Brain Reading Using Full Brain Support Vector Machines for Object Recognition: There Is No "Face" Identification Area

Stephen José Hanson
jose@tractatus.rutgers.edu
Yaroslav O. Halchenko
yoh@psychology.rutgers.edu
Rutgers Mind/Brain Analysis Laboratories, Psychology Department,
Rutgers University, Newark, NJ 07102, U.S.A.

Over the past decade, object recognition work has confounded voxel response detection with potential voxel class identification. Consequently, the claim that there are areas of the brain that are necessary and sufficient for object identification cannot be resolved with existing associative methods (e.g., the general linear model) that are dominant in brain imaging methods. In order to explore this controversy we trained full brain (40,000 voxels) single TR (repetition time) classifiers on data from 10 subjects in two different recognition tasks on the most controversial classes of stimuli (house and face) and show 97.4% median out-of-sample (unseen TRs) generalization. This performance allowed us to reliably and uniquely assay the classifier's voxel diagnosticity in all individual subjects' brains. In this two-class case, there may be specific areas diagnostic for house stimuli (e.g., LO) or for face stimuli (e.g., STS); however, in contrast to the detection results common in this literature, neither the fusiform face area nor parahippocampal place area is shown to be uniquely diagnostic for faces or places, respectively.

1 Introduction

Recent work in visual object recognition using neuroimaging measures has been controversial concerning the way that information may be stored or encoded in the brain. In particular, some (Kanwisher, 2006; Kanwisher, McDermott, & Chun, 1997; Spiridon & Kanwisher, 2002) have proposed that the underlying representations of visual object categories are localized in specific brain tissue associated with specific types or tokens (e.g., "faces," "places," "body parts or motion"). Other work has suggested that coding for object identity in the brain is more distributed, similar to a relief or topographic map in which some pattern of activity may be associated with

This research supported by a grant from the McDonnell Foundation.

specific object types or tokens (Haxby et al., 2001).[1] And in more recent work along these lines, Hanson, Matsuka, & Haxby (2004) indicate that the category codes may actually be combinatoric and thus allow the coding to efficiently reuse voxel information (e.g., neural populations) for each specific type or token.

Statistical classifiers, which have become much more powerful in the past decade, we believe, have the potential to resolve this controversy. At this point, only a handful of researchers have used just a few kinds of classifiers to attempt to "read out" the contents of brain activity as indexed by fMRI (Carlson, Schrater, & He, 2003; Cox & Savoy, 2003; Hanson et al., 2004; Haxby et al., 2001; Haynes & Rees, 2006; LaConte, Strother, Cherkassky, Anderson, & Hu, 2005; Mitchell et al., 2004; O'Toole, Jiang, Abdi, & Haxby, 2005). Because the BOLD signal is very noisy, typically less than 5% of the actual signal is useful for classification, and in a pairwise or small multiclass (e.g., 10 classes) discrimination task, the classifier can possess final error from 10% (based on averaging single scans in blocks of 10 or 20 scans) to as high as 40% (on single scans) with unseen sample scans (Hanson et al., 2004; O'Toole et al., 2005). Standard linear methods based on the general linear model (GLM) for identification of potentially selective brain regions are neither crossvalidated nor optimized for performance (Kanwisher et al., 1997; Kanwisher, Stanley, & Harris, 1999; Spiridon & Kanwisher, 2002). In a typical block design study, the GLM is used to fit selective time series, typically accounting for less than 50% of the total variance, despite statistically significant excursions from a predefined baseline (note that the significance of a statistical test also implies a high variance accounted for in the fit of the linear model). This allows for considerable indeterminacy regarding the identification of object-selective voxels or regions. A key conceptual error in this type of method is using the strength of regression coefficients as measures of identification when what they are merely indicating is the presence of brain tissue conditional on the presence of the stimulus exemplar $p(\mathrm{VOXEL} \mid \mathrm{FACE} > \textit{chance})$, or what many in this field have called "selective." Kanwisher (2006) seems aware of this limitation, as she has most recently described the significance of what she first dubbed as the "fusiform face area": "More recently, fMRI has revealed a particular region in the human brain where this special face perception machinery apparently resides: the fusiform face area, a blueberry-sized region on the bottom surface of the posterior right hemisphere that *responds significantly more strongly* when people look at faces than when they look at any other stimulus class yet tested [emphasis added]."

[1] Still other work in this field (Gauthier, Skudlarski, Gore, & Anderson, 2000) makes a claim in an orthogonal direction to the analysis here and has proposed that the fusiform face area seems to code for more general properties of stimulus types (e.g., "expertise"). Nonetheless, the same argument concerning the confounding identification with detection discussed below applies and very well may affect the interpretation of their basic claim. This is not at present the focus of this letter.

The difference in detectability between faces in a prelocalized fusiform face area (FFA) is at its maximum around 2% above baseline, while for all other objects except for whole humans, headless bodies, or animal heads, it is closer to 1%, a factor of 2 in detectability that is quite impressive. Kanwisher has maintained that although other areas of the brain might be face selective—that is, they detect the presence of a face stimulus—they are also not uniquely face selective as she maintains the FFA is. Unfortunately for this claim, face detection is not face identification. Recall that identification refers to the ability in a set of alternatives to identify a stimulus as belonging to a specific category, which may be as small as a specific token (e.g., recognizing your cat). The degree to which errors occur in identification results in a classic confusion matrix, which further delineates the discriminability and similarity of various object tokens (Krantz, Luce, Suppes, & Tversky, 1971). Detectability is related to identification in that it is necessary for identification but unfortunately is not sufficient.

Luce (Krantz et al., 1971) has formalized this relationship in what is often referred to the Luce choice axiom:

$$P(i|v) = \frac{w_{(i|v)}}{\sum w_{(n|v)}} \quad \textit{for } i \in n.$$

Note that choice identification depends on conditional probabilities of the object given the voxel or feature, however relative to all other potential alternative category choices. This type of calculation can be made with strength or detectability measures such as regression coefficients, and in fact, when it was done correctly by Haxby et al. (2001), they found that the voxels involved in face identification were actually distributed over much of the inferior temporal lobe. Kanwisher, not deterred by this demonstration, pointed out that despite the obvious distributed nature of the code for most objects she also tested,

> although categorical information is apparently spread over a broad expanse of the ventral visual pathway, our analysis finds little or no evidence that *the FFA and PPA carry discriminative information about nonpreferred stimuli.* This finding is inconsistent with the suggestion of Haxby et al. (2001) that "regions such as the parahippocampal place area or the fusiform face area are not dedicated to representing only spatial arrangements of human faces but, rather, are part of a more extended representation for all objects" [Spiridon & Kanwisher, 2002, p. 2427; emphasis added].

Indeed, we agree that this is the key test: Do FFA and parahippocampal place area (PPA) carry discriminative information about "nonpreferred stimuli" (house and face, respectively)? In this letter, using single scans with a whole brain statistical classifier we show that they do.

2 Selectivity Measures Are Confounded

In the previous discussion, we pointed out that the work in visual object recognition using fMRI measures uses the percentage above baseline activity in voxels to indicate selectivity. Although this may seem like a benign choice, we will argue this is at the heart of the confusion in this literature and has tended to perseverate a view that selectivity of cortical tissue can be measured unambiguously with normalized BOLD responses. Our problem with these measures is actually twofold: first, in the context of the GLM (this has nothing necessarily to do with localizers per se), a detection is calculated based on a contrast between one stimulus condition against one or possibly others. This type of measure consequently deals only with the fact that the voxel is implicated as a regressor. It does not indicate that it is diagnostic or that identification of the category is conditionally dependent on that feature. Worse, the intensity, which covaries with the strength of the coefficient, is confounded with its location. Consider this: What if blood flow to cortical regions actually signaled the amount of "work" or "energetics" involved in processing states? In this case, those BOLD intensities that were not at peak levels might be processing those particular stimulus patterns more efficiently than those with a peak levels that are actually "working harder." This problem requires selectivity measures that are designed toward diagnostic identification rather than strength of an association that could reflect voxels that were lower in intensity than others but were more reliably useful in predicting the correct category for a given stimulus. Fundamentally, cognitive neuroscientists who study visual object recognition are interested in whether the putative brain mechanisms are used to correctly classify or identify the stimulus, not simply respond more in some location.

3 Identification Tasks

Note that we are discussing voxel identification in contrast to behavioral identification of a subject who is asked to identify a stimulus that may be noise masked or otherwise obscured (Grill-Spector & Kanwisher, 2005). With decreasing mask values, a standard psychophysical function will result, which indicates that the subject has correctly identified the specific stimulus, and of course one can measure associated fMRI. Our point here again is that independent of the task, even in a human subject identification task, a GLM analysis of voxel values can produce voxels only that are associated or most similar to systematic variations of the independent variable. This should be seen as distinct from a voxel identification function, which if one chose to pursue and as we have discussed above puts us squarely in the realm of statistical classifiers.

4 Statistical Classifiers

There has been enormous progress over the past few years in statistical classifiers: recent work often focuses on problems in extremely large feature spaces (more than 100K to 1M) and poor signal-to-noise measurement, often with remarkable out-of-sample generalization (Guyon, Weston, Barnhill, & Vapnik, 2002). In our, case, the BOLD signal has notoriously low signal-to-noise gain and may reflect a mixture of various neural signals and, in particular, field potentials. There is also considerable evidence that the underlying distribution is nongaussian (Chen, Tyler, & Baseler, 2003; Hanson & Bly, 2001), indicating that parametric classifiers may underestimate valid signal excursions, making approaches such as Haxby et al. (2001) employed biased toward conservative generalization estimates. In fact, when Hanson et al. (2004) compared Haxby's nearest-neighbor classifier to neural network classifiers, using a conservative cross-validation estimate showed improvements by as much as 30% to 40% in generalization. Much of the improvement in statistical classification often can be attributed to both a more appropriate classification function (e.g., specific nonlinearity) and feature weighting or selection (Hanson & Burr, 1990). Given that we are interested in discovering features (areas of the brain) that are discriminative for specific stimulus types, we will also want to focus on whole brain classification. To date, in this visual object recognition domain (see Mitchell et al., 2004, for use of SVM in a more generic "brain reading" context and LaConte et al., 2005, who also used single scans in order to filter support vectors), no one has attempted whole brain, single scan classification instead tending to focus on regions of interest, not surprisingly in the inferior temporal (IT) lobe. Nonetheless, it has been pointed out a number of times in the object recognition field that there are object-selective or face-selective areas of the brain outside IT (Chao, Martin, & Haxby, 1999; Gauthier, Tarr, Moylan, Skudlarski, Gore, & Anderson, 2000; Malach, Levy, & Hasson, 2002). However, it still remains unclear whether these selective areas are also areas that uniquely identify these object types: Do these other brain areas that are neither FFA or PPA carry discriminative information about exemplars like faces or houses? Clearly, given what we have already established, the only definitive way to answer this type of question is in the domain of statistical classifiers.

In tests explored here, we employed a number of classifiers; however, in this letter, we focus on support vector methods (SVMs) (Guyon, Boser, & Vapnik, 1993). SVMs have many desirable properties: they can learn even in huge (more than 1 million) feature spaces, they will produce a unique solution due to problem formulation as a constrained quadratic problem, they are known to generalize well while working with feature spaces many orders larger than the data sample size, they can be highly robust over significant levels of noise, and they can learn subtle distinctions near the separating hyperplane that are most ambiguous about each category sample.

This is all done without having to invest in the costly learning of the complete distribution of each category (on the other hand, this will turn out to be a disadvantage for visualization).

In order to resolve the question of brain area identification we use the full brain (approximately 40,000 voxels, we use white matter as noise background to increase reliability of estimates) and then incrementally searched for voxels that were uniquely identifying either face or house stimuli. This was done exhaustively per subject per scan and cross-validated on independent sets of data. In order to increase the generalizability of our approach, we also used data from two separate object identification experiments where 10 subjects performed judgments on the two key object types (FACE, HOUSE) in two different tasks. The first task was from a benchmark (Haxby et al., 2001), from five subjects while they viewed pictures of faces and houses as well as four other categories of man-made objects (chairs, scissors, shoes, and bottles). In this case, subjects performed a one-back repetition detection task (we designate this as the N-back, or NB, task). In this letter, we focus on the key stimuli for the identification task of FACE and HOUSE. In the second task, which avoids spatial memory demand and focuses a simple perceptual judgment, we used similar high-contrast black and white stimuli in an oddball task (we will designate task 2 as the oddball, or OB, task), where trials consisted of simultaneous presentation of three stimuli in different orientations, in which subjects had to identify the one that was different (either three faces in FACE trials or three houses in HOUSE trials) from the others. In both tasks, subjects achieved behavioral accuracy rates identifying objects above 80% correct. In both cases with full brain (approximately 40,000–50,000 voxels) data, there were 144 (77 of each type) samples in task 1 and 200 (100 of each type) in task 2. Both experiments used a block design with blocks of size 7 in task 1 and size 17 (initially there were 20 trials as we eliminated repetition time (TR(s)) from the block end points to reduce autocorrelation effects) in task 2 to train and generalize the classifier (leave-two blocks-out; although all classification was done with single scans or equivalently single TRs; see Hanson et al., 2004). In order to achieve the lowest possible error in generalization, backward or recursive feature elimination (RFE) was performed (the strategy is after Guyon et al., 2002; also see Ishak & Ghattas, 2005; Rakotomamonjy, 2003). This approach has the advantage of detecting the specific object identification brain areas by harvesting the most sensitive voxels after training and doing subsequent retraining on this more sensitive reduced set. This process was continued until there are no more voxels left to test and therefore was exhaustive over the single-scan brain voxel set.

To avoid potential cross-block contamination of hemodynamic BOLD signal and therefore accidentally creating a generalization bias, scans at the beginning and the end of all blocks (category or rest conditions) were discarded. We also routinely hold out whole blocks and test single TRs from those blocks (holding out all other TRs in that block until they are sampled later). This reduces any upward bias from possible correlation within blocks. A 20 second rest block was also used in both experiments to prevent any temporal contamination or confusion between category responses. In order for the SVM classifier to operate on the data, data had to be converted from raw voxel values to lie primarily within the [−1,+1] range. Two possible conversion schemes were tested: scaled percentage change relative to baseline and *z*-scores relative to the baseline (rest condition blocks) statistics. SVMs trained on *z*-scores provided better a generalization on chosen test cases and thus were chosen for further analysis. Each 3D scan (brain voxels only) containing roughly up to 40,000 voxels was used as an input sample for the classification, having a label of the corresponding stimulus condition (face or house) when that scan was acquired. Specifically all subjects' data were submitted to a soft-margin SVM with on average of 500 to 1000 voxels (approximately a 12% to 15% exponential removal rate until we reached 100 voxels and then removed one voxel at a time until only one voxel was left) per step backward feature elimination. On each step, each voxel's diagnosticity (see next section) based on the standard SM-SVM was used to keep or eliminate that voxel. The voxels with the smallest weights on each step were eliminated.

In order to obtain an unbiased estimate of SVM generalization performance, each data set was split into training and testing data sets. Similar to Hanson et al. (2004), an N-1 block bootstrap procedure was implemented; specifically, for each training set, a single block from each category (blocks of both FACE and HOUSE) was taken out for testing, which left $B - 1$ per category used for training. All possible combinations of testing blocks from the two categories were taken, which made up (12×12 blocks = 144 and 10×10 blocks = 100 bootstrap cases for NB and OB, respectively) training and testing data sets.[2] Each training and testing data set proceeded through recursive feature elimination independently, and at the end, their performances were averaged over all classifiers and bootstraps to obtain generalization estimate for a given subject/SVM. Shown in Figure 1 is the generalization average error as a function of voxels that remain left in the training set. Each line shows a single subject performance on out-of-sample pairs of HOUSE and FACE exemplars. The strength of the line indicates the group of subjects in either NB (bold lines) or OB (light lines) tasks. Note that near the minimum, they significantly overlap, with the OB task starting at a lower error on initial learning.[3]

[2] For two of the subjects—one in the OB task and the other in NB—a single block in each case was found to be corrupted; in those cases, the number of bootstrap opportunities was 81 and 121, respectively.

[3] One possible reason for error advantage in the oddball task could simply be the difference in scanner strength, which for N-Back was 1.5T and for OB was 3T, although there are other task-related explanations, including that the oddball task required a category judgment, while the N-Back task is focused on single stimulus identification, hence involving

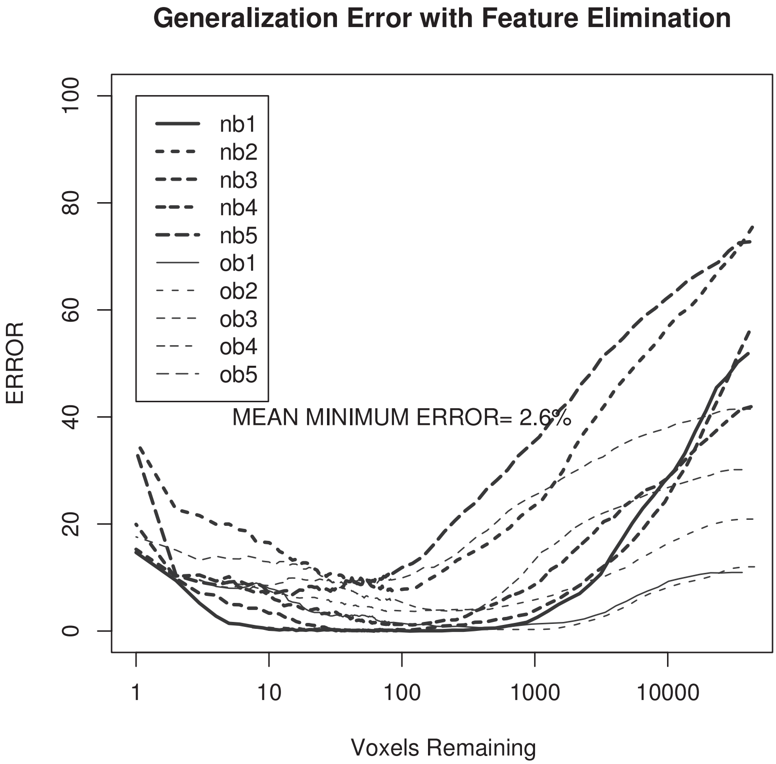

Figure 1: Generalization error for all 10 subjects on single scans for both kinds of stimuli held out from a training exemplars. Recursive voxel elimination produced a minimum for half the subjects at zero error, while over all subjects on single scans, across both tasks, there is nearly a 98% correct generalization on unseen scans.

Table 1: Minimal Error and Associated Voxels Count.

Generalization/ Subject	NB1	NB2	NB3	NB4	NB5	OB1	OB2	OB3	OB4	OB5
Voxels	200	104	117	120	34	378	125	81	313	225
Error	0.0	6.8	0.0	0.0	6.9	0.0	2.7	6.7	3.0	0.0

Table 1 shows the exact minimum error and associated voxels remaining at that point. The minimum average error for all subjects in both tasks

no behavioral contrast between stimuli. In either case, a parametric experiment that varied the difficulty of the categorization judgment might resolve this intercept difference.

indicated nearly 98% correct out-of-sample cases with three subjects in the NB task and 2 from the OB task showing zero generalization error.

5 Brain Area Identification

Which areas of the brain are uniquely required for the identification of faces versus houses? We harvest the classifiers by doing a sensitivity analysis (similar to Hanson et al., 2004) by effectively asking what their contribution to classification error is when they are removed versus when present. If category error significantly changes when the voxel is removed, then we can infer that it is contributing in a proportionally diagnostic way to the identification of the FACE or HOUSE category. We discuss next a detailed description of the visualization methods.

6 Visualization

6.1 Atlas Registration. All brain volumes (tasks, subjects) were first registered in the following steps: (1) the subject's sample bold scan, anatomical, and Montreal Neurological Institute (MNI) anatomical were skull stripped using BET from FSL tools (Smith et al., 2004); then (2) coregistered stripped anatomical to stripped MNI using FLIRT (from FSL) to obtain anatomical-to-MNI transformation; and then (3) coregistered stripped anatomical to stripped BOLD using FLIRT to obtain anatomical-to-BOLD transformation; and finally (4) transformed full anatomical into BOLD space using anatomical to BOLD transformation for easy visualization of activation patterns.

6.2 Localizers for the FFA and PPA. In order to localize the FFA and PPA for analysis, we adopted the standard method in the field to find voxels that significantly responded to FACE > "other nonFACE objects" or HOUSE> "other non-house objects." In the case of the N-Back task, we used the original FACE and HOUSE masks in Haxby et al.'s (2001) study, which were created in the way we are about to describe and consisted of masks that ranged from 20 voxels to about 100 voxels. In the case of the oddball study, we used independent localizer scans in a standard GLM contrasts for FACE > HOUSE and HOUSE > FACE. Although for someone first hearing this, these procedures may seem tautologous, it is nonetheless the standard procedure within the literature to establish selective regions of interest (ROI)s for subsequent testing. There is a controversy that has erupted recently about this method (Friston, Rotshtein, Geng, Sterzer, & Henson, 2006; Saxe, Brett, & Kanwisher, 2006), but nonetheless, we maintain that this procedure is a reliable way to construct candidate voxels for more diagnostic approaches as advocated in this letter even though its validity might be in question. In Hanson et al. (2004), for example, we started with an ROI mask of 1500 voxels (much larger than the FFA in any particular subject), and using sensitivity analysis, we were able to reduce the diagnostic voxel set to 40% to 50% of the original masks. In what we describe below, we will intersect the FFA and PPA masks with our sensitivity masks to determine percentage voxel overlap. The difficulty in identifying the FFA in particular is that there can be considerable variability in the size of masks across subjects, and in the original paper (Kanwisher et al., 1997) that first reported FFA, only 12 of the 16 subjects actually had FFA activity (functional sampling error?). In any case, finding the FFA by GLM contrasts of localizer scans is state of the art in the neuroimaging field for candidate voxel selectivity.

6.3 SVM Sensitivity. There are a number of possibilities for identification of diagnostic brain areas. An obvious choice might be all the support vectors themselves. However, it should be clear that using support vectors strictly in the margin may be inappropriate in characterizing the diagnostic brain areas, as they can represent brain volumes that are near the separating surface and therefore are "near-misses" in some broader sense of the category that would not be representative of the brain response to house stimuli or face stimuli. The other possibility is the nonsupport vectors (NSV), which was done by LaConte et al. (2005). These are vectors that are distributed beyond the support vectors; indeed, some may be prototypical of the brain response, but unfortunately many will not be. In fact, in general, even assuming a gaussian spread of the NSVs, only a small minority will be typical or "best" members of the category. Because SVM optimizes a margin between the two categories, the actual distribution of members of the categories is in effect ignored. Hence, ironically the same property that makes SVM an excellent candidate for classification in high dimensions is the one that also makes it tricky to visually interpret. We therefore chose a visualization approach in between these two extreme possibilities; effectively we implemented a sensitivity/perturbation approach (e.g., Hanson et al., 2004), which measures the error for a given category (face, house) when the voxel is present versus when it is removed. Specifically, to estimate SVM-based sensitivity, we used one of the simplest criteria proposed (Guyon et al., 2002; Ishak & Ghattas, 2005; Rakotomamonjy, 2003), which is simply the reciprocal of the separating margin width $W = \|w\|$, where $w = \sum_i \alpha_i y_i x_i$. Minimization of this criterion leads to maximization of the margin width. In the case of linear SVM, the squared values of the separating plane normal coefficients (i.e., w_i^2), as stated, effectively correspond to the change of the criteria W as if the voxel i is removed. Therefore, the classifier is less sensitive to the features with low w_i^2. During recursive feature elimination, we sequentially eliminated features with the smallest w_i^2. Additionally, in order to increase diagnostic selectivity, we derived weights for FACE category by using only FACE SVs and for HOUSE category by using only HOUSE category SVs. Thus, higher voxel value tends toward typical regions in the classification space for the SV appropriate category. We will refer to these direction selective voxel coefficients as *diagnosticity* of the voxel direction for or against HOUSE (in blue) and FACE (in red). We show in Figure 3 (discussed in more detail in section 6.4) diagnosticity measures for two typical subjects (one from NB and one from OB). We used a nonparametric method for thresholding these diagnosticity distributions at $p < .01$ for each subject, taking into account the nongaussianity of the underlying diagnosticity distributions (compare with the type 4 method constructing empirical cumulative distributions; Hyndman & Fan, 1996). Slices are shown for each subject with specific voxel clusters. Note that SVM finds fairly contiguous regions, despite no specific bias to do so; in fact, at lower thresholds (.05), the diagnosticities begin to fractionate through the whole brain. Unlike a regression analysis (e.g., GLM), which finds "selective" or detection areas, these voxel patterns are unique identification areas in that they are contributing to a correct classification of one category against the other and have been cross-validated in independent data sets, implying that they will generalize to other unseen cases of either faces or houses.

On this basis, we now can answer our question by assaying the above threshold identification areas in order to determine whether (1) there are unique identification areas for face and house and (2) whether the FFA or PPA is carrying other discriminative information about other nonpreferred stimuli (in this case, either house or face, respectively).

6.4 Brain Areas Identifying FACEs and HOUSEs. In Figure 2, we show all areas harvested that were common (intersection set)[4] to all subjects in both tasks that were either highly diagnostic of faces or of houses based on the sensitivity and diagnosticity analysis as described. The bar plot shows the percentage of voxels associated with each area at the .01 threshold using the nonparametric methods. The total in each bar can be computed by multiplying the percentage against the total number of sensitivities above threshold in each category (FACE = 363 and HOUSE = 358), so, for example, about half of the voxels in both HOUSE and FACE, or about 150, fall into the FFA. Note that these areas are based on the best cross-validated single-scan classifiers that were nearly 98% correct on out-of-sample exemplars. Hence, the areas we identify under these constraints are not based on the usual object-selective interpretation and therefore are not subject to the resultant ambiguity with methods that are based on similarity or association. They are, in fact, albeit in a probabilistic sense, necessary and sufficient for the

[4] We used a very conservative harvesting to include only voxels that were in all 10 subjects and voxels that were above $p < .01$ and therefore appeared in both tasks. We also initially averaged over all generalization runs in order to increase the sample power of the voxel sets per subject. In any case, there was significant overlap (64%) of the same voxels across generalization runs, and this tended to covary with the minimum generalization error reached for that classifier. It is appropriate to average across the runs, since any differences in voxel sets are due to sampling error in classifier estimation and data noise.

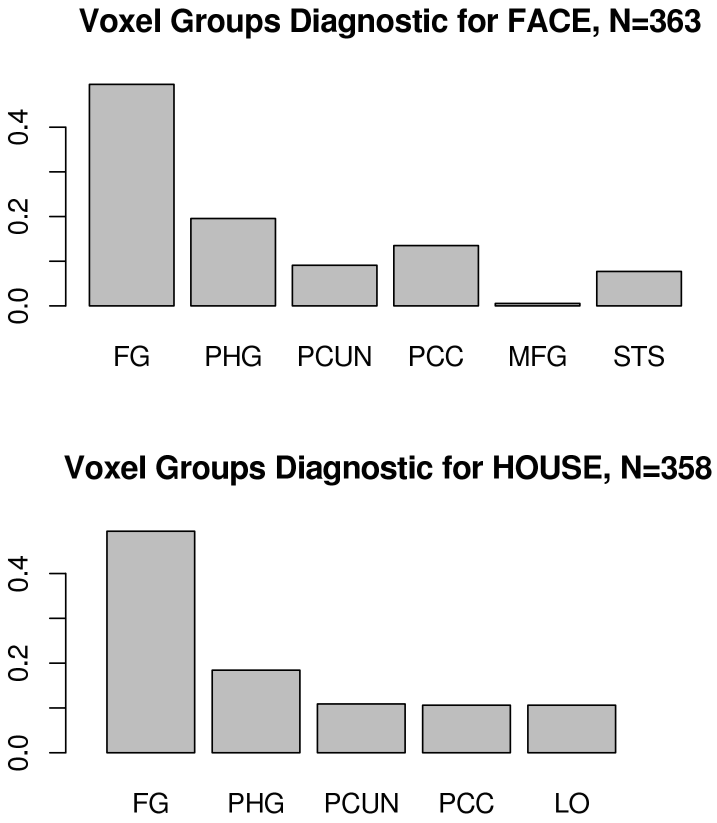

Figure 2: Voxel area cluster relative frequencies intersected over all subjects and both tasks. For FACE diagnosticities (top), six areas were identified at $p < .01$ (FG = fusiform gyrus; PHG = parahippocampal gyrus; PCC = posterior cingulate cortex; MFG = middle frontal gyrus; STS = superior temporal sulcus; PCUN = precuneus) with relative frequencies based total number of voxels ($N = 363$) over all areas. For HOUSE diagnosticities (bottom), there were five areas (at $p < .01$, $N = 358$)—some the same and some different: FG = fusiform gyrus; PHG = parahippocampal gyrus; PCC = posterior cingulate cortex; PCUN = precuneus; and LO = lateral occipital. Note that overlapping FACE and HOUSE areas are FG, PHG PCUN, and PCC. Distinctive areas for FACE included STS, MFG, while for HOUSE, distinctive areas included only LO.

identification of faces or houses. It is important at this point to clarify that the classifiers could have found single areas or single voxels as predictive of a single category. In fact, linear methods tend to be biased toward using single dimensions (voxels in this case) to minimize classifier error, especially if area correlations tend to be small between features or voxels. If there are large correlations between voxels, these could be minimized by using a neural network, which can decorrelate features as it classifies. Nearest-neighbor classifiers such as used by Haxby et al. (2001) in fact are more biased toward sets of features since their similarity increases incrementally with more common overlap of voxels. Consequently, we can begin to answer the question posed in the Introduction: Are there other areas of the brain that are diagnostic for the identification of faces or houses? Do other areas of the brain carry discriminative information other than FFA and PPA? Clearly from our analysis, the answer is yes.

Overall, the diagnostic profile of relevant areas for both FACEs and HOUSEs are somewhat similar but not identical. Nonetheless there are distinct areas between each category that are part of a larger network of areas. In fact, in Figure 3 we show a selection of representative examples from different subjects across both tasks (OB and NB). In the figure, each paired set of figures in a row is the same subject showing the diagnosticities for FACE in red (on the left of each pair) and HOUSE in blue (on the right of each pair). In the first two paired sets, we show FFA sensitivity in two different subjects across the two different tasks in both FACE and HOUSE stimulus presentations. For all subjects, fusiform gyrus (and FFA masks overlapped with 90% of the FG voxels for all subjects and both tasks) appeared for both FACE and HOUSE, indicating diagnostic value for this area that was neither specialized nor unique. Also for all subjects across the tasks, the parahippocampal gyrus (and PPA masks overlapped with more than 70%) and the posterior cingulate cortex (PCC) was also diagnostic of FACE and HOUSE. The next three sets show distinctive diagnostic areas, including unique STS sensitivity for FACE but not HOUSE. Other common diagnostic areas are middle occipital gyrus (including area LO) and middle frontal gyrus (BA 9). In general, a network of areas was identified as diagnostic of these two stimulus types, with FACE having a prefrontal area involved, while HOUSE appeared to involve an area known for visual shape and texture processing. Note that we are not implying that FACE stimuli do not require specialized functions (e.g., shape, texture) processing; rather, the areas that have been identified in specialized experiments induce us to use labels that have some lexical familiarity with the stimuli we have used, which otherwise may be misleading in another context where those brain areas are interacting with many other brain areas. We take up this labeling problem and other implications from this study in the following discussion.

7 Discussion

This letter started out to answer a simple question that has been plaguing the object recognition field for the last 10 years: Is there a unique area of the brain whose sole purpose is to identify faces? Further, is there a unique area of the brain whose sole purpose is to identify houses? Based on the

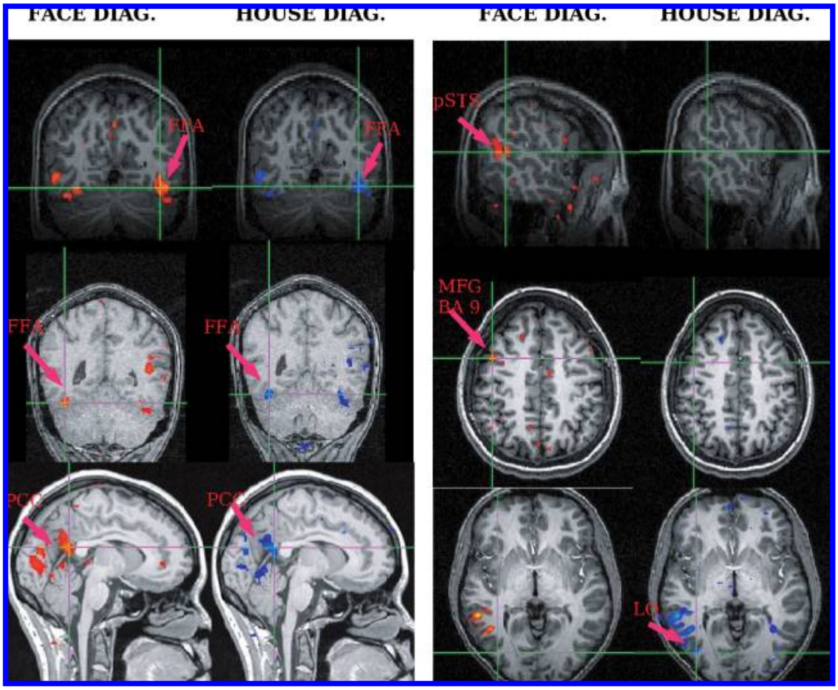

Figure 3: Voxel sensitivities for brain areas that are diagnostic for FACE (red and on left of each panel) and for HOUSE (blue and right of each panel). In the top left corner, we show FFA (fusiform face area as measured by a localizer task) in the first paired panel (same subject, same slice, same task), which shows subject NB1. Below that is subject OB3, also showing FFA in both slices. On the left bottom panel, we show PCC, which is diagnostic for HOUSE and FACE, which both show sensitivity (NB2). The next three paired images show distinctive areas between diagnostic areas of FACE and HOUSE. The next panel above on the right shows pSTS (posterior superior temporal sulcus), which is diagnostic for FACE against HOUSE (subject OB1). The next panel below shows another FACE diagnostic area (MFG BA9, subject OB5). And in the last paired panel on the right, we show a HOUSE diagnostic area (LO) for subject NB5.

analysis, the answer is no. To be fair, we should qualify this "no" in two ways. First, we are not claiming some general property of combinatorial coding throughout the brain; rather, our results are specific to inferior temporal lobe and in particular for object recognition processes, and second, we are using standard resolution fMRI (3 mm) which might change the results dramatically as we descend to more and more detail within inferior temporal lobe and FFA in particular. Nonetheless, what we did find is a disjoint network of brain areas that are diagnostic for either face stimuli or house stimuli. This is consistent with Haxby, Hoffman, and Gobbini's (2000) model of face perception in that the areas we found are in fact involved in a common network for processing faces. The common areas in this analysis include the fusiform face area, specific areas of the lingual gyrus (which showed up weakly in our analysis), the middle occipital gyrus (LO), and precuneus; distinctive areas for face stimuli included pSTS and middle frontal gyrus (BA9). Distinctive areas for house stimuli included only area LO. However, for faces, there seems to be no evidence in this analysis that the FFA is unique diagnostically, nor does it appear to be specialized since it was identified in every subject responding to the house stimuli.

So this observation would seem to put a rest to the controversy that there are unique areas of the brain that respond only to specific tokens or types. But not exactly. Kanwisher's claim is more complicated than just this observation of the fact that many areas of the brain seem to be required for identification of these object types, faces and houses, for example. In fact, the claim is whether presumptive areas that are already selected through an independent localizer test are uniquely and solely involved in identifying these object types. So to paraphrase Kanwisher, do the FFA and PPA provide discriminative information about object types other than face and house, respectively? Clearly again, the answer based on the present classifiers is yes. Since most of the voxels identified for either face or house were squarely sitting in the FFA of each subject, this would seem to be definitive evidence that the FFA does involve discriminative information about object types other than faces—in this case, houses. We also must note that the exact overlap of the areas between FFA-face and FFA-house is not exactly the same, despite voxels that are squarely in the same place (see cross-hairs); nonetheless, this is normative in the field, as the FFA will have different locations and shapes across subjects and even within a subject across sessions or experiments will vary in strength, location, and shape. Given that houses and faces do look different, it is not impossible that the FFA does code these stimuli differently, and a high-resolution experiment might very well produce different distributions of recruited voxels in the FFA for each stimulus, which might be more consistent with Kanwisher's claims. Nonetheless, within the state of the art, our localizations of diagnostic cortical areas have no more variability or lack of precision than that which appears in the standard literature. A potentially more difficult issue for Kanwisher is that in our results, there were also brain areas that were distinctive for face or house other than FFA or PPA. For example, it would be possible to argue that pSTS is the superior temporal sulcus FACE area, the pSTSFA! We know that this area is sensitive to biological motion, Why could there not be a part of it specifically dedicated to the unique identification of faces? Other category tests would be very likely to show these areas are not distinctive, but more likely part of a larger network of some kind of face identification system. In terms of uniqueness and given the specificity of the claim, that the FFA is claimed to be unique and necessary for FACE identification, then if it were also to be diagnostic for any other category such as houses, the claim must be refuted. Again, this would seem to lay the matter to rest: there is no face area per se, at least in the way that Kanwisher has defined it. Clearly, we are forced to conclude there is no area that responds uniquely and solely to faces that could be found in a whole brain assay of a high-performance single TR classifier.

So that would seem to be the end of the story, and this letter, but as usual in this problematical and persistent claim, there always seem to be more twists and turns. In fact, more recently, Kanwisher and her supporters have retreated to a more nuanced claim that was always inherent in the original claim: that of higher-resolution detection. They argue that the reason that one can find the FFA responding to other object types is that the fMRI methodology is simply not high resolution enough to identify the tiny "blueberry" size cluster of cells that are in fact, the fusiform face area. Unfortunately for this claim, more recent work involving very high-resolution fMRI analysis seems to show the opposite of what one might hope if the blueberry FFA existed (Grill-Spector, Sayres, & Ress, 2006; see also Haxby, 2006, for an interesting discussion of this result). In fact, if the code for object recognition was distributed in the way that Haxby (e.g., Hanson et al., 2004; Haxby et al., 2001) for one has contended, one would expect to see a fractionating of signal at higher and higher resolutions as opposed to a single "blueberry beacon" of face recognition as Kanwisher's position must hold. Even a cursory examination of Grill-Spector's scans shows a low-resolution smooth, convex cluster breaking into smaller islands of distributed structure, making the retreat to high resolution unlikely to clearly support Kanwisher's proposal or perhaps any simple object-related proposal (although she reports small face-selective voxels, these are "selective" in the standard sense, and it remains to be seen how "selective" they may be once reassessed properly with a classifier). Finally, though, let us suppose for the sake of argument that Kanwisher is correct. Suppose that there is a very small area (or many small areas) of the brain, maybe even the size of a single neuron, that detects faces and only faces. But then in the final analysis, why would there be? What exactly was the plausible computational mechanism that could make any sense of such an unlikely representation for something as complex and important as recognizing someone's face?

Acknowledgments

We thank Jim Haxby and Catherine Hanson for discussions and two anonymous reviewers for edits that make this version of the letter considerably better than when it had started. We are also grateful for generous support from the McDonnell Foundation and NSF.

References

Carlson, T. A., Schrater, P., & He, S. (2003). Patterns of activity in the categorical representations of objects. *J. Cogn. Neurosci.*, *15*(5), 704–717.

Chao, L. L., Martin, A., & Haxby, J. V. (1999). Are face-responsive regions selective only for faces? *Neuroreport*, *10*(14), 2945–2950.

Chen, C. C., Tyler, C. W., & Baseler, H. A. (2003). Statistical properties of BOLD magnetic resonance activity in the human brain. *Neuroimage*, *20*(2), 1096–1099.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, *19*(2 Pt. 1), 261–270.

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *Neuroimage*, *30*(4), 1077–1087.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.*, *3*(2), 191–197.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *J. Cogn. Neurosci.*, *12*(3), 495–504.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychol. Sci.*, *16*(2), 152–160.

Grill-Spector, K., Sayres, R., & Ress, D. (2006). High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nat. Neurosci.*, *9*(9), 1177–1185.

Guyon, I., Boser, B., & Vapnik, V. (1993). Automatic capacity tuning of very large VC-dimension classifiers. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in neural information processing systems*, *5* (pp. 147–155). San Mateo, CA: Morgan Kaufmann.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.*, *46*(1–3), 389–422.

Hanson, S. J., & Bly, B. M. (2001). The distribution of BOLD susceptibility effects in the brain is non-gaussian. *Neuroreport*, *12*(9), 1971–1977.

Hanson, S. J., & Burr, D. J. (1990). What connectionist models learn: Toward a theory of representation in connectionist networks. *Behavioral and Brain Sciences*, *13*, 471–518.

Hanson, S. J., Matsuka, T., & Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: Is there a "face" area? *Neuroimage*, *23*(1), 156–166.

Haxby, J. V. (2006). Fine structure in representations of faces and objects. *Nat. Neurosci.*, *9*(9), 1084–1086.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.*, *4*(6), 223–233.

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.*, *7*(7), 523–534.

Hyndman, R. J., & Fan, Y. (1996). Sample quantiles in statistical packages. *American Statistician*, *50*(4), 361–366.

Ishak, B., & Ghattas, B. (2005). An efficient method for variable selection using SVM based criteria. *Journal of Machine Learning Research*, *6*, 1357–1370.

Kanwisher, N. (2006). Neuroscience. What's in a face? *Science*, *311*(5761), 617–618.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.*, *17*(11), 4302–4311.

Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals. *Neuroreport*, *10*(1), 183–187.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). San Diego: Academic Press.

LaConte, S., Strother, S., Cherkassky, V., Anderson, J., & Hu, X. (2005). Support vector machines for temporal classification of block design fMRI data. *Neuroimage*, *26*(2), 317–329.

Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends. Cogn. Sci.*, *6*(4), 176–184.

Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M., & Newman, S. (2004). Learning to decode cognitive states from brain images. *Machine Learning*, *57*, 145–175.

O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *J. Cogn. Neurosci.*, *17*(4), 580–590.

Rakotomamonjy, A. (2003). Variable selection using SVM-based criteria. *Journal of Machine Learning Research*, *3*, 1357–1370.

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *Neuroimage*, *30*(4), 1088–1099.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, *23* (Suppl. 1), S208–S219.

Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron*, *35*(6), 1157–1165.

Received September 8, 2006; accepted December 16, 2006.

This article has been cited by:

1. Russell A. Poldrack. 2011. Inferring Mental States from Neuroimaging Data: From Reverse Inference to Large-Scale Decoding. *Neuron* **72**:5, 692-697. [CrossRef]
2. Christopher L. Suhler, Patricia Churchland. 2011. Can Innate, Modular “Foundations” Explain Morality? Challenges for Haidt's Moral Foundations Theory. *Journal of Cognitive Neuroscience* **23**:9, 2103-2116. [Abstract] [Full Text] [PDF] [PDF Plus]
3. Peter M. Rasmussen, Lars K. Hansen, Kristoffer H. Madsen, Nathan W. Churchill, Stephen C. Strother. 2011. Model sparsity and brain pattern interpretation of classification models in neuroimaging. *Pattern Recognition* . [CrossRef]
4. J. D. Carlin, J. B. Rowe, N. Kriegeskorte, R. Thompson, A. J. Calder. 2011. Direction-Sensitive Codes for Observed Head Turns in Human Superior Temporal Sulcus. *Cerebral Cortex* . [CrossRef]
5. A. Manelis, L. M. Reder, S. J. Hanson. 2011. Dynamic Changes In The Medial Temporal Lobe During Incidental Learning Of Object-Location Associations. *Cerebral Cortex* . [CrossRef]
6. A. Nestor, D. C. Plaut, M. Behrmann. 2011. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences* **108**:24, 9998-10003. [CrossRef]
7. Francisco Pereira, Matthew Botvinick. 2011. Information mapping with pattern classifiers: A comparative study. *NeuroImage* **56**:2, 476-496. [CrossRef]
8. Georg Langs, Bjoern H. Menze, Danial Lashkari, Polina Golland. 2011. Detecting stable distributed patterns of brain activation using Gini contrast. *NeuroImage* **56**:2, 497-507. [CrossRef]
9. Peter Mondrup Rasmussen, Kristoffer Hougaard Madsen, Torben Ellegaard Lund, Lars Kai Hansen. 2011. Visualization of nonlinear kernel models in neuroimaging by sensitivity maps. *NeuroImage* **55**:3, 1120-1131. [CrossRef]
10. Andre F Marquand, Sara De Simoni, Owen G O'Daly, Steven CR Williams, Janaina Mourão-Miranda, Mitul A Mehta. 2011. Pattern Classification of Working Memory Networks Reveals Differential Effects of Methylphenidate, Atomoxetine, and Placebo in Healthy Volunteers. *Neuropsychopharmacology* . [CrossRef]
11. Svetlana V. Shinkareva, Vicente L. Malave, Robert A. Mason, Tom M. Mitchell, Marcel Adam Just. 2011. Commonality of neural representations of words and pictures. *NeuroImage* **54**:3, 2418-2425. [CrossRef]
12. Stephen José Hanson, Arielle Schmidt. 2011. High-resolution imaging of the fusiform face area (FFA) using multivariate non-linear classifiers shows diagnosticity for non-face categories. *NeuroImage* **54**:2, 1715-1734. [CrossRef]

13. Anna Manelis, Catherine Hanson, Stephen José Hanson. 2011. Implicit memory for object locations depends on reactivation of encoding-related brain regions. *Human Brain Mapping* **32**:1, 32-50. [CrossRef]
14. Svetlana V. Shinkareva, Vicente L. Malave, Marcel Adam Just, Tom M. Mitchell. 2011. Exploring commonalities across participants in the neural representation of objects. *Human Brain Mapping* n/a-n/a. [CrossRef]
15. Tanya Schmah, Grigori Yourganov, Richard S. Zemel, Geoffrey E. Hinton, Steven L. Small, Stephen C. Strother. 2010. Comparing Classification Methods for Longitudinal fMRI Studies. *Neural Computation* **22**:11, 2729-2762. [Abstract] [Full Text] [PDF] [PDF Plus] [Supplementary Content]
16. Kevin S. Weiner, Kalanit Grill-Spector. 2010. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *NeuroImage* **52**:4, 1559-1573. [CrossRef]
17. Patrick Suppes, Marcos Perreau-Guimaraes, Dik Kin Wong. 2009. Partial Orders of Similarity Differences Invariant Between EEG-Recorded Brain and Perceptual Representations of Language. *Neural Computation* **21**:11, 3228-3269. [Abstract] [Full Text] [PDF] [PDF Plus] [Supplementary Content]
18. Russell A. Poldrack, Yaroslav O. Halchenko, Stephen Jos## Hanson. 2009. Decoding the Large-Scale Structure of Brain Function by Classifying Mental States Across Individuals. *Psychological Science* **20**:11, 1364-1372. [CrossRef]
19. Joset A. Etzel, Valeria Gazzola, Christian Keysers. 2009. An introduction to anatomical ROI-based fMRI classification analysis. *Brain Research* **1282**, 114-125. [CrossRef]
20. Michael Hanke, Yaroslav O. Halchenko, Per B. Sederberg, Stephen José Hanson, James V. Haxby, Stefan Pollmann. 2009. PyMVPA: a Python Toolbox for Multivariate Pattern Analysis of fMRI Data. *Neuroinformatics* **7**:1, 37-53. [CrossRef]
21. Francisco Pereira, Tom Mitchell, Matthew Botvinick. 2009. Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage* **45**:1, S199-S209. [CrossRef]
22. David Pitcher, Lucie Charles, Joseph T. Devlin, Vincent Walsh, Bradley Duchaine. 2009. Triple Dissociation of Faces, Bodies, and Objects in Extrastriate Cortex. *Current Biology* **19**:4, 319-324. [CrossRef]
23. F DEMARTINO, G VALENTE, N STAEREN, J ASHBURNER, R GOEBEL, E FORMISANO. 2008. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage* **43**:1, 44-58. [CrossRef]
24. R POLDRACK. 2008. The role of fMRI in Cognitive Neuroscience: where do we stand?. *Current Opinion in Neurobiology* **18**:2, 223-227. [CrossRef]
25. Michael L. Mack, Jennifer J. Richler, Thomas J. Palmeri, Isabel GauthierCategorization . [CrossRef]