and parahippocampal gyrus are all sensitive to the mouth. More specifically, it seems that the latter regions are responding to a conjunction of eye and mouth information on fear trials. All these regions have been implicated in facial expression processing (e.g. Fitzgerald et al., 2006; Fu et al., 2007).

If we examine which regions are sensitive on fearful trials for observer ETS, under the liberal threshold (see Supplementary Table 1; see Supplemenatry Table 2 for EGA under a liberal threshold), we find reliable fusiform gyrus sensitivity to both the eyes and the mouth. This is consistent with previous reports of enhanced rFFA (Fusiform Face Area; see Kanwisher & Yovel., 2006) activation to fearful faces (e.g. Vuilleumier et al., 2003; Winston et al., 2003b; Fitzgerald et al., 2006). This contrasts with the pattern observed on happy trials, where there is no evidence of rFFA sensitivity whatsoever. It might seem, moreover, somewhat surprising that we find no reliable amygdala sensitivity on fearful trials, for either observer (even under the more liberal threshold) given the wealth of evidence highlighting the connection between amygdala activity and fearful faces (e.g. Whalen et al., 2004; Vuilleumier et al., 2003; Morris et al., 1996). It is important to realise, however, that our scanning protocol was not optimised for targeting this region: we sought to optimise the signal originating from the occipito-temporal areas.

Thus, in summary, we have shown that the Bubbles method can be used with fMRI to localise sensitive brain regions, in theoretically meaningful brain areas, and to depict the visual information processing strategies of each such region.

INSERT FIGURE 5 ABOUT HERE

Reverse (GLM) Analysis

In order to corroborate the results obtained with the Bubbles approach, we also performed a reverse analysis (see Method).[1] As opposed to sorting bubble masks as a function of BOLD amplitude, deriving a classification image and correlating this with feature templates (as one does in the forward analysis) we here assign each trial a continuous value which measures the extent to which the bubble mask for that trial is a good representation of either the eyes or the mouth (different regressors), independently for each expression (four regressors in total). This allows us to search for brain regions displaying a strong relation between BOLD amplitude and the presence of an important visual feature (i.e. the eyes or the mouth) by using a standard GLM.

We present, in Figure 5, a comparison of the two methods of analysis for an HRF sample point of 4s (for observer ETS): we observe a good degree of agreement for each expression. Note the clear agreement, for diagnostic happy, in bilateral anterior cingulate and right middle temporal gyrus (good agreement is also present for posterior cingulate and inferior occipital gyrus though harder to visualise on the flatmap projection), and for diagnostic fear, in the superior frontal gyrus

[1] We are most grateful to an anonymous reviewer for pointing out the feasibility of such an approach.

(left). Thus we have corroborated our main results by two different approaches. The results for the second observer (not shown) are comparable.

INSERT FIGURE 6 ABOUT HERE

In addition to corroborating our main results with two different methods, we present in Figure 6, a (whole-brain) comparison between our two observers for a HRF sample point of 4s (note we set different thresholds per observer due to the different number of trials collected). We observe an overlap in sensitivity in three main regions: bilateral anterior cingulate, right superior temporal sulcus (around Middle Temporal Gyrus), and a region around cuneus / pre-cuneus bordering posterior cingulate. Given the different number of trials obtained for each observer we believe the agreement seen is reasonable. Thus the new whole-brain projection, at offset thresholds, has shown that there are several regions of consistent sensitivity across observers.

INSERT FIGURE 7 ABOUT HERE

Furthermore, due to the relative ease of performing the GLM, we were able to run this analysis over a series of HRF sample points. We show the results of such an analysis in Figure 7a. Note the gradual emergence and disappearance of strong bilateral activity around the middle

temporal gyri, and the region encompassing the cuneus / precuneus bordering posterior cingulate, at a time frame of 3 to 7 s post-stimulus. We observe this pattern more strongly for diagnostic happy than diagnostic fear but it seems to be present in both cases. Note, in addition, the strong activity in the anterior cingulate region for diagnostic happy (this is only observed fleetingly for fear). Thus, we can be confident of the important role these regions play in the present task.

In Figure 7b, we also present a time course analysis for several ROIs defined significant at an HRF sample point of 4s. These plots clearly show the relevant areas responding selectively to either diagnostic happy (left anterior cingulate; right middle temporal gyrus) or diagnostic fear (left superior frontal gyrus). Thus it seems that the highlighted regions do show selective sensitivity for the diagnostic feature of one given expression.

INSERT FIGURE 8 ABOUT HERE

Information of Sensitive Brain Regions

We now highlight the potential which lies in directly analysing the actual visual information that modulates each sensitive brain region, for a given observer, expression and template. Figure 8 shows, on the right hand side, the classification images for all brain regions sensitive to diagnostic information (there are 18 such regions) for observer ETS on happy trials (defined from the Bubbles analysis method). We have picked out this combination of observer, expression and template for the purposes of providing a demonstration. Although there may not seem to be large differences in visual information use across these 18 regions, there are certainly subtle differences in the use of information, especially as concerns the mouth. It can clearly be seen that the

similarity of information use between related brain areas, in particular between the left and right anterior cingulate (Figure 8, fourth row) and the left and right lingual gyrus (Figure 8, first row) is high. We formalised these notions by performing a cluster analysis on the classification images shown on the right hand side in Figure 8. The images are in fact shown organised by clusters in Figure 8, where each row represents a different cluster, with the cluster centroid presented second from the left hand side. It is clear that on all possible occasions where bilateral sensitivity was evident those bilateral structures are grouped together by the cluster analysis. The probability of observing this pattern by chance for each pair of bilateral structures is low.[2] Thus our method may provide a way to identify candidate networks of brain regions solely in terms of the stimulus information they are maximally responsive to.

General Discussion

In summary, we have mapped out the information sensitivity of the brain for the categorization of two emotions, happy and fearful, and two observers, depicting the brain regions which are significantly modulated by the diagnostic (and non-diagnostic) information for each categorisation. In addition, we have shown the 'information state' of each such brain region: that is, the visual information in a face which modulates the activity of the given region. Furthermore we have corroborated our results by performing a reverse (GLM) analysis and extended that analysis to range across different HRF sample points. Finally, we have shown the potential of analysing the function of a set of brain regions in terms of the visual information they process. We now turn to comment upon some specific aspects of our methodology.

We contrasted two complementary methods of analysis in the present work: one based on pre-existing work with Bubbles (see e.g. Schyns et al., 2007) and a novel randomization test, the

[2] The probability, assuming independence of all selected brain regions, would be 0.2^8. Note that the same pattern (each pair of bilateral structures being grouped together in the same cluster) is evident for all cluster sizes ranging from 2 to 6, and that for cluster sizes of 7 to 9, the same three

other based on the GLM. The most important point is the degree of agreement between the two methods: although the two techniques are not always equal in terms of power (the GLM would seem to have more, on the whole) by using offset thresholds the overlap is very high. The fact that we can replicate our results with two different analysis methods gives us confidence in the validity of applying Bubbles to fMRI. We also note that while the Bubbles analysis and the GLM give similar results the Bubbles method provides a richer representation than the employed GLM – i.e. a two dimensional classification image representing the visual information modulating the BOLD signal for each voxel in the brain. Although we have simplified here by correlating such classification images with feature templates this is not the only approach one could take.

A different type of approach, and one which is perhaps intuitively more appealing, would be to perform a PCA or a spatial ICA on the voxel-based classification images. This would give a natural means of grouping a specific configuration of visual information with a specific set of brain regions whilst making use, simultaneously, of the whole visual information space. As the present method does not make use of the whole information space, it is possible that sensitivity to different configurations of visual information exists within the brain but that we have not picked it up: we can, however, be sure that we have captured the information sensitivity of the brain to both the eyes and the mouth for each expression considered (i.e. the critical visual information in the present task). Thus Bubbles does provide, in principle, a richer representation of the response properties of brain voxels. It would also be interesting, as an aside, to perform the Bubble analysis in a multivariate manner since different types of information have been shown to be detectable by univariate and multivariate methods of analysis (see e.g. Kriegeskorte et al., 2006).

We have, in addition, shown the potential of analysing the activity of a set of brain regions in terms of the information they process: our analysis successfully grouped together bilateral brain areas more often than would be expected by chance. Extending this analysis to

---

out of four of the bilateral structures are grouped together. Thus the ability to classify bilateral structures in terms of the visual information they are modulated by is non-chance.

incorporate time as a factor (as we have done with the GLM method) we will potentially be able to trace the flow of visual information from one brain region to another.

On a different note, now that the validity of the basic method has been established we can foresee several potentially fruitful applications: for instance, what visual information is the amygdala modulated by in a multiple facial expressions task? There have been suggestions that the eyes are critical in fear (e.g. Whalen et al., 2004) but other work suggests that the amygdala is activated similarly in response to all emotions (e.g. Winston et al., 2003a). Thus it would be interesting to discover whether the amygdala is sensitive to the diagnostic information for the judgement at hand (e.g. the eyes in fear, the mouth in happy) or to a particular configuration of visual information across expression (perhaps the eyes). We would, however, have to adapt our scanning procedure to maximise the signal coming from the amygdala area to answer this question.

Thus, in sum, we have put forward a new method with which to study the brain in the realm of high-level vision experiments. We have shown that it is possible to depict the visual information which modulates activity in different regions of the brain, separating out real sensitivity from noise. Further, our technique suggests that many important regions involved in facial expression processing are sensitive to the diagnostic information of the judgement at hand. Finally, by contrasting the information processing strategies of different brain regions, our method may provide a new way to identify candidate neural networks. We have now a new set of tools which allow us to analyse human brain function in an information processing space.

References

Adolphs, R. 2002. Recognising Emotion from Facial Expressions: Psychological and Neurological Mechanisms. Behav Cogn Neurosci Rev., 1, 21-62.

Adolphs, R., Gosselin, F., Buchanan, T.W., Tranel, D., Schyns, P., Damasio, A.R. 2005. A mechanism for impaired fear recognition after amygdala damage. Nature, 433, 22-23.

Britton, J.C., Taylor, S.F., Sudheimer, K.D., Liberzon, I. 2006. Facial expressions and complex IAPS pictures: Common and differential networks. NeuroImage, 31, 906-919.

Bush, G., Luu, P., Posner, M.I. 2000. Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn Sci. 2000. 4, 215-222.

Chauvin, A., Worsley, K.J., Schyns, P.G., Arguin, M., Gosselin, F. 2006. Accurate statistical tests for smooth classification images. J Vis. 2005, 5, 659-667.

Dailey, M., Cottrell, G. W., Reilly, J., 2001. *California Facial Expressions, CAFE,* unpublished digital images, UCSD Computer Science and Engineering Department.

Ekman, P., Friesen W.V., 1978. The facial action coding system (FACS): A technique for the measurement of facial action. Palo Alto, CA: Consulting Psychologists Press.

Fitzgerald, D.A., Angstadt, M., Jelsone, L.M., Nathan, P.J., Phan, K.L. 2005. Beyond threat: Amygdala reactivity across multiple expressions of facial affect. NeuroImage, 30, 1441-1448.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C. 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn Reson Med. 33, 636-647.

Fu, C.H., Williams, S.C., Brammer, M.J., Suckling, J., Kim, J., Cleare, A.J., Walsh, N.D., Mitterschiffthaler, M.T., Andrew, C.M., Pich, E.M., Bullmore, E.T. 2007. Neural responses to happy facial expressions in major depression following antidepressant treatment. Am J Psychiatry. 164, 540-542.

Goebel, R., Esposito, F., Formisano, E. 2006. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Hum Brain Mapp. 2006. 27, 392-401.

Gosselin, F., Schyns, P.G. 2001. Bubbles: A technique to reveal the use of information in recognition tasks. Vision Research, 41, 2261–2271.

Habel, U., Klein, M., Kellermann, T., Shah, N.J., Schneider, F. 2005. Same or different? Neural correlates of happy and sad mood in healthy males. NeuroImage. 26, 206-214.

Haynes, J., Rees, G. 2005. Decoding mental states from brain activity in humans. Nat. Rev. Neuroscience, 7, 523-534.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishal, A., Schouten, J.L., Pietrini, P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science, 293, 2425-2430.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I. 2000. The distributed human neural system for face perception. Trends Cogn Sci. 2000. 4, 223-233.

Kamitani, Y., Tong, F. 2005. Decoding the visual and subjective contents of the human brain. Nature Neuroscience, 8, 679-685.

Kanwisher, N., Yovel, G. 2006. The fusiform face area: a cortical region specialized for the perception of faces. Philos Trans R Soc Lond B Biol Sci. 361, 2109-2128.

Kriegeskorte, N., Goebel, R., Bandettini, P. 2006. Information-based functional brain mapping. PNAS, 103, 3863-3868.

Morris, J.S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R.J.. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, *383*, 812-815.

Nielsen, K.J., Logothetis, N.K., Rainer, G. 2006. Discrimination strategies of humans and rhesus monkeys for complex visual displays. Curr Biol., 16, 814-820.

Nielsen KJ, Rainer G. 2007. Object recognition: similar visual strategies of birds and mammals. Curr Biol., 17, R174-R176.

Ringach, D., Shapley, R. 2004. Reverse Correlation in Neurophysiology. Cognitive Science, 28, 147-166.

Schyns, P.G., Jentzsch, I., Johnson, M., Schweinberger, S.R., Gosselin, F. 2003. A principled method for determining the functionality of brain responses. NeuroReport, 14, 1665-1669.

Sekuler, A.B., Gaspar, C.M., Gold, J.M., Bennett, P.J. 2004. Inversion leads to quantitative, not qualitative, changes in face processing. Curr Biol. 14, 391-396.

Smith, M.L., Gosselin, F., Schyns, P.G. 2004. Receptive fields for flexible face categorizations. Psych Sci, 15, 753-761.

Smith M.L., Cottrell, G.W., Gosselin, F., Schyns, P.G. (2005). Transmitting and decoding facial expressions. Psychol Sci, 16, 184-189.

Smith, M.L., Gosselin, F., Schyns, P.G. 2006. Perceptual moments of conscious visual experience inferred from oscillatory brain activity. PNAS, 103, 5626-5631.

Smith, M.L., Gosselin, F., Schyns, P.G. 2007a. From a face to its category via a few information processing states in the brain. NeuroImage, epub ahead of print.

Smith, M.L., Fries, P., Goebel, R., Gosselin, F., Schyns, P.G. 2007b. The information domain – A new approach to interpreting MEG brain signals. NeuroImage, 36, S21.

Wang, A.T., Dapretto, M., Hariri, A.R., Sigman, M., Bookheimer, S.Y. (2004). Neural correlates of facial affect processing in children and adolescents with autism spectrum disorder. J Am Acad Child Adolesc Psychiatry. 43, 481-490.

Whalen, P.J., Kagan, J., Cook, R.G., Davis, F.C., Kim, H., Polis, S., McLaren, D.G., Somerville, L.H., McLean, A.A., Maxwell, J.S., Johnstone, T. 2004. Human amygdala responsivity to masked fearful eye whites. Science, 306, 2061.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience, 6,* 624-631.

Winston, J.S., O'Doherty, J., Dolan, R.J. 2003a. Common and distinct neural responses during direct and incidental processing of multiple facial emotions. NeuroImage, 20, 84-97.

Winston, J.S., Vuilleumier, P., Dolan, R.J. 2003b. Effects of low-spatial frequency components of fearful faces on fusiform cortex activity. Curr Biol. 13, 1824-1829.

Figure Captions

Figure 1: Examples of original (first row) and sparse stimuli used (second row), and behavioural classification images for each observer (bottom two rows) and expression (columns).

Figure 2: Flow diagram indicating the main stages of fMRI data analysis. For a given observer, expression and voxel, we sort the bubble masks into two sums, those associated with greater and those associated with less than median BOLD amplitude. We form the voxel-based classification image as the difference of these two image sums. We then correlate all the voxel-based classification images, for a given expression and observer, with both a mouth (diagnostic for happy) and an eyes (diagnostic for fear) feature template, resulting in a mouth and an eyes r-map for that specific expression and observer.

Figure 3: Regions displaying significant diagnostic information sensitivity on happy trials. (A) Regions displaying significant sensitivity to the mouth (diagnostic; first row - excluding right anterior cingulate) and the visual information each region is responding to (second row) for ETS. (B) Regions displaying significant sensitivity to the mouth (diagnostic; first row) and the visual information each region is responding to (second row) for EGA.

Figure 4: Regions displaying significant sensitivity on fear trials for observer ETS. The first column shows the brain region significantly sensitive to the eyes (diagnostic; upper row) along with the visual information this region responds to (lower row) while the remaining columns show those brain regions significantly sensitive to the mouth (non-diagnostic; upper row) and the visual information each region is responding to.

Figure 5: Comparison of forward (Bubbles) and reverse (GLM) analysis methods. (A) Comparison of the two methods for diagnostic (mouth) information on happy trials (Bubbles - blue; GLM - green). (B) Comparison of the two methods for diagnostic (eyes) information on fear trials (Bubbles – red; GLM – yellow). Note that we use a different threshold for each method (Bubbles t>1.5; GLM t>1.8) reflecting the relative power of each method to detect significant regions.

Figure 6: Comparison of ETS and EGA for an HRF sample point of 4s post-stimulus (GLM). (A) Areas observed sensitive for EGA to the diagnostic feature for each expression (happy – green; fear - yellow). (B) Areas observed sensitive for ETS to the diagnostic feature for each expression (happy – green; fear - yellow). Note we use a different threshold for the two observers (EGA t>1.5; ETS t>1.8) reflecting the fact that there are different numbers of trials per observer. STS = Superior Temporal Sulcus; AC = Anterior Cingulate; C/PrC/PC = Region including cuneus, pre-cuneus bordering posterior cingulate.

Figure 7: Diagnostic information sensitivity as a function of HRF sample point (observer ETS). (A) Flatmap projection of the sensitive diagnostic areas for each expression (happy – green; fear – yellow – both t>1.8) across a range of HRF sample points for observer ETS. (B) Time course of beta weights for selected ROIs for each expression (happy: left anterior cingulate, right middle temporal gyrus; fear: left superior frontal gyrus). The ROIs are defined as significant (t>1.99, cluster threshold of 300 voxels) at a HRF sample point of 4s post-stimulus. (C) The diagnostic information for each expression modulating the brain areas as shown in (A) and (B). MTG = Middle Temporal Gyrus; C/PrC/PC = Region including cuneus, pre-cuneus bordering posterior cingulate.

Figure 8: Hierarchical cluster analysis (single linkage algorithm using correlation distance metric) of the regions displaying significant diagnostic information sensitivity for ETS on happy trials ($p < .1$ voxel wise, arbitrary cluster threshold of 300 voxels, Bubbles analysis). Each row represents a different cluster of brain regions, with the second image in each row representing the cluster centroid of the given cluster (the first image is the thresholded version of the given centroid). The right hand side of each cluster shows the classification images for each different region assigned to that cluster.

Table 1

Regions showing significant information sensitivity for each combination of observer, expression and feature template.

Observer	*Expression*	*Template*	*Region*	*Laterality*	*TAL*	*Cluster Size*
ETS	Happy	D (mouth)	Anterior Cingulate	Left	-4 42 5	645
				Right	7 36 8	384
			Posterior Cingulate	Left	3 –59 17	303
				Middle Temporal Gyrus	Right	41 –64 28 412
			Inferior Occipital Gyrus	Left	-10 –92 –7	490

ETS	Happy	AD (eyes)	Nil					
ETS	Fear	D (eyes)	Superior Frontal Gyrus	Left	-23 41 39	340		
ETS	Fear	AD (mouth)	Lingual Gyrus	Right	17 -89 –4	506		
			Cuneus	Left	-4 -72 21	313		
			Parahippocampal gyrus	Left	-16 -30 –3	308		
…								
EGA	Happy	D (mouth)	Insula	Right	44 -2 12	328		
			Precuneus	Right	13 –57 19	400		

EGA	Happy	AD (eyes)	Nil
EGA	Fear	D (eyes)	Nil
EGA	Fear	AD (mouth)	Nil

P-maps were thresholded at p <=.05 voxel wise and further cluster size thresholded to ensure the probability of observing a false cluster was <=.05. D = diagnostic template, AD = anti-diagnostic template.

Table 2

Regions showing significant information sensitivity, for observer ETS, for each combination of expression and template at a liberal threshold.

Observer	*Expression*	*Template*	*Region*	*Laterality*	*TAL*	*Cluster Size*
ETS	Happy	D (mouth)	Anterior Cingulate	Left	-1 40 6	2033(1)
				Right	5 37 25	1033(1)
			Posterior Cingulate	Left	3 –59 17	759(1)
			Cingulate Gyrus	Left	-1 –47 26	466(1)
			Parahippocampal Gyrus (Hippocampus)	Left	-25 –18 –15	1132(4)
			Parahippocampal Gyrus (Amygdala)	Right	32 –2 –13	605(4)
			Inferior Occipital Gyrus	Left	-11 –92 –7	898(1)
			Lingual Gyrus	Left	-13 –65 –2	331(1)
				Right	14 –68 –4	558(1)

			Inferior Parietal Lobule	Right	58 –24 24	336(2)	
			Middle Temporal Gyrus	Right	41 –62 25	1103(3)	
				Left	-47 –66 21	307(2)	
			Superior Temporal Gyrus	Left	-46 –56 26	468(5)	
			Supramarginal Gyrus	Right	56 –52 20	322(4)	
			Temporal Lobe (subgyral)	Right	46 –3 –9	916(13)	
			Inferior Frontal Gyrus	Right	41 25 –1	766(1)	
			Insula	Right	36 –13 24	469(31)	
ETS	Happy	AD (eyes)	Nil				
ETS	Fear	D (eyes)	Superior Frontal Gyrus	Left	-26 41 39	925(1)	
				Right	13 37 45	376(38)	

			Inferior Frontal Gyrus	Left	-41 22 9	573(8)
			Middle Frontal Gyrus	Right	41 25 44	317(9)
			Middle Temporal Gyrus	Left	-44 –55 13	310(1)
			Superior Temporal Gyrus	Right	44 –55 16	372(12)
			Posterior Cingulate	Left	-3 –49 14	368(1)
			Fusiform Gyrus	Right	44 –47 –20	343
			Left Brainstem (Red Nucleus)	Left	0 –22 –4	832(8)
ETS	Fear	AD (mouth)	Medial Frontal Gyrus	Left	-12 40 11	353(1)
			Middle Frontal Gyrus	Left	-45 40 18	458(1)

Lingual Gyrus	Right	9 –90 –3	1702(6)
Cuneus	Left	-5 –72 21	676(1)
Middle Occipital Gyrus	Right	39 –68 7	401(20)
Inferior Occipital Gyrus	Left	-26 –89 –11	478(6)
Fusiform Gyrus	Right	40 –47 –20	959
Occipital Face Area	Right	32 –62 –18	626
Cingulate Gyrus	Right	7 –27 35	441(21)
Thalamus	Left	-16 –30 –2	601(5)
Insula	Left	-48 –19 22	325(12)
Superior Temporal Gyrus	Left	-55 –52 16	303(24)

P maps were thresholded at p <=.1 voxel wise and a cluster threshold of >= 300 voxels was imposed to limit the chance of finding clusters by chance. We do this purely to observe trends evident in regions active across expressions and observers.

The number in parentheses is an error measure of the location of the region (1 is best). D = diagnostic, AD = anti-diagnostic.

Table 3

Regions showing significant information sensitivity, for observer EGA, for each combination of expression and template at a liberal threshold.

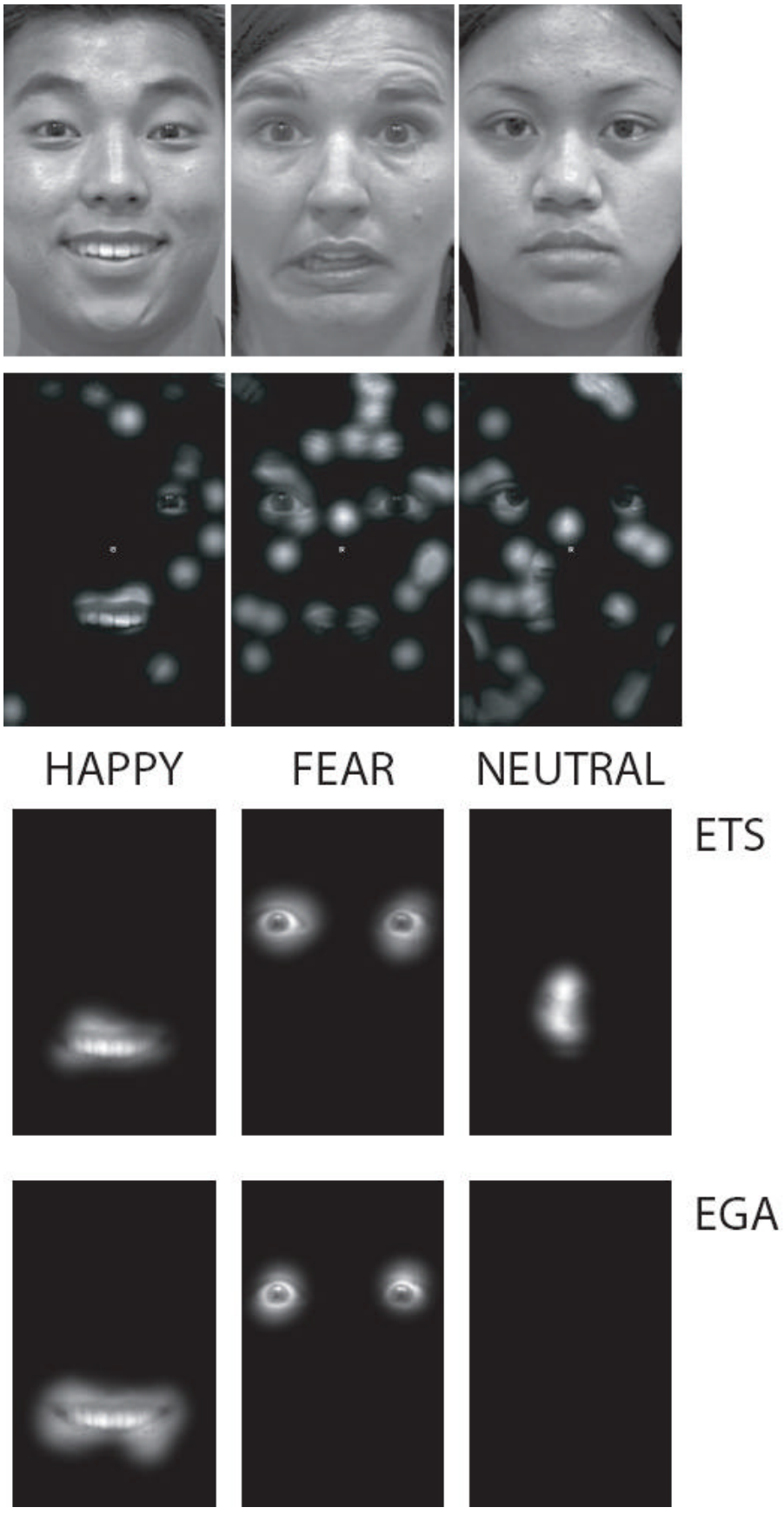
Observer	*Expression*	*Template*	*Region*	*Laterality*	*TAL*	*Cluster Size*
EGA	Happy	D (mouth)	Anterior Cingulate	Left	0 25 2	654(4)
			Posterior Cingulate	Right	7 –43 7	304(6)
			Cingulate Gyrus	Left	-15 –47 21	324(18)
			Parahippocampal Gyrus	Right	21 –37 –5	770(7)

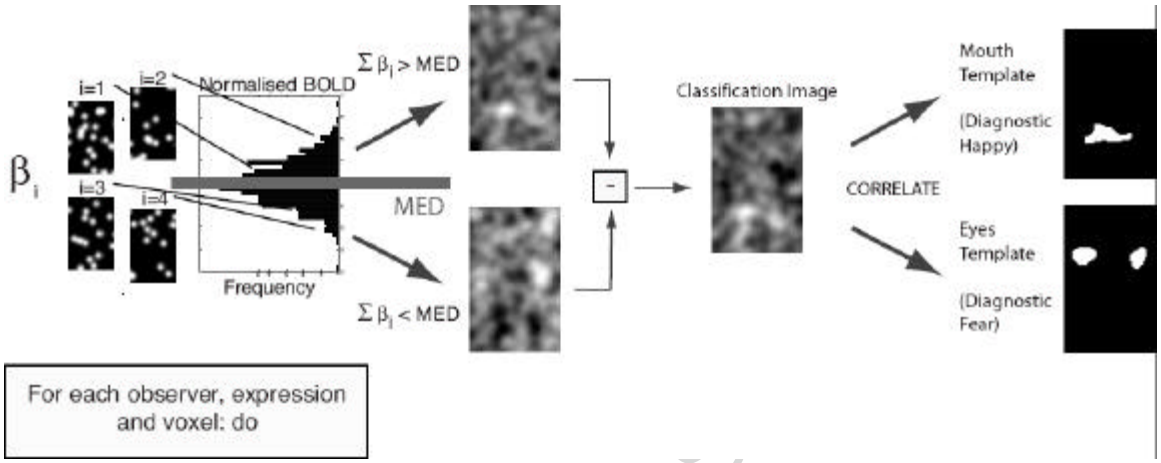
			Lingual Gyrus	Left	-13 –69 –2	509(8)
			Inferior Parietal Lobule	Right	58 –28 26	388(10)
			Precuneus	Right	13 –57 19	765(10)
			Postcentral Gyrus	Left	-49 1-5 20	690(24)
			Insula	Left	-40 –3 13	424(14)
				Right	46 -2 11	726(14)
			Superior Temporal Gyrus	Right	47 –32 5	1013(9)
			Cerebellum	Right	2 -63 –26	344(1)
				Left	-11 –36 –5	744(11)
EGA	Happy	AD (eyes)	Anterior Cingulate	Right	5 30 10	336(1)

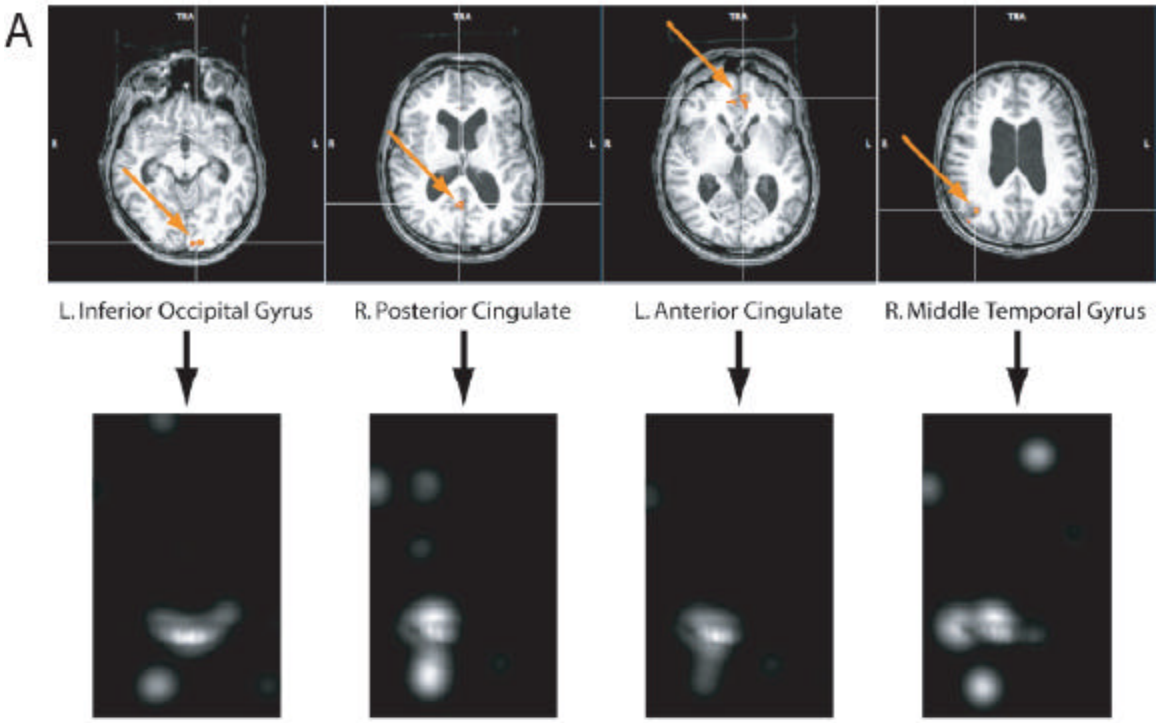
EGA	Fear	D (eyes)	Nil			
EGA	Fear	AD (mouth)	Posterior Cingulate	Right	8 –58 10	335(3)
			Inferior Frontal Gyrus	Left	-39 28 10	997(9)

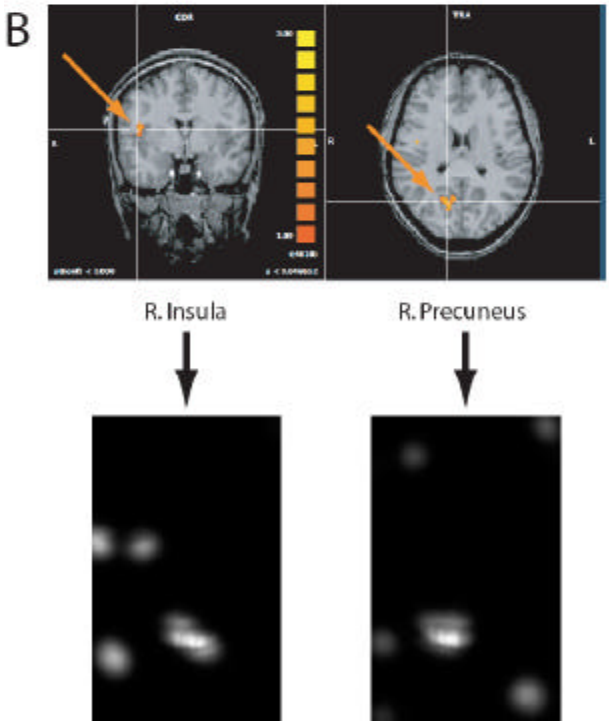
P maps were thresholded at p <=.1 voxel wise and a cluster threshold of >= 300 voxels was imposed to limit the chance of finding clusters by chance. We do this purely to observe trends evident in regions active across expressions and observers.

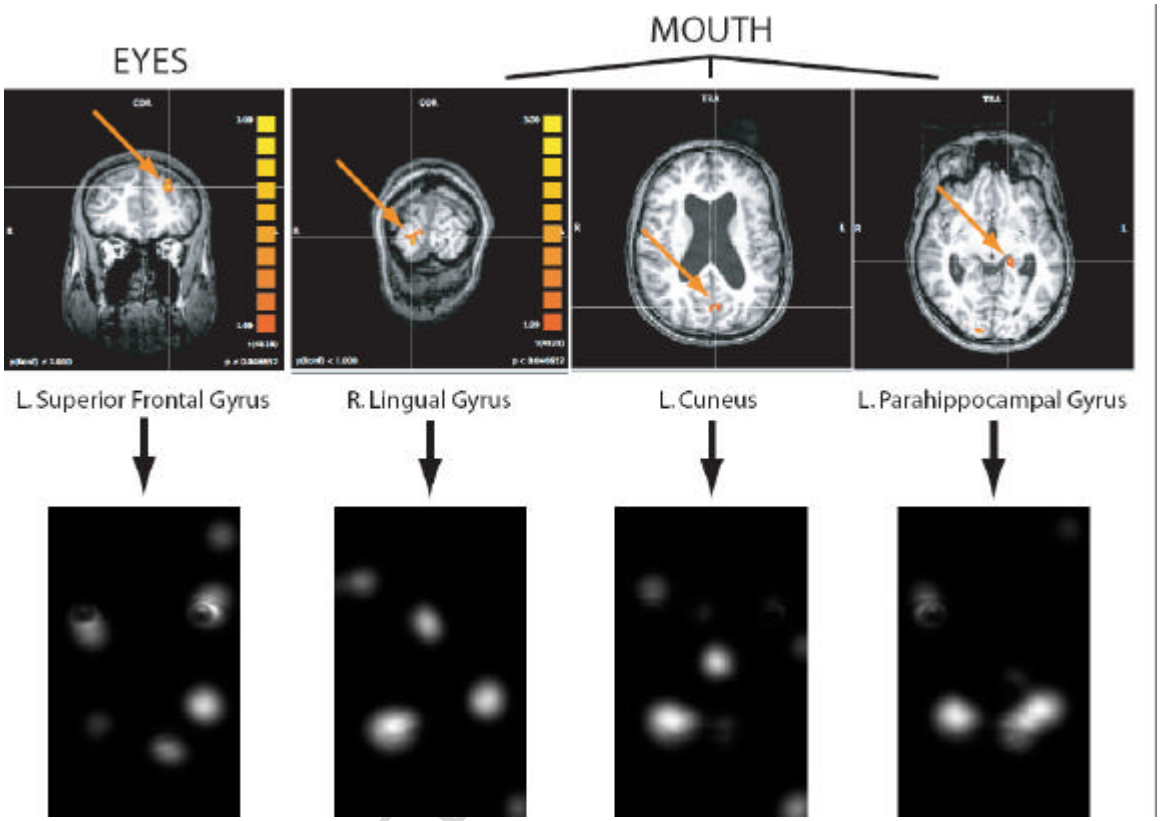
The number in parentheses is an error measure of the location of the region (1 is best). D = diagnostic, AD = anti-diagnostic.

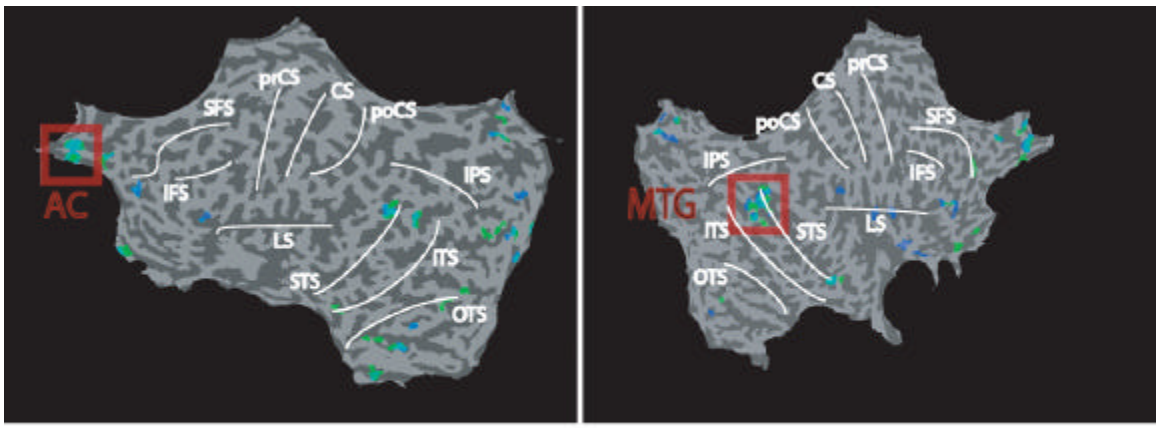




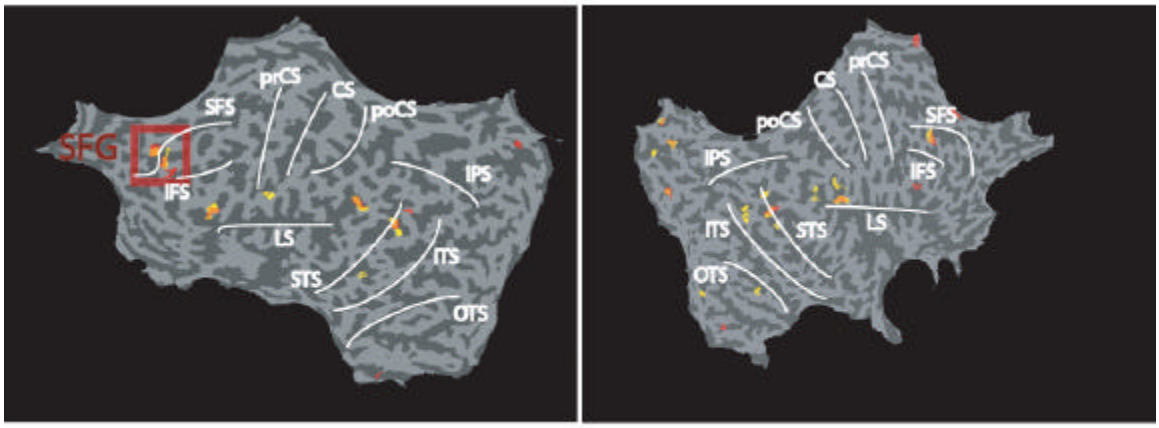








Happy - Bubbles (Randomization) Method
Happy - GLM Method



Fear - Bubbles (Randomization Test) Method
Fear - GLM Method