To err is human: correlating fMRI decoding and behavioral errors to probe the neural representation of natural scene categories

Dirk B. Walther
Beckman Institute
University of Illinois at Urbana-Champaign

Diane M. Beck
Beckman Institute and Department of Psychology
University of Illinois at Urbana-Champaign

Li Fei-Fei
Department of Computer Science
Stanford University

Index words:
natural scenes
scene categories
categorization
fMRI decoding
error correlations
searchlight analysis
confusion matrix
correlation with behavior
behavioral correlation

Ah ne'er so dire a Thirst of Glory boast,
Nor in the Critick let the Man be lost!
Good-Nature and Good-Sense must ever join;
To err is Humane; to Forgive, Divine.

(An Essay on Criticism, Alexander Pope, 1711)

Abstract

New multivariate methods for the analysis of functional magnetic resonance imaging (fMRI) data have enabled us to decode neural representations of visual information with unprecedented fidelity. But how do we know if humans make use of the information that we decode from the fMRI data for their behavioral response? In this chapter we propose a method for correlating the errors from fMRI decoding with the errors made by subjects in a behavioral task. High correlations suggest that subjects use information that is closely related to the content of the fMRI signal to make their behavioral response. We demonstrate the viability of this method using the example of natural scene categorization. Humans are extremely efficient at categorizing natural scenes (such as forests, highways, or beaches), despite the fact that different classes of natural scenes often share similar image statistics. Where and how does this happen in the brain? By applying multi-voxel pattern analysis to fMRI data recorded while subjects viewed natural scenes we found that the primary visual cortex (V1), the parahippocampal place area (PPA), retrosplenial cortex (RSC), and the lateral occipital complex (LOC) all contain information that distinguishes among natural scene categories. Correlation of the decoding errors with errors made by the subjects in a behavioral experiment revealed that only the information about scene categories contained in the PPA, RSC, and LOC is directly related to behavior, but not the information in area V1. A match between behavioral performance and accuracy of decoding scene categories from the PPA and RSC for two manipulations of the stimuli (scene inversion and quality of category exemplar) underscores the central role of these two areas in natural scene categorization.

1. Introduction

Vision science has made tremendous progress in understanding how the brain processes various components of our visual world. Part of this progress is owed to the advent of functional magnetic resonance imaging (fMRI), a non-invasive neuroimaging method that can image activity in the whole brain. Indeed, fMRI has enabled the mapping of several important visual areas in the human brain, for instance, retinotopic visual cortex including primary visual cortex and extrastriate regions (Engel et al., 1994), the lateral occipital cortex for object perception (Malach et al., 1995), the fusiform face area (Kanwisher et al., 1997), and the parahippocampal place area (Epstein and Kanwisher, 1998). In these seminal studies univariate statistics was employed (i.e. each voxel was analyzed independently) to produce maps of functional activity. However, our understanding of stimulus representations in the brain has greatly expanded with the application of multi-voxel pattern recognition techniques to fMRI data. It has now been shown, for example, that the neural representation of a much larger range of stimuli can be decoded if one considers the *pattern* of activity across voxels in response to visually presented objects, leveraging the distributed nature object representations in the human brain (Haxby et al., 2001). This discovery has spurred a surge of studies applying multi-voxel pattern analysis (MVPA) techniques to many questions in visual neuroscience and beyond.

In this chapter we aim to demonstrate the importance of tying MVPA to behavioral data using the example of a challenging high-level visual recognition task: natural scene perception. Throughout our lives we are exposed to a large number of different scenes, indoors and outdoors, human-made and natural, from different viewpoints and in various lighting conditions. Yet, in spite of this large variability in visual appearance (see fig. 1 for an illustration), it is fast and effortless for us to determine that a scene is a forest or a city or a beach. Being able to categorize natural scenes quickly is crucial for our orientation in new as well as familiar environments and for visual tasks such as navigation, fast object recognition, or obstacle avoidance (Tversky and Hemenway, 1983). Humans can categorize natural scenes with high reliability even if they have never seen the particular scene before (Thorpe et al., 1996; Fei-Fei et al., 2007).

[figure 1 about here]

Behavioral studies utilizing rapid serial visual presentation (RSVP) streams of scene photographs have highlighted the remarkable speed and ease with which humans can process natural scenes. Early work established that viewers required as little as 250 ms to achieve satisfactory recognition performance (e.g., Potter and Levy, 1969). Biederman (1972) showed that a brief presentation (300 ms–700 ms) of a coherent natural scene improves memory for objects at locations that are subsequently cued, when compared to performance on the same task with jumbled scenes. In keeping with the literature documenting observers' ability to extract information from very brief presentations of natural scenes, event-related potentials recorded while subjects made judgments about whether an image contained an animal or not (or a vehicle or not) revealed a differential signal for categorical decisions as early as 150 ms after stimulus onset (Thorpe et al., 1996; VanRullen and Thorpe, 2001). Moreover, observers can categorize scenes rapidly presented in peripheral vision even when attentional resources are engaged in a highly demanding task at fixation (Li et al., 2002; Fei-Fei et al., 2005), and they can effortlessly switch the detection task from one category to another in as little as 50 ms (Walther and Fei-Fei, 2007). Finally, even in a single glance, many details of natural scenes are accessible to observers (Fei-Fei et al., 2007).

Although this body of research has established that the human visual system is adept at processing scenes, very little is known about the neural mechanisms that underlie this ability. How is it, for instance, that the brain determines whether it is looking at a forest or a city skyline? Because the neural representation of natural scene categories is likely to be far more distributed than such superordinate categories as places or faces, it is less amenable to discovery by traditional neuroscience methods. This makes natural scene categorization an ideal case for the application of MVPA techniques. Specifically, we use pattern recognition algorithms to ask whether information distinguishing different scene categories is present in various regions of the brain.

Our primary goal is the exploration of the human ability to categorize natural scenes. We therefore need to look beyond the mere readout of scene categories from the brain. We want to determine which activation patterns are most closely related to humans' perception. We achieve this by comparing the types of errors made in decoding scene category from the fMRI data with the types of behavioral errors made by the subjects when categorizing natural scenes.

2. The role of pattern analysis in interpreting fMRI data

Pattern analysis has become popular as an analysis tool for fMRI data in recent years. It has been used, for instance, to demonstrate the distributed nature of the representation of objects and faces (Haxby et al., 2001; Carlson et al., 2003; Cox and Savoy, 2003; O'Toole et al., 2005; Kriegeskorte et al., 2008), to access population codes of the representation of visual information in primary visual cortex (Haynes and Rees, 2005; Kamitani and Tong, 2005, 2006; Kay et al., 2008), and to decode the mental state of subjects (Polyn et al., 2005; Haynes and Rees, 2006; Haynes et al., 2007). Rather than analyzing the overall level of activity in a particular area of the brain, in these studies the specific pattern of activity was analyzed.

To this end, data are represented as vectors in high-dimensional spaces, whose dimensions typically represent the values of individual voxels or, after dimensionality reduction, the related rotated and scaled coordinates. These data vectors are sorted into groups according to experimental conditions such as the orientation of gratings, identity of objects or scenes etc. It is the goal of pattern recognition algorithms to separate those vectors belonging to one condition from those belonging to another in a manner that generalizes well to new data not seen before by the algorithms. This goal is achieved by computing the parameters of decision boundaries between the vectors belonging to the experimental conditions. In fMRI analysis we typically have many dimensions (voxels) to deal with, but only relatively few data points (brain acquisitions) for each condition. For this reason, most MVPA algorithms use hyperplanes as the simplest decision boundaries possible. Such algorithms are called “linear” and include, among others, correlation analysis, support vector machines with linear kernels, Gaussian Naïve Bayesian classifiers (when used with shared covariances among categories), and linear discriminant analysis. The hyperplanes found by these algorithms bisect feature space: the data in one half correspond to one experimental condition, the data in the other half to another condition. The goal is for this division of feature space to generalize to new experimental data, allowing for the generation of predictions, which can then be confirmed or refuted.

An important question remains, however: How is the decision boundary in this high-dimensional voxel space related to the representation of, say, visual percepts that people access when they respond to the stimuli? These pattern recognition algorithms are powerful statistical constructs, pulling out regularities in the data patterns that allow for the discrimination of conditions, irrespective of their relations to human behavior. In doing so these algorithms may rely on properties of the data that are irrelevant to the human participant, and whose variations among experimental conditions could even be a mere coincident.

This point is nicely illustrated by an anecdote from the early years of pattern recognition research[1]. Researchers designed neural networks for image recognition for the military. They wanted their neural networks to recognize tanks in photographs. They trained the networks with many pictures of tanks as well as pictures without tanks and let the network weights converge. After training was complete, the network was almost perfect at telling which photographs contained tanks and which did not – the researchers rejoiced. When they finally decided to go out and take new pictures to test their tank recognition system, however, the researchers were shocked to discover that the predictions of the network were no better than random guesses. Further investigation revealed that the neural network had not learned anything about tanks at all. It turned out that all of the original training pictures with tanks had been shot on a sunny day and all of the negative examples on a cloudy day. The network had in fact learned to discriminate between bright and dark images and not between images with and without tanks.

Since those days, the theory of pattern recognition has come a long way (e.g., Bishop, 2006). We are now much more aware of problems such as overfitting, quality of generalization, and clean separation of training and test data (Hanson et al., 2004; O'Toole et al., 2007). We also have statistical techniques at our disposal that help us ameliorate these pitfalls (Mitchell et al., 2004).

However, we are still faced with the question: Do the patterns learned by our algorithms correspond to the patterns of neural activity being used by the brain? We are not simply asking whether the algorithms have learned sensible decision boundaries that accurately generalize to new data. We are interested in using the algorithms to tell us something about how the human brain works beyond the knowledge of experimental conditions being decoded from a brain area. How do we know that the algorithms are using the same information that human subjects are accessing in the experiment?

Fortunately, the nature of the problem also gives us constraints that help us solve it. In many cases we already know how the brain processes certain types of stimuli. And if we have a model of the function of particular brain regions, we can use it to guide the interpretation of the fMRI data. This approach has been demonstrated by Kay et al. (2008) for fMRI data read out from primary visual cortex (V1). In the training phase of their experiment, Kay et al. showed their two subjects 1750 images of natural scenes and used the resulting fMRI data from V1 to fit a receptive field structure encompassing location, orientation, and spatial frequency selectivity for

[1] http://www.bbc.co.uk/dna/h2g2/A413687

each voxel. Equipped with this knowledge, they could predict with high accuracy (82% average over two subjects) which of 120 previously unseen images the subjects were viewing in the test phase of the experiment. It was crucial to this effort that Kay et al. had a good model of the receptive field properties in V1.

A similarly-minded approach was used by Mitchell et al. (2008) to predict the fMRI activity elicited by 60 concrete nouns. A model of intermediate semantic features was fitted based on the frequency of co-occurrence of the nouns with a set of 25 sensory-motor verbs in a large text corpus. The activity of each voxel was modeled as a linear combination of these 25 semantic features. The fMRI activity of 58 of the 60 nouns was used to find the weights for this linear combination, which were then used to predict the activity patterns elicited by the two left-out nouns. This prediction was better than chance: matching up the predicted and the actually elicited activity patterns, allowed the researchers to tell which of the two left-out fMRI images belonged to which of the two left-out nouns in 77% of the cases (chance level: 50%). Here again, having a prior model based on a large corpus of background information was instrumental for achieving good predictions.

But for the majority of the cortical areas in the brain and for most brain functions, we do not have such explicit models. Here we demonstrate an entirely different, equally powerful approach for moving beyond decoding accuracies. To scrutinize whether patterns of fMRI activation are fundamentally related to subject performance, we evaluate the MVPA algorithms against patterns of human behavior. There has been a long history of behavioral experiments to study visual perception going back, at least, to Wilhelm Wundt (Wundt, 1874) and his teacher Hermann von Helmholtz (von Helmholtz, 1925/1909). Comparisons with behavior can serve as an important constraint when using powerful statistical methods to extract signal from brain data. We propose that relating activity patterns to behavioral data with MVPA can be used as an effective tool to gain new insights into the neural mechanisms of visual perception and cognitive processing. We illustrate this point with examples from our own work on natural scene categorization.

3. Relationship between behavioral data and fMRI decoding

Behavior arises from brain activity. It is thus natural to combine measurements of brain activity and behavior in order to learn more about neural function. Neuroimaging has the advantage of providing us with information about patterns of activity in the entire brain all at once at a reasonable spatial resolution. What neuroimaging does not tell us is whether subjects actually use these patterns to perceive the stimulus, solve the task, or generate a response. Analysis of behavior, on the other hand, gives us insight into a subject's experience, and this can be used to constrain the interpretation of fMRI data.

In the concrete example of natural scene categorization we will see how comparisons of error patterns between fMRI decoding and behavior help us to establish which brain areas are more closely related to human perception. To get subjects to make errors in categorizing natural scenes, the presentation times of the images needed to be short and followed by a perceptual mask. These experimental conditions are not optimal for fMRI experiments, however. Instead, we used image presentation times of more than a second to ensure a strong visual signal for each image. Furthermore, since we expect scene categories to differ in subtle ways, we are relying on the power of a block design to strengthen our category-related signal. However, since our ability to decode scene category in our fMRI experiment and participants' ability to discriminate briefly presented scene categories are both predicated on the quality of the category-specific representation in the brain, we predict that these two measures will be correlated

Decoding patterns of fMRI activity assumes that there is a category-specific signal present in the brain, and that we can use classification algorithms to tease some part of the signal out of the noisy fMRI data. The weaker or less distinct the category-specific signal is, the harder it will be to read out that signal. This will cause the decoding algorithm to make more mistakes. If the subject is relying on the same category-specific signal to make a decision about the category membership of a presented scene, then her error rate, too, should go up if the signal is weaker. This is true in particular for short presentation times when accumulation of evidence about the stimulus is disrupted by a perceptual mask. Therefore, if certain stimuli are more confusable than others, then we should see higher error rates for these stimuli in the behavioral response as well as in decoding from fMRI activity in areas that contribute to the behavioral decision.

[figure 2 about here]

For a more formal treatment of the relationship between accuracy measures in fast forced-choice experiments and the signal decoded from fMRI data as a race-diffusion model of decision making in the brain (Ratcliff, 1985; Bogacz, 2007) see figure 2. This simple model illustrates how a noisier representation of the stimulus in the brain (weaker evidence for the stimulus) can lead to higher error rates in a fast two-alternative forced-choice experiment with short presentation times. By the same argument we would expect to see higher rates of confusion for stimuli whose representations are more similar in experiments with more than two choices.

It is clear that behavior should be taken into account when trying to understanding brain functions intimately tied to behavioral output, such as motor responses or decision processes. In fact, signals decoded from fMRI activity in the motor area have been used to control a robotic hand in near-real time (Kawato, 2008). Similarly, in recent years, MVPA has been used to study executive brain functions such as decision making and planning (Haynes et al., 2007; Clithero et al., 2009).

On the other hand, when the representation of particular stimulus properties in primary sensory areas is the main interest, it may be more useful to correlate neural activity with physical properties of the stimuli rather than subject behavior. For investigating early visual processing, for instance in the primary visual cortex, it may be adequate to only use the input data as ground truth, since only one processing step is involved. The behavioral response, removed by many more processing steps from early visual processing, is presumably not as useful as ground truth in this case. Indeed several MVPA studies have used this approach to describe processing in the primary visual cortex (Haynes and Rees, 2005; Kamitani and Tong, 2005, 2006; Kay et al., 2008; Miyawaki et al., 2008). However, even in these cases, comparing neural activity with reported subjective experience has generated new insights (Haynes and Rees, 2005; Kamitani and Tong, 2005).

In the next section, we describe in more detail the fMRI and behavioral experiments on natural scene categorization that allowed us to compare a fine grained pattern of errors produced by the MVPA algorithm and human behavioral errors.

4. Natural scene categorization

Much of the progress in the visual sciences has been due to a strategy of decomposing visual scenes into simple, more tractable components. Of course, such features and objects are greatly simplified in comparison to the complex visual scenes we interact with every day. Despite their ecological importance, however, we know surprisingly little about how, or even where in the brain, we process scenes as a whole.

Humans are extremely efficient at perceiving natural scenes and understanding their contents. We tested this ability in a behavioral experiment, in which we showed five subjects photographs of natural scenes (see fig. 1 for examples) followed by a perceptual mask (Walther et al., 2009). We asked subjects to press one of six keys to indicate to which of the six categories *beaches, buildings*, *forests*, *highways*, *industry*, or *mountains* each image belonged. To make the task more difficult we decreased the presentation time (stimulus onset asynchrony, SOA) of the images in a staircasing procedure using the Quest algorithm (King-Smith et al., 1994) down to a target performance of 65% correct responses. We needed to use an SOA of as little as 11–45 ms (depending on the individual) to get subjects to make enough mistakes in categorizing the scene images.

[figure 3 about here]

We recorded the mistakes in a confusion matrix (fig. 3), which shows, for instance, in what fraction of cases of being shown a beach (first row) subjects responded by pressing the key that belongs to highways (fourth column). For a subject with perfect performance, the confusion matrix would have ones on the diagonal and zeros in all off-diagonal fields. As can be seen in figure 3, the diagonal entries are considerably larger than the off-diagonals. The mean of the diagonal elements (77%) is the accuracy of subjects in the behavioral experiment, which is significantly above the chance level of 1/6. In fact, categorization accuracy was significantly above chance for each of the six categories.

We also conducted an fMRI experiment, in which subjects passively viewed blocks of 10 images of the same category with each image presented for 1.6 s. Six blocks (one for each category) were combined in a run, and alternating runs contained upright and up-down inverted images. We discuss the utility of the inverted images in the next section. Functional imaging was performed on a 3 Tesla Siemens Allegra Scanner. After minimal preprocessing (motion correction and normalization to the temporal mean of each run) we extracted the brain volumes corresponding to the blocks of image presentation with a time lag of 4 s to approximate the lag in the hemodynamic response.

We trained a linear support vector machine (SVM) in a 6-way classification task to predict natural scene category based on the fMRI activity during blocks with upright images. In a leave-one-run-out (LORO) cross validation procedure, one of the upright runs was held out, the SVM was trained on the data from the other runs, and predictions were created for the scene categories viewed in the left-out run. The process was repeated until each of the upright runs was left out in turn, thus generating predictions for each of them. Correct classification rate was computed as the fraction of blocks in which the predicted scene category matched ground truth.

We applied LORO cross validation to the voxels of several regions of interest (ROIs), which were determined in separate localizer scans. Known to be involved in the processing of scenes in general, the parahippocampal place area (PPA) and the retrospenial cortex (RSC) were included as likely candidate regions for scene categorization (Aguirre et al., 1996; Epstein and Kanwisher, 1998; O'Craven and Kanwisher, 2000). Its sensitivity to a variety of objects (Malach et al., 1995) made the lateral occipital complex (LOC) a potentially interesting region, because scenes can be construed as collections of objects. We also included the fusiform face area (FFA) as an area involved in the processing of complex visual stimuli (Grill-Spector et al., 1999; Gauthier et al., 2000; Tarr and Gauthier, 2000; Haxby et al., 2001), although its primary sensitivity to faces (Kanwisher et al., 1997) might not suggest a role in scene categorization. The primary visual cortex (V1) was included in the analysis, because different scene categories may differ in properties coded in V1, such as their spatial frequency content (Oliva and Torralba, 2001) or the distribution of local texture (Fei-Fei and Perona, 2005; Bosch et al., 2006).

[figure 4 about here]

Results for these regions of interest (ROI) are shown in figure 4. Classification rate from LORO cross validation was significantly above chance in V1, LOC, RSC, and PPA, but not in FFA. What does this tell us? It means that there is some kind of information present in the voxel patterns in each of these ROIs that allows a linear classifier to predict scene category more accurately than the throw of a die. It does not tell us, however, whether this is the kind of information that human subjects would use to make the category decision. This information could be correlated with stimulus attributes but not instrumental in subjects' judgments.

It would therefore be desirable to have a closer link between the scene category-specific information in the fMRI activation patterns and what humans actually do. We achieve this by comparing the pattern of errors. We can establish a confusion matrix for the classifier predictions, similar to the one we obtained from the behavioral experiment (fig. 5). In the fMRI decoding confusion matrix, rows represent the categories shown to subjects, and each entry gives the fraction of blocks, for which the classifier predicted this block to belong to the category indicated by the column. As before, the diagonal entries are correct classifications (omitted in fig. 5), and all off-diagonal entries are errors. It is the particular pattern of errors that allows us to compare fMRI decoding results with behavioral performance. As can be seen in figure 5, the error pattern from decoding PPA activity is more similar to the behavioral error pattern than the one decoded from the FFA. Pairwise correlations of the off-diagonal entries of the confusion matrices give us a quantitative measure of the similarity of error patterns. We find high correlation with behavior for PPA, RSC, and LOC, but not for V1 and FFA (fig. 5).

[figure 5 about here]

This result indicates that the patterns of fMRI activity that we see in PPA, RSC, and LOC are more closely related to the information used by human subjects in the behavioral experiment. This is compatible with the known selectivity of PPA and RSC for natural scenes (Aguirre et al., 1996; Epstein and Kanwisher, 1998; O'Craven and Kanwisher, 2000; Maguire, 2001; Epstein and Higgins, 2007). So far we can only speculate about the role of object-sensitive LOC in natural categorization. Objects can often indicate particular scene categories (Hollingworth and Henderson, 2002; Davenport and Potter, 2004). For example a beach umbrella indicates that the scene is a beach, and a particular type of traffic sign may indicate that the scene is a highway scene. Indeed, Bar and Aminoff (2003) have reported activity in LOC for a comparison of strong object-scene associations (i.e., object is very typical for the scene) versus weak associations (i.e., object can occur in this context but also in many others). It is also possible that information flows the opposite way, with global scene information providing context for object detection (Biederman, 1972; Bar, 2004). Our current experiments do not allow us to determine the nature of the connections between these areas.

5. Effects of image manipulations on fMRI decoding and behavior

Another way of comparing the neural signal decoded from patterns of fMRI activation with subject behavior is the manipulation of stimuli in such a way as to change behavior. If we can establish a change in the behavioral performance, then we can compare how this change affects the neural representation that we can decode from the fMRI activity in different parts of the brain.

Scene inversion

For example, in our experiment with natural scenes we also presented images that were up-down inverted. In the behavioral experiment, categorization accuracy was lower for inverted scenes compared to upright scenes (65% versus 77%). Presumably, this is because the layout information in the image, which contributes to the correct categorization, gets disrupted by the inversion process. Our subjects can still perform the task well above chance, either by using other image features that are invariant to inversion such as textures and colors, or by executing some form of mental rotation, but the drop in accuracy due to inversion is significant ($p < 0.001$).

In the fMRI experiment we included a run with inverted images following each run with upright images. In each inverted run, the same images were presented in the same order as in the preceding upright run, except that they were up-down inverted. We applied the same leave-one-run-out cross validation procedure as described in the previous section to the inverted runs and compared the classification accuracies obtained from our ROIs with the accuracies from the upright runs.

The significant drop in behavioral performance for inverted relative to upright scenes was mirrored in the decoding performance in some but not all of our ROIs. Decoding accuracy for inverted runs was significantly lower in the PPA (fig. 4). It also decreased in the LOC, RSC, and FFA. However, there was no difference in the decoding accuracy for V1. In other words, it did not make a difference to the V1 decoder whether it was trained and tested on upright or inverted image; it decoded them all equally well. These results are consistent with our assertion that inversion disrupts global scene layout but not local image features such as texture and color. PPA is thought to be sensitive to scene layout, and indeed it is most affected by inversion. In contrast, V1, which encodes local features, showed no decrement in decoding accuracy for inverted scenes. Furthermore, based on these correlations with behavior, we can conclude that the information distinguishing natural scene categories in the PPA is much closer to what humans use to make their decision than is the category-related information in V1.

Just as for upright images, we can look for matches between decoding and behavior for inverted images. If it is true that for inverted scenes we rely more on local texture cues than on layout and context, then we should expect a decrease of the error correlations in areas thought to process layout and context, such as PPA and RSC, while error correlation in V1 should remain largely unaffected. This is in fact what we see for these three areas: error correlation in RSC drops from 0.34 to -0.09 for inverted images, and in PPA from 0.57 to 0.32, while error correlation in V1 is slightly higher for inverted images at 0.31, compared to 0.21 for upright images. It is also possible to train the decoder on the data from upright images and test it on data from inverted images. We discuss this possibility in detail elsewhere (Walther et al., 2009).

Note that while the presence of correlation with behavior provides evidence in favor of a brain region's participation in the behavior, its absence does not mean that the region does not contribute to the task in question. There is no doubt that V1 and other retinotopic areas are involved in the processing of natural scenes, for instance. Finding correlation of fMRI decoding with behavior means that the information in the fMRI data is similar to the information used by the subject to generate her behavioral response.

Good and bad category exemplars

Another way of manipulating behavioral performance of natural scene categorization is controlling how well the images presented in the experiment conform to their categories. For images that are bad exemplars we should expect a weaker category-related signal in the fMRI data than for images that are good exemplars. To address this question we had 4025 color images from 6 different categories (*beaches*, *city streets*, *forests*, *highways*, *mountains* and *offices*) rated as to how good of an exemplar each image was for its category on a scale from 1 to 5 (Torralbo et al., 2009; Torralbo et al., under review). Observers also had the option to say that the image was not an exemplar for the particular category at all. In this case the image was eliminated from the set. Based on the ratings we selected 80 good and 80 bad images for each of six natural scene categories to be used in subsequent behavioral and fMRI experiments (see fig. 6 for examples).

[figure 6 about here]

In a behavioral experiment similar to the one in the aforementioned inversion study, subjects performed a six-alternative, forced-choice classification of the images with short presentation times and masks for the images. We found significantly higher classification accuracy for good than for bad exemplars.

We then conducted an fMRI experiment similar to the inversion experiment, except that instead of inverted and upright scenes we presented good and bad upright scenes. Runs containing only good or only bad images were randomly interleaved. In a leave-two-runs-out cross validation procedure we trained a decoder on runs with good and bad exemplars, leaving out one good and one bad run. The decoder was then tested on the good and the bad run separately, generating predictions of the scene categories in the left-out runs. The procedure was repeated until each run was left out once.

In agreement with the behavioral categorization results, scene categories were predicted significantly more accurately for the runs with good exemplars than for the runs with bad exemplars – but only in the PPA and the RSC (Torralbo et al., 2009; Torralbo et al., under review). Decoding accuracy in the LOC and in area V1 did not differ in the two conditions. The higher decoding accuracy for good than for bad exemplars in the PPA and RSC was not due to a higher BOLD signal in these areas for good exemplars. This match between behavioral results and decoding results from PPA and RSC provides more evidence that these two areas contain information about scene categories that is vital for making decisions about category membership of natural scenes.

All of the examples described so far have one important fact in common: they illustrate the importance of comparing fMRI decoding results with behavioral experiments as a means of assessing how likely it is that decoding performance seen in particular brain areas is related to human perceptual decisions. Note that these experiments were hypothesis-driven: they looked at the activity in ROIs that were likely to be involved in the task at hand. In the next section we show how comparisons with behavior can be used as an exploratory tool.

6. Correlation with behavior as an exploratory tool

Much of the neuroimaging work in visual perception now typically assesses signal in a list of candidate ROIs, which are established in separate mapping or localizer scans. This reflects a degree of maturity in our understanding of the human and non-human primate visual system. However, in developing a new measurement tool, as we are here, we would like to move beyond ROIs established by univariate methods and discover new regions with multi-voxel pattern analysis, akin to a whole-brain analysis in conventional fMRI processing.

Kriegeskorte et al. (2006) introduced the idea of moving a spherical “searchlight” region (known as a “scanning window” in computer vision) to all potential locations in the brain and performing statistical analysis on the voxels in each small spherical neighborhood. In their study, Kriegeskorte et al. compared experimental conditions by locally computing the Mahalanobis distance of the voxel activities in one condition from the activity of the same voxels in another condition. This distance was interpreted as a measure for the information contained in a particular small neighborhood with respect to the experimental conditions. A similar method was used by Haynes et al. (2007) to decode hidden intentions of subjects planning to add or subtract two subsequently presented numbers. In this study, a support vector machine (SVM) was used to discriminate between the two possible states of intent. The searchlight procedure was applied to both the planning phase and the execution phase of the algebraic operation, and different frontal regions were identified in these two cases.

Here we demonstrate how the same idea of analyzing the voxels in a small local region can be used to establish a map of correlations with behavior in the entire brain. To this end we analyzed the data from our 6-way natural scene categorization experiment with a new behavior-correlation searchlight procedure. We established a spherical searchlight region with a radius of 2.5 voxels (8.6 mm), containing 81 voxels, which is similar in size to the ROIs used above. We centered the sphere on each voxel in the brain in turn and performed the same leave-one-run-out cross validation procedure described before, but using the voxel values within the sphere instead of the pre-defined ROIs (fig. 7).

[figure 7 about here]

As a result we obtain two complete brain maps, one of decoding accuracy, and one of correlation coefficients of decoding errors with behavioral errors. To generate these maps we first compute separate maps for each of the 36 entries in the confusion matrix. We combine these maps across subjects by registering them into standard MNI space, spatially smoothing, and averaging them across five subjects. Then we average the six maps corresponding to the diagonal entries of the confusion matrix (i.e., correct decoding) to yield the brain map of decoding accuracy. The values of the maps corresponding to the 30 off-diagonal entries of the confusion matrix (decoding errors) are correlated with the off-diagonal entries of the combined behavioral confusion matrix at each voxel location, and the Pearson correlation coefficients are stored in the behavior-correlation brain map. For further analysis we applied a threshold of $P < 0.01$ (significance of Pearson correlation) to the behavior-correlation map. Correction for multiple comparisons was performed at the cluster level based on an estimate of the spatial correlation among voxels, implemented with AlphaSim from the AFNI toolbox (Cox, 1996), resulting in a minimum cluster size of 19 voxels.

[figure 8 about here]

Since comparison of decoding with behavior only makes sense in brain regions that contain information related to scene category in the first place, we only consider regions that also have decoding accuracy significantly above chance ($P < 0.01$, t test over five subjects). Regions that survived significance tests for error correlation as well as decoding accuracy were located in the posterior parahippocampal gyri, the fusiform gyri, the right posterior inferior temporal gyrus, anterior parts of both middle occipital gyri, and in the posterior right precuneus. These areas included, but were not limited to some of the ROIs from the original study. For comparison, we computed the overlap of our original ROIs with the intersection of the above-chance maps for behavior correlation and decoding accuracy for each subject separately in MNI space. Figure 8 shows the overlap with the ROIs for one subject. Table 1 gives a summary of the results over all five subjects.

Table 1. Overlap of ROIs with the searchlight map for behavior-correlation intersected with decoding accuracy in percent of ROI voxels, summarized over five subjects.

ROI	*mean*	*standard deviation*
V1	4.7 %	1.7 %
FFA	4.1 %	8.4 %
LOC	3.5 %	4.7 %
RSC	25.7 %	4.4 %
PPA	28.6 %	13.1 %

These overlap results indicate that a large number of voxels uncovered by our searchlight analysis coincided with our PPA and RSC voxels. In keeping with the lower behavioral correlation in V1 and FFA, the behavioral correlation searchlight found very few voxels in these areas. The searchlight analysis also uncovered very few voxels in LOC, despite the significant behavioral correlation revealed in the ROI analysis. This discrepancy is likely due to small differences in the size of the LOC ROIs and the searchlight window used. Overall, this comparison serves as a validation of the searchlight analysis of correlation with behavior.

The regions found with this exploratory searchlight analysis further uncovered new potential ROIs, including parts of the fusiform gyri, the right posterior inferior temporal gyrus and anterior parts of both middle occipital gyri not included in our ROIs, as well as the posterior right precuneus. Future experiments might test the properties of these regions with an ROI-based approach. It is important to note that it would not be valid to perform this ROI-based analysis on the same data set that was used for the searchlight analysis. The correct category labels were already used to compute the confusion matrix entries for each searchlight location, irrevocably spoiling the data for any further classification analysis. Using these results to select voxels for an ROI-based analysis would create fallacies of the kind recently highlighted by Kriegeskorte et al. (2009).

7. Correlations with behavior in other domains

Correlating fMRI decoding with behavior has also been applied successfully to object and shape perception. Williams et al. (2007) have incorporated behavioral information in an analysis of object representations in retinotopic cortex and the LOC. In their experiment, Williams et al. scanned subjects performing an object categorization task. They then analyzed the correlations of activity patterns for same versus different categories separately for trials with correct responses and trials with incorrect responses. They found that retinotopic cortex did not care whether the subject responded correctly or not: the correlation for same category stimuli was significantly higher than for different category stimuli irrespective of subjects' response. In the LOC, on the other hand, the correlation for same category stimuli was higher than for different category stimuli only in correct trials, but not in incorrect trials. This suggests that the relation of fMRI activity in the LOC to subjective perception is closer than that of early visual areas. Here again we should note that these results do not indicate that a region apparently unrelated to behavior (as retinotopic cortex) does not play a critical role in the task. Indeed retinotopic cortex must have provided the major portion of the signal to LOC.

Another study investigated the distributed representation of shape in different parts of the LOC using a combination of fMRI scanning and behavioral experiments (Haushofer et al., 2008). In the behavioral part of the study, participants performed a two-alternative forced-choice task to decide whether two successively shown shapes were the same or different. The similarity matrix from this task was compared with the correlation matrix obtained from the activity elicited by these shapes in an fMRI experiment. It was found that activity in the anterior LOC correlated much better with behavior than activity from posterior LOC. When the fMRI data were compared with a similarity matrix derived from the physical properties of the shapes, the opposite pattern was observed: activity from posterior LOC correlated better with physical similarity than activity from anterior LOC. This suggests that the shape representation in the posterior LOC is driven mostly by the physical shape of the presented objects, whereas activity in the anterior LOC includes subjective perception by the subjects.

Another example of correlating fMRI decoding with behavior is a recent study addressing the ability of native speakers of English and Japanese to discriminate /la/ from /ra/ sounds (Raizada et al., 2009). In this study, patterns of fMRI activity in right Heschl's gyrus (primary auditory cortex) were used to predict differences in the behavioral ability among individuals to discriminate among these sounds.

The correlation of error patterns between fMRI decoding and behavior is closely related to similarity analysis (Aguirre, 2007; Kiani et al., 2007; Kriegeskorte et al., 2008). Instead of representational dissimilarity matrices (RDMs), we use confusion matrices obtained from a classification-based analysis as a measure of similarity. Unlike the correlation-based RDMs, the confusion matrices are not necessarily symmetric. That is, there can be biases for confusing, say, *beaches* for *highways* but not, or not as often, *highways* for *beaches*. Therefore, confusion matrices are more closely related to decision processes about the stimuli than correlation matrices, which can be seen as comparing stimulus similarity. While Kriegeskorte et al. (2008) compared similarity structures across species (humans and macaque monkeys) and across neurophysiological measurement techniques (fMRI and multiple unit recording), here we compare the similarity structures between fMRI decoding and behavior of the same human subjects. This allows us to draw inferences about the relation of the neural representations in particular areas with the information accessed by humans in generating their behavioral response.

8. Conclusion

In this chapter we have demonstrated how relating fMRI decoding to behavior can give us new insights into the neural representation of visual information. Specifically, we have shown that the PPA, RSC, and LOC contain natural scene category information that is closely related to the information that human subjects access when they categorize natural scenes in a behavioral experiment. The patterns of errors made when decoding from these areas match well with the error patterns in the behavioral experiment. The prominent role of the PPA in scene categorization is further highlighted by the fact that accuracy of decoding from PPA significantly decreases for inverted scenes, just as accuracy in the behavioral experiment does, and by the fact that, similar to behavioral performance, decoding accuracy from PPA is significantly lower for bad than for good exemplars of scene categories.

The specific pattern of errors provides a richer description of the data than mere decoding accuracy. While we have focused on comparing fMRI data with behavioral data in this chapter, the same principle of correlating error patterns can be applied to uncover relations between brain regions, or even to compare data across different measurement techniques. As long as the data can be used to generate a prediction about the stimulus it is possible to generate a confusion matrix. The pattern of errors in the confusion matrix gives us one more standard with which to judge the contribution of a region to the brain function under investigation.

Finally, we have extended the use of the correlation of error pattern from ROI-based analysis to an exploratory whole-brain searchlight approach, which allows us to find new regions that are potentially involved in a behavior. We believe that these techniques have the potential for more wide-spread applications in visual neuroscience and beyond.

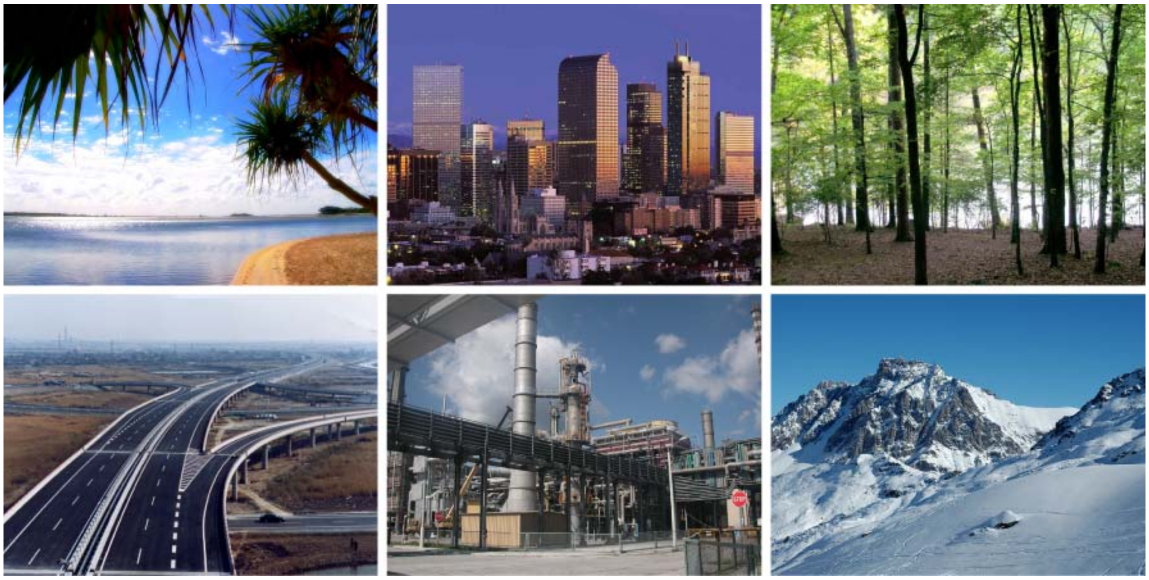


Figure 1. Example images for categories in our scene categorization experiment: *beaches*, *buildings*, *forests*, *highways*, *industry*, and *mountains* (from left to right and from top to bottom).

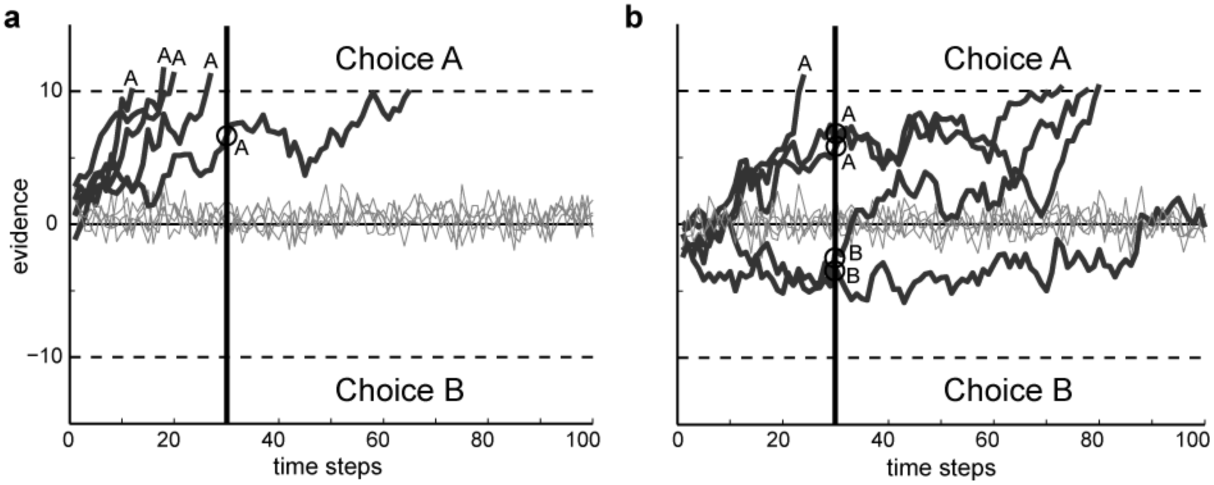
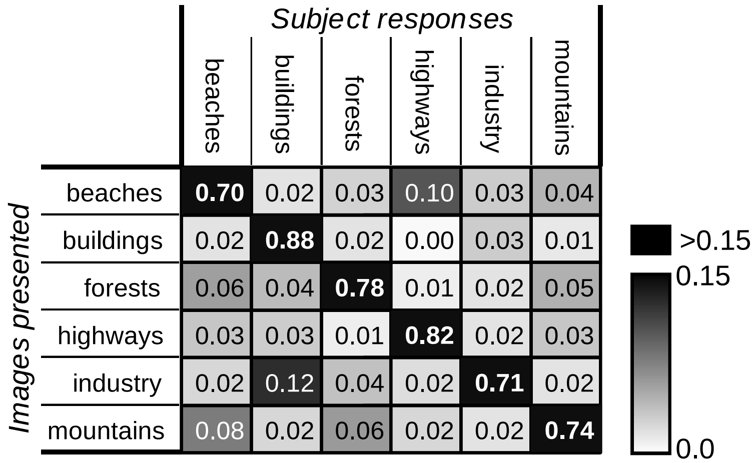


Figure 2. Race-diffusion model of decision making. Evidence in favor of choice A or B is accumulated over time until the cumulative evidence reaches a decision threshold for either choice (dashed lines). Each figure shows five instances of evidence drawn from a normal distribution (light gray curves) and the corresponding cumulative evidence (black curves). If accumulation of evidence is interrupted due to short stimulus presentation and masking (vertical black line), a decision must be made based on the evidence accumulated thus far. (a) Evidence is drawn from a normal distribution with a strong bias toward choice A ($\mu_{bias} = 0.5$, $\sigma = 1$). In all five simulation runs, the correct decision in favor of choice A is made, even in the case of interrupted evidence accumulation. (b) In the presence of weak evidence in favor of choice A ($\mu_{bias} = 0.05$, $\sigma = 1$), an erroneous decision is made for choice B in two out of five simulation runs when evidence accumulation is interrupted.



Images presented \ Subject responses	beaches	buildings	forests	highways	industry	mountains
beaches	**0.70**	0.02	0.03	0.10	0.03	0.04
buildings	0.02	**0.88**	0.02	0.00	0.03	0.01
forests	0.06	0.04	**0.78**	0.01	0.02	0.05
highways	0.03	0.03	0.01	**0.82**	0.02	0.03
industry	0.02	0.12	0.04	0.02	**0.71**	0.02
mountains	0.08	0.02	0.06	0.02	0.02	**0.74**

Figure 3. Confusion matrix for the behavioral scene categorization experiment. Diagonal entries are correct categorizations, off-diagonal entries are errors. The gray values reflect the frequencies of errors. For instances, *beaches* are frequently mistaken for *highways*, and images of *industry* are often confused with *buildings*.

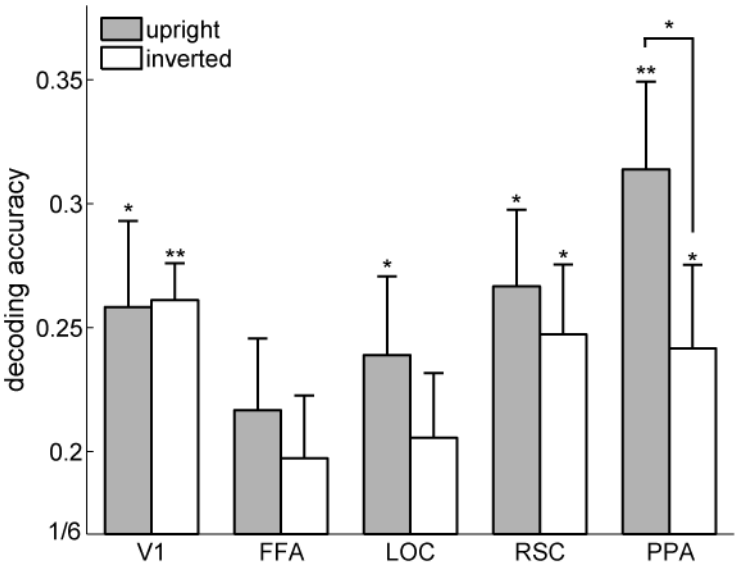


Figure 4. Accuracy of decoding scene categories from five regions of interest for training and testing using upright images (gray), and for training and testing using inverted images (white). Significance levels are with respect to baseline chance performance of 1/6th. Decoding accuracy is significantly above chance for V1, LOC, RSC, and PPA. However, only the PPA shows a significant difference in decoding accuracy between upright and inverted images. Error bars are s.e.m. over five subjects. *$p < 0.05$; **$p < 0.01$.

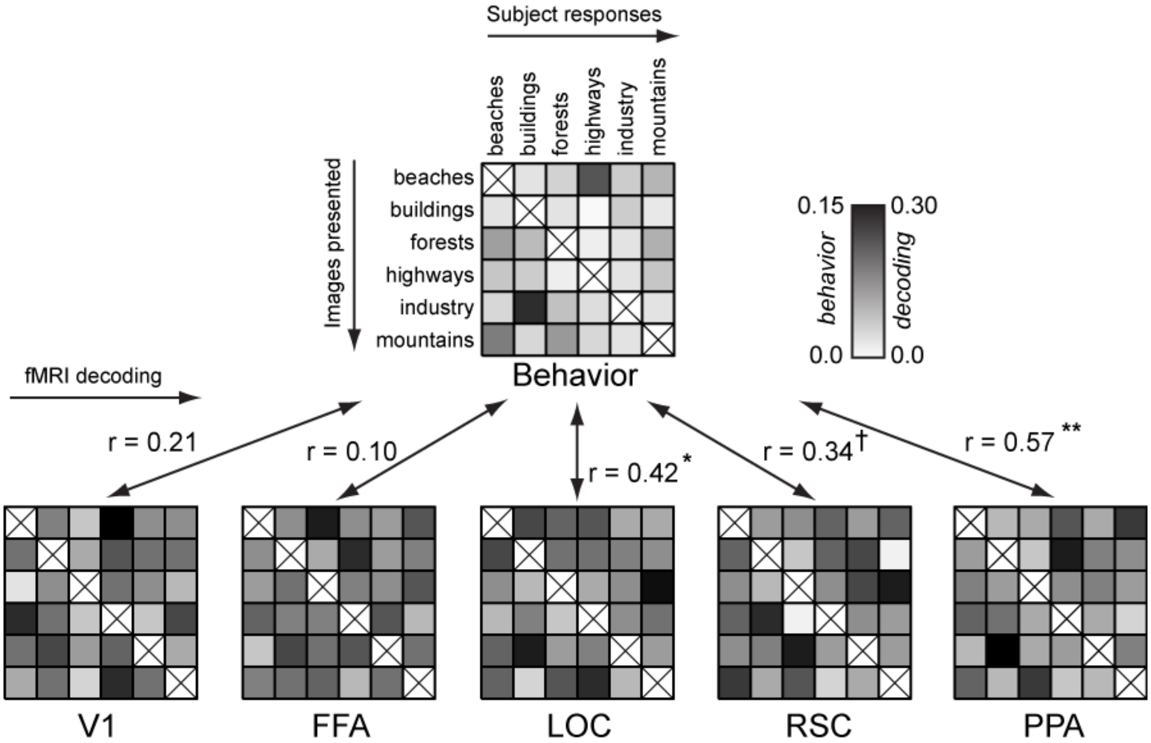


Figure 5. Confusion matrices for behavior and fMRI decoding with correlations of error patterns (off-diagonal elements). The rows of the matrices correspond to the image categories presented to the subjects. The columns in the behavioral confusion matrix indicate how frequently subjects responded with the respective category (see fig. 3). Correct responses are on the diagonal and are not shown in this illustration of error patterns. The entries in the fMRI decoding confusion matrices indicate the frequency of the decoding algorithm predicting the category corresponding to the column when the subject in fact saw images of the category corresponding to the row. All confusion matrices are averages over five subjects. Matches between fMRI decoding and behavior are computed as Pearson correlations of the off-diagonal elements. Good correlations are obtained for LOC, RSC, and PPA. *$p < 0.05$; **$p < 0.01$; $^{\dagger}p = 0.069$.

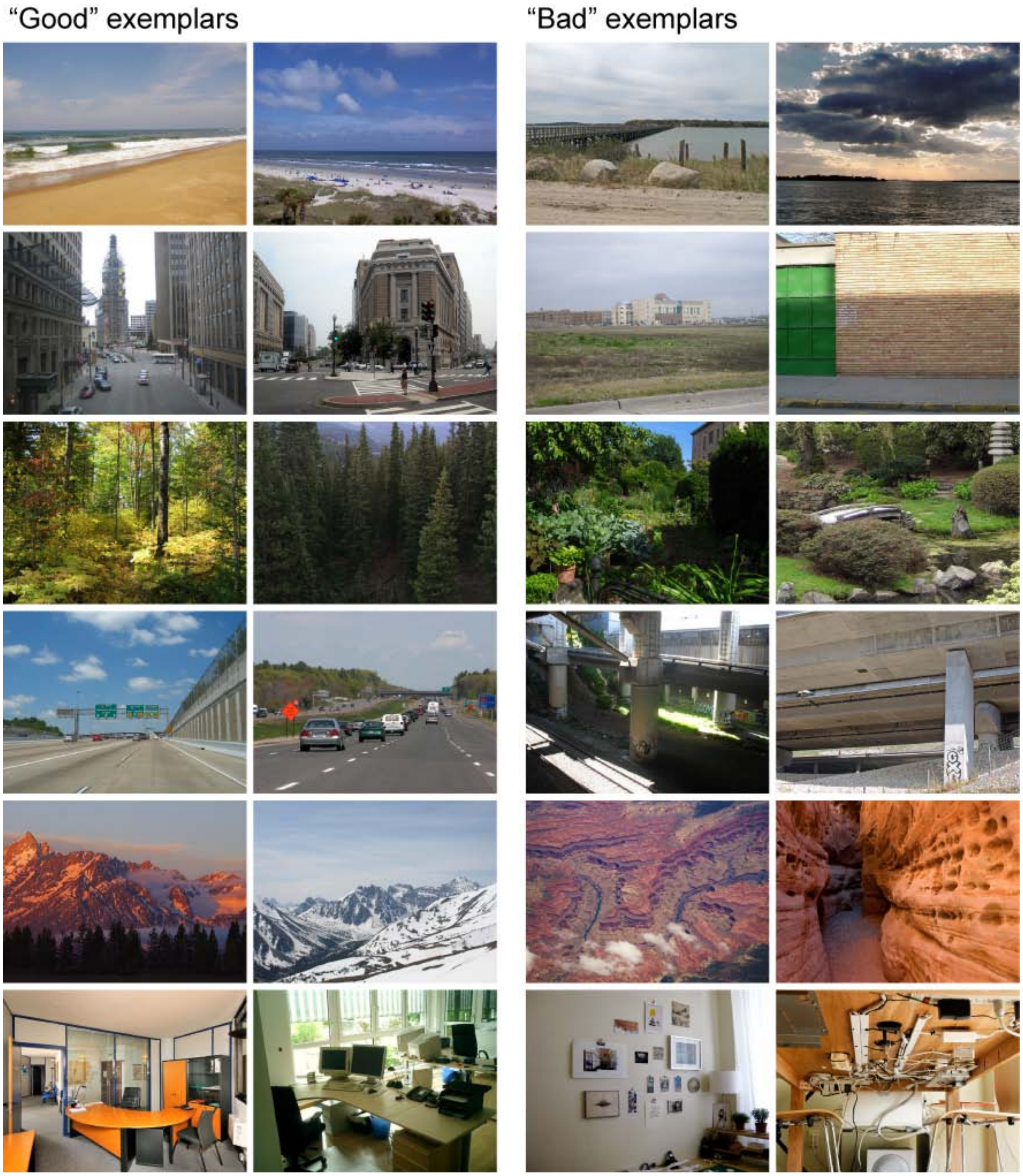


Figure 6. Examples of good (left two columns) and bad (right two columns) exemplars of natural scene categories *beaches*, *cities*, *forests*, *highways*, *mountains*, and *offices* (from top to bottom).

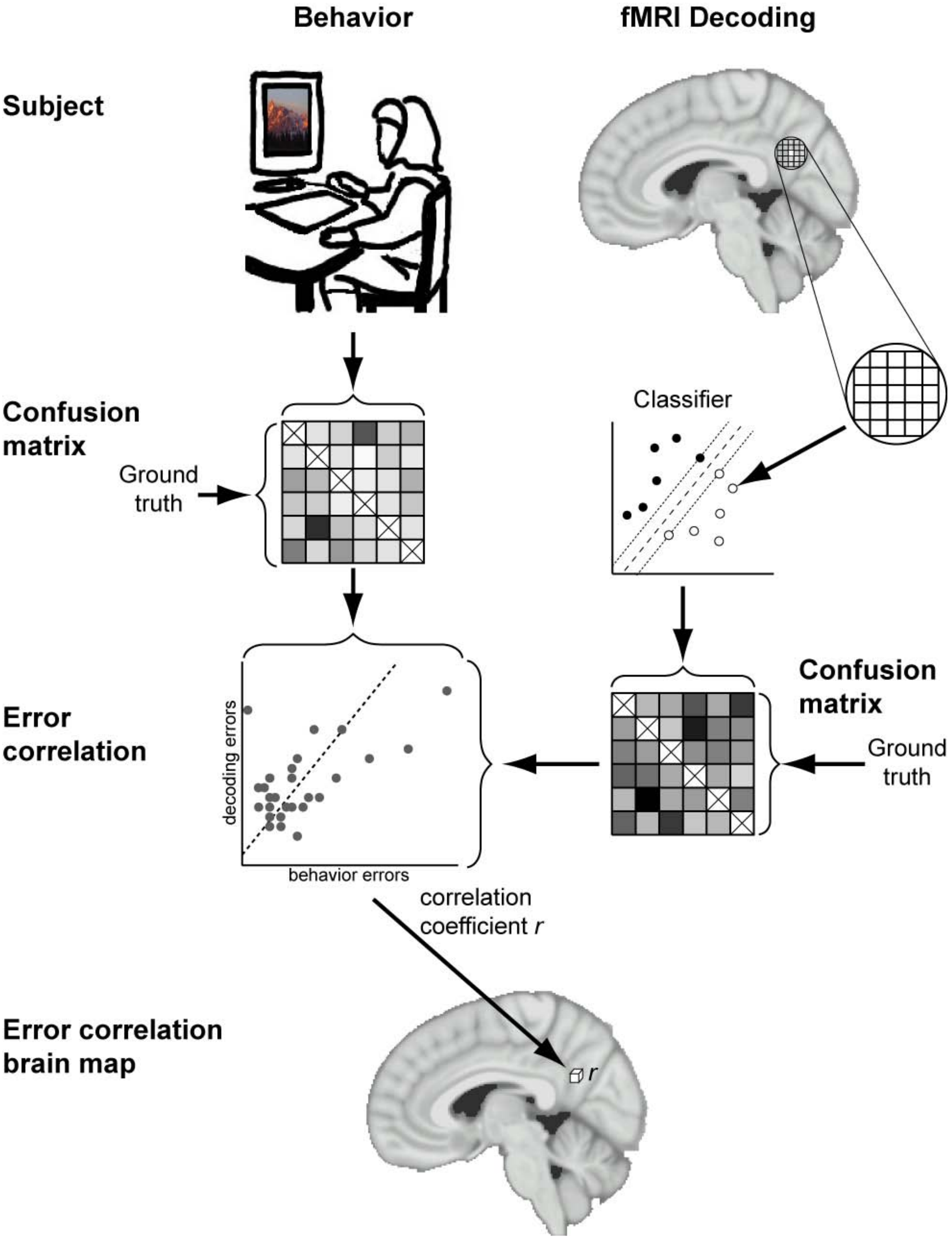


Figure 7. Illustration of searchlight analysis of correlation with behavior. The searchlight template is positioned at every location in the brain in turn, generating a whole-brain error correlation map.

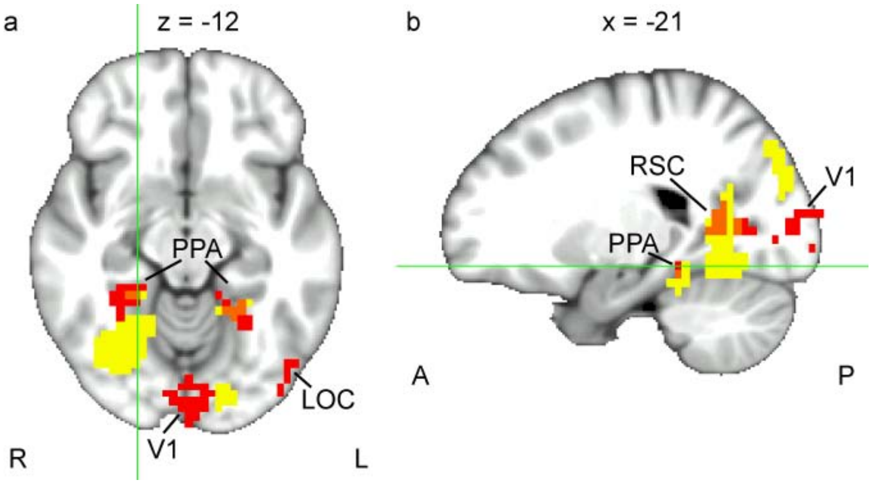


Figure 8. Searchlight map of error correlation intersected with decoding accuracy. (a) axial view (z = –12 mm); (b) sagittal view (x= –21 mm). This analysis was performed for five subjects in MNI space. Brain regions that show significant error correlation ($p < 0.01$) as well as significant decoding accuracy ($p < 0.01$) are marked in yellow. For comparison, the locations of PPA, RSC, LOC, and V1 for one of the five subjects are shown in red. Overlap between the subject ROIs and the searchlight area is marked orange. Multiple comparison correction was performed on the cluster level using AlphaSim.

References:

Aguirre GK (2007) Continuous carry-over designs for fMRI. Neuroimage 35:1480-1494.

Aguirre GK, Detre JA, Alsop DC, D'Esposito M (1996) The parahippocampus subserves topographical learning in man. Cerebral Cortex 6:823-829.

Bar M (2004) Visual objects in context. Nature Reviews Neuroscience 5:617-629.

Bar M, Aminoff E (2003) Cortical analysis of visual context. Neuron 38:347-358.

Biederman I (1972) Perceiving real-world scenes. Science 177:77-80.

Bishop CM (2006) Pattern Recognition and Machine Learning. New York, NY: Springer.

Bogacz R (2007) Optimal decision-making theories: linking neurobiology with behaviour. Trends in Cognitive Sciences 11:118-125.

Bosch A, Zisserman A, Munoz X (2006) Scene Classification Via pLSA. In: European Conference of Computer Vision.

Carlson TA, Schrater P, He S (2003) Patterns of activity in the categorical representations of objects. Journal of Cognitive Neuroscience 15:704-717.

Clithero JA, R.M. C, Huettel SA (2009) Local pattern classification differentiates processes of economic valuation. Neuroimage 45:1329-1338.

Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19:261-270.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research 29:162-173.

Davenport JL, Potter MC (2004) Scene consistency in object and background perception. Psychological Science 15:559-564.

Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN (1994) fMRI of human visual cortex. Nature 369:525.

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392:598-601.

Epstein RA, Higgins JS (2007) Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. Cerebral Cortex 17:1680-1693.

Fei-Fei L, Perona P (2005) A Bayesian Hierarchical Model for Learning Natural Scene Categories. In: IEEE International Conference on Computer Vision and Pattern Recognition.

Fei-Fei L, VanRullen R, Koch C, Perona P (2005) Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. Visual Cognition 12:893-924.

Fei-Fei L, Iyer A, Koch C, Perona P (2007) What do we perceive in a glance of a real-world scene? Journal of Vision 7:10, 11-29.

Gauthier I, Skudlarski P, Gore JC, Anderson AW (2000) Expertise for cars and birds recruits brain areas involved in face recognition. Nature Neuroscience 3:191-197.

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 24:187-203.

Hanson SJ, Matsuka T, Haxby JV (2004) Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? Neuroimage 23:156-166.

Haushofer J, Livingstone MS, Kanwisher N (2008) Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. PLoS Biology 6:e187.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425-2430.

Haynes JD, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nature Neuroscience 8:686-691.

Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. Nature Reviews Neuroscience 7:523-534.

Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE (2007) Reading hidden intentions in the human brain. Current Biology 17:323-328.

Hollingworth A, Henderson JM (2002) Accurate visual memory for previously attended objects in natural scenes. Journal of Experimental Psychology: Human Perception and Performance 28.

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

Kamitani Y, Tong F (2006) Decoding seen and attended motion directions from activity in the human visual cortex. Current Biology 16:1096-1102.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of Neuroscience 17:4302-4311.

Kawato M (2008) Brain-controlled robots. In: IEEE International Conference on Robotics and Automation. Pasadena, CA.

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352-355.

Kiani R, Esteky H, Mirpour K, Tanaka K (2007) Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J Neurophysiol 97:4296-4309.

King-Smith PE, Grigsby SS, Vingrys AJ, Benes SC, Supowit A (1994) Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. Vision Research 34:885-912.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proceedings of the National Academy of Sciences of the USA 103:3863-3868.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12:535-540.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60:1126-1141.

Li FF, VanRullen R, Koch C, Perona P (2002) Rapid natural scene categorization in the near absence of attention. Proceedings of the National Academy of Sciences of the United States of America 99:9596-9601.

Maguire EA (2001) The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. Scandinavian Journal of Psychology 42:225-238.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proceedings of the National Academy of Sciences of the United States of America 92:8135-8139.

Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X (2004) Learning to Decode Cognitive States from Brain Images. Machine Learning 57:145-175.

Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA (2008) Predicting human brain activity associated with the meanings of nouns. Science 320:1191-1195.

Miyawaki Y, Uchida H, Yamashita O, Sato MA, Morito Y, Tanabe HC, Sadato N, Kamitani Y (2008) Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron 60:915-929.

O'Craven KM, Kanwisher N (2000) Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. Journal of Cognitive Neuroscience 12:1013-1023.

O'Toole AJ, Jiang F, Abdi H, Haxby JV (2005) Partially distributed representations of objects and faces in ventral temporal cortex. Journal of Cognitive Neuroscience 17:580-590.

O'Toole AJ, Jiang F, Abdi H, Penard N, Dunlop JP, Parent MA (2007) Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. Journal of Cognitive Neuroscience 19:1735-1752.

Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision 42:145-175.

Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-specific cortical activity precedes retrieval during memory search. Science 310:1963-1966.

Potter MC, Levy EI (1969) Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology 81:10-15.

Raizada RD, Tsao FM, Liu HM, Kuhl PK (2009) Quantifying the Adequacy of Neural Representations for a Cross-Language Phonetic Discrimination Task: Prediction of Individual Differences. Cereb Cortex.

Ratcliff R (1985) Theoretical interpretations of the speed and accuracy of positive and negative responses. Psychol Rev 92:212-225.

Tarr MJ, Gauthier I (2000) FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. Nature Neuroscience 3:764-769.

Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. Nature 381:520-522.

Torralbo A, Chai B, Caddigan E, Walther DB, Beck DM, Fei-Fei L (2009) Categorization of good and bad examples of natural scene categories. In: Annual Meeting of the Vision Sciences Society. Naples, FL.

Torralbo A, Walther DB, Chai B, Caddigan E, Fei-Fei L, Beck DM (under review) Decoding good and bad examples of natural scene categories.

Tversky B, Hemenway K (1983) Categories of Environmental Scenes. Cognitive Psychology 15:121-149.

VanRullen R, Thorpe SJ (2001) Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. Perception 30:655-668.

von Helmholtz H (1925/1909) Physiological Optics (vol. 3). Rochester, NY: Optical Society of America.

Walther DB, Fei-Fei L (2007) Task-set switching with natural scenes: measuring the cost of deploying top-down attention. Journal of Vision 7:9, 1-12.

Walther DB, Caddigan E, Fei-Fei L, Beck DM (2009) Natural scene categories revealed in distributed patterns of activity in the human brain Journal of Neuroscience 29:10573-10581.

Williams MA, Dang S, Kanwisher NG (2007) Only some spatial patterns of fMRI response are read out in task performance. Nature Neuroscience 10:685-686.

Wundt W (1874) Grundzüge der physiologischen Psychologie. Leipzig: Wilhelm Engelmann.