Kernels of Learning: Tracking the emergence of visual recognition through multivariate approaches

by

Scott Gorlin

B.A., Boston University (2005)

Submitted to the Department of Brain and Cognitive Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computational and Systems Neuroscience

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2011

Author .....
Department of Brain and Cognitive Sciences
August 5th, 2011

Certified by .....
Pawan Sinha
Professor
Thesis Supervisor

Accepted by .....
Earl K. Miller
Picower Professor of Neuroscience
Director, BCS Graduate Program

Kernels of Learning: Tracking the emergence of visual recognition through multivariate approaches

by

Scott Gorlin

Submitted to the Department of Brain and Cognitive Sciences
on August 5th, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Computational and Systems Neuroscience

Abstract

The visual system is a dynamic entity whose response properties depend on context and experience. In this thesis, I examine how the brain changes as we learn to see – what changes occur during the onset of recognition, in the mature visual system on the one hand, and in a developmentally nascent one, on the other?

Working with normal adults, I focus on the processes that underlie the interpretation of images as meaningful entities. This interpretation is greatly facilitated by prior information about a stimulus. What are the neural sites that exhibit experience dependent changes? Using multivariate decoding techniques, I find pervasive evidence of such changes throughout the visual system. Critically, cortical regions previously implicated in such learning are not the same loci as sites of increased information.

Examining the temporal mechanisms of recognition, I identify the perceptual state transitions corresponding to the onset of meaning in an observed image. Furthermore, decoding techniques reveal the flow of information during this 'eureka moment.' I find feedback processing when a degraded image is first meaningfully interpreted, and then a rapid transition into feed-forward processing for more coherent images.

Complementing the studies with mature subjects, my work with developmentally nascent observers explores the genesis of visual interpretation. What neural changes accompany the earliest stages of visual learning? I show that children treated for congenital blindness exhibit significant cortical re-organization after sight onset, in contrast to the classical notion of a critical period for visual plasticity. The specific kind of reorganization suggests that visual experience enhances information coding efficiency in visual cortex. Additionally, I present evidence of rapid development of functionally specialized cortical regions.

Overall, the thesis presents two complementary perspectives on the genesis of visual meaning. The results help advance our understanding of how short-term experience, as well as developmental history, shapes our interpretation of the complex visual world.

Thesis Supervisor: Pawan Sinha
Title: Professor

Contents

[5]

List of Figures

List of Tables

Acknowledgments

First and foremost, I must express my gratitude to my fiancée Angela for her never-ending support as I completed this work.

Additionally, this work could not have been polished without the battery of ideas, questions, and challenges raised by all those I have interacted with over the past several years. This specifically includes the members of the Sur and Sinha labs, who through many presentations and discussions have pinpointed innumerous ways for me to better my work. Additionally, there are many members of the BCS community, and beyond MIT, whose input and feedback have critically shaped this work. I am proud to be a part of such a helpful community.

Finally, I would like to thank the army that runs our department for their ceaseless toil ensuring a smooth graduate curriculum. I could not have asked for a better experience in graduate school.

Part I

Introduction

Chapter 1

Introduction

1.1 Introduction

Much of the research effort in visual neuroscience has been devoted to uncovering what the component features of vision are – what stimuli are encoded in which regions of the brain? Which components feed into the next layer to represent a more complex object? Such were the questions addressed by Hubel and Wiesel, in the late 1950's, when they discovered that cortical neurons respond primarily to edges of light [Hubel and Wiesel, 1959] – and 'lines' became accepted as the component feature for most of vision. In fact, models of the human visual system based solely on the hierarchical organization of edge detectors, feeding into basic neural networks, have made excellent progress in solving complex visual problems [Serre et al., 2007].

However, to treat the visual system as a static collection of feature detectors ignores several fundamental questions related to the dynamics and development of the system. What I mean by 'dynamics' is this: every visual image we encounter undergoes a process of perceptual organization which enables us to interpret it as not just a collection of pixels, or image features, but rather as a set of meaningful objects. This typically happens so fast that we are not even aware that there is a dynamic progression in our perception from meaninglessness to meaningfulness. Nevertheless, this is a necessary and fundamental aspect of our visual analysis. What are the neural correlates of such dynamic perceptual state transitions? On a slightly longer time-scale, our visual experience with an image at one moment helps us more easily and accurately interpret it at a later occasion even when the image is degraded in some way, say via occlusion or a reduction in the signal to noise ratio. What are the neural loci involved in such perceptual learning? Finally, on a developmental time-scale, how does the ability to interpret images meaningfully emerge in the first place? What is the timecourse of such development and how susceptible is to perturbations or delays?

These questions frame the visual system not as a collection of feature detectors, but as a dynamic entity that is influenced by two factors: 1) the history of visual input into the system over short and long timescales, and 2) the context in which the system operates. Research into these two factors has become more and more important over the past several

years, especially as artificial systems reach the limit of what a static, feed-forward system can accomplish.

It is in this vein that I focus my thesis: what changes in the brain as we learn to see? What happens during the moment of recognition, and what steps lead up to these changes? Specifically, I will ask three questions related to the dynamic properties of vision.

1.2 Question 1: How is recognition shaped by prior information?

At what stage of visual processing does bottom-up information combine with top-down expectations to yield the eventual percept? This question lies at the heart of a mechanistic understanding of feed-forward/feed-back interaction, as it is implemented in the brain and as it might be instantiated by computational vision systems. Furthermore, this question is of central significance not only for vision, but for all sensory processing since the combination of current and prior data is ubiquitous as a processing principle.

A compelling demonstration of the role of prior experience is obtained with images so degraded that they are initially perceived as devoid of anything meaningful, as seen in Figure 1-1. However, after being shown their coherent versions, observers are readily able to parse the previously uninterpretable image. The well-known Dalmatian dog picture [Gregory, 1970] – a black-and-white thresholded photograph, or Mooney image [Mooney, 1957] – is a classic example of this phenomenon.

Figure 1-1: Images which are hard to recognize at first viewing, but become easily interpretable after priming or a 'visual eureka' moment.

In this set of experiments, we use these Mooney images to explore the role of top-down

influences in visual perception. Specifically, we show that prior knowledge of the coherent image facilitates processing by increasing information throughout the visual system, including low- and high-level visual cortical areas. Interestingly, we also show that prior information is carried over complex, and not simple, visual features. Critically, these results were not found in regions of the brain previously reported as involved in priming, indicating that an information-theoretic approach reveals intricacies in the data unseen through traditional analyses.

1.3 Question 2: What changes in the brain when an object is recognized?

What happens when you recognize an image for the first time? Such a question is difficult to answer because the visual system must solve two problems at once – faithfully representing a stimulus, and interpreting, or recognizing, the results in a meaningful way. Of particular interest is the visual 'eureka' moment, where viewers perceive an image for an extended period of time before suddenly recognizing it. We seek to understand the difference between these two processes using special images that are not immediately recognizable, which allow us to tease apart the effects of low-level image representations and the process of visual recognition.

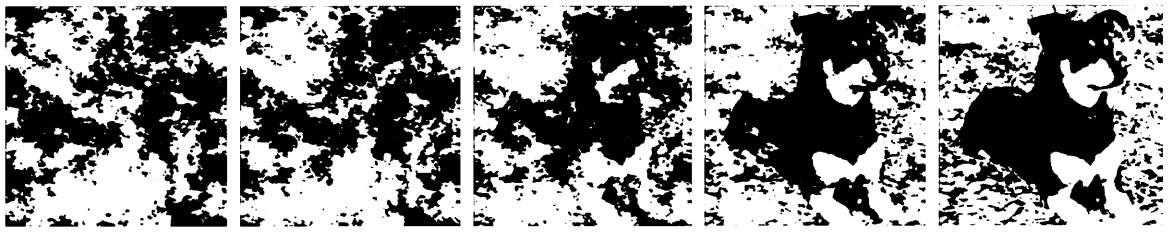


Figure 1-2: An example image series inducing a visual eureka in a controlled manner.

Figure 1-2 shows a series of images that gradually induce this eureka moment, in a controlled manner, by presenting sequentially more information to the viewer on a frame-by-frame basis. Using these sequences, we can compare neural activity immediately before and after a perceptual recognition event, allowing us find neural markers distinguishing

between visual representation and true perception.

This goal is motivated by both basic and applied considerations.

From the basic science perspective, the identification of such a marker would enable novel studies of the development of recognition (tracking how the marker changes as a function of age or visual experience) and also allow an investigation of the process by which the brain organizes complex visual arrays into meaningful percepts. Indeed, the results we present here have already begun to provide some insights into not just the markers but also the mechanisms of perceptual organization.

From the applied perspective, having a neural marker for the onset of recognition will be invaluable for assessing high-level vision in patients with certain kinds of neurological disorders. Conditions like autism, visual agnosia and Alzheimer's disease are associated with impaired recognition abilities [Behrmann et al., 2006a, Behrmann et al., 2006b, Farah, 1990, Hof and Bouras, 1991], but a quantification of these impairments has been hard to come by because we currently lack any reliable biomarkers of recognition.

In addition to presenting evidence of such markers, we were able to utilize multivariate decoding techniques to trace the directional flow of information in the brain. We find that the eureka moment induces a feedback cascade of information flowing from frontal to occipital cortex, implicating frontal cortex in a 'hypothesize-and-test' role of visual recognition. However, as the image becomes more coherent, this effect rapidly disappears and information is processed in a feed-forward manner.

1.4 Question 3: What changes in the brain as we learn to recognize the world?

Critical to our understanding of how the visual system works is the knowledge of how it develops. Humans acquire significant visual abilities within the first few months of life; however, the neural changes accompanying these early stages of development have been hard to elucidate due to the operational difficulties of conducting neuroimaging studies in very young children. By the time the children are old enough to be cooperative subjects in

such studies, they have already passed major developmental milestones.

The ideal research subject to study visual learning, therefore, is not an infant - but a developmentally mature individual who has been deprived of normal visual input. Such subjects are rare because of the difficulty of restoring sight to blind individuals; where the cure is readily available, it is usually provided at birth. However, in developing countries the advances of modern medicine are not as accessible – India, for example, has a relatively large population of children who are born with cataracts due to prenatal infections or nutritional deficiencies. Modern eye clinics are easily able to restore vision to these children, but for various cultural and economic reasons, many do not receive treatment until well into their teens. In providing sight to such individuals, through an outreach program called Project Prakash, we are presented with an unusual scientific opportunity to study their visual development from the moment sight is restored.

It cannot be taken for granted that an individual who has been blind from birth for several years will benefit in any functionally significant manner from an optical correction of his or her eyes. The classic literature on critical periods paints a pessimistic picture of the prospects of recovery after extended visual deprivation. In this field of research, a typical experiment consists of suturing shut one eye of a kitten or mouse before natural eye opening, and examining the acuity or neural response to light of the deprived eye after an extended period of time. The results of such deprivation are by no means trivial. While viewing the world through the deprived eye, animals will behave as if blind; recordings in primary visual cortex reveal a vast reduction in cells driven by the deprived eye, cell growth in the lateral geniculate nucleus is greatly reduced, and the response properties of the few cortical neurons that are driven by the deprived eye are grossly abnormal [Wiesel and Hubel, 1965b].

Binocular deprivation is a more complicated story. Similar to the monocularly deprived cats, when viewing the world for the first time through a deprived eye, cats will behave as if blind; this is accompanied by a reduction in cell body size in the lateral geniculate body and an increase in abnormally firing or non-responsive cells in visual cortex [Wiesel and Hubel, 1965a]. However, there remain a small number of cells that retain normal firing properties. Additionally, recovery of visual acuity from binocular depri-

vation has been observed over a period of time proportional to the term of deprivation [Wiesel and Hubel, 1965b, Timney et al., 1978, Smith et al., 1980, Zernicki, 1991].

Data on human visual recovery after extended blindness have so far been very limited. A handful of cases paint a similarly bleak picture, where early blindness prevents the afflicted from recovering any significant amount of functional sight [Fine et al., 2003, Gregory and Wallace, 1963, Sacks, 1995, Valvo, 1971]. In one particularly well studied case, subject MM developed normal low spatial frequency vision, but was insensitive to frequencies above 1 cycle per degree [Fine et al., 2003]. He performed well in several measures of visual perception, including the perception of simple shapes and colors; more advanced skills, however, like the segmentation of texture, 3D perception, subjective contours, and perspective cues, never developed even after years of visual experience. Interestingly, a more careful reading of the same literature provides clues that rehabilitation can occur; MM developed significant spatial localization and visual discrimination skills [Kurson, 2008], and another early-blind patient eventually learned to recognize his colleagues and cross the road based on visual skills alone [Gregory and Wallace, 1963]. The sparsity of data on visual recovery, along with the uncertainty about the extent and time-course of recovery, motivate a principled examination of how visual skills develop with the onset of sight, and whether or not human vision is restricted by a critical development period.

With the launching of Project Prakash, work from our laboratory has documented the significant gains in visual proficiency that congenitally blind individuals exhibit as a function of time after sight-restorative surgery [Bouvrie and Sinha, 2007, Held et al., 2011, Ostrovsky et al., 2006, Ostrovsky et al., 2009]. These results motivate an examination of changes in brain organization after sight onset. In this regard, it is especially interesting to investigate the changes brought about within cortical areas that in the normal brain are visually driven. Would the influx of patterned visual information after the onset of sight result in organizational changes within these areas?

Using functional neuroimaging, we probe how the brain changes as these children learn to see. We find vast evidence for cortical reorganization in several different manifests. First, we measure the local connectivity of visual cortex and find that peripheral visual cortex

becomes progressively decorrelated with itself over time, indicating an increase in efficient coding mechanisms in visual cortex. Second, we examine how specific functional modules develop in the ventral visual stream, and show that face-selective cortex grows rapidly as the subjects learn to recognize faces. Additionally, we show that precursors to face- and object-selective cortex exist within days after surgery, suggesting that these regions require a very modest amount of visual experience before they are established in the brain.

1.5 The emergence of recognition

These three questions are fundamental to understanding the genesis of visual meaning in our percepts. This thesis seeks to address all of these questions through a combination of studies with unusual subjects (those who gained sight after prolonged blindness) and sophisticated computational analytics, specifically multivariate analyses of neural data. We begin, in the next chapter, with a description of the multivariate approaches we have used in our work and which are likely to play an increasingly important role in neuroscience more broadly.

Chapter 2

Multivariate analysis

Abstract

Understanding how the mind works is a daunting task – and one that is severely limited by the difficulty of observing activity in the brain without modifying it. Traditionally, experiments in neuroscience have focused on acute techniques, like extracellular recording and optical imaging, which require extensive surgical preparation in animal subjects. Noninvasive techniques, like fMRI and EEG recordings, make experiments in humans possible, at the loss of spatial and temporal resolution achievable in techniques when the safety of the subject is not the primary concern. Recently, a new wave of analytical techniques focused on multivariate analysis has given scientists a better understanding of how groups of neurons behave together at a larger scale, and is bridging the gap between experiments in humans and animals – bringing a rich literature of systems neuroscience to bear on the behavioral complexities of the human mind.

2.1 Introduction

The ultimate goal of a cognitive neuroscientist is to understand how a circuit of neurons behave as to produce a given thought or behavior. This is a deceptively daunting task, and one which begins with the difficulty of measuring neural activity in an alert, behaving animal. As physicists have Heisenberg, so too are neuroscientists faced with an indelible amount of uncertainty in their recordings – the more accurate the recording technique, typically, the more invasive the procedure and the less likely the neuron is to behave in a normal, and relevant, way.

A prudent example of this is a classic extracellular recording preparation. In order to measure the activity of single neurons, a scientist might lower sharp electrodes into the cortex of, say, a macaque monkey to measure the electrical potential of neurons as they release action potentials in response to input. A single electrode causes minimal damage in and of itself, but can only measure the activity of, typically, one to four neurons. So, to get a better understanding of how a group of neurons behave, more electrodes are often used – between 4 and 100 cortical penetrations are common depending on the experiment. Needless to say, there is a point of diminishing returns where lowering more electrodes causes more damage to the system and yields less valuable information to the researcher.

Other preparations have similar limitations. Optical and two-photon imaging, for instance, allow for measuring the activity of an entire flat patch of cortex several millimeters in diameter, and the resolution easily allows viewing how groups of neurons interact with their neighbors. However, these techniques require a craniotomy in order to view the cortical surface in the first place – and it is impractical to enlarge the craniotomy beyond a certain point.

With all the information gleaned from these experimental techniques, however, its usefulness is still critically limited – we cannot assume that the human brain works identically as in these animal models, and we cannot perform the same experiments in humans, as surgically invasive procedures are neither legal nor ethical to perform in normal people. Exceptions are made when subjects require invasive procedures for other purposes, but even in these scenarios basic research is not in the interest of the patient. All told, there is

very little to be said about how individual neurons behave in humans, or how neurons in any species behave in complex tasks an animal will not perform reliably in a lab.

Instead, noninvasive techniques have become popular with researchers interested in higher level processes such as complex perception and executive behavior. One technique, known as Functional Magnetic Resonance Imaging (fMRI), has become increasingly popular over the past twenty years due to its relatively high spatial and temporal resolution. This technique was first used to localize visual cortex in humans by simply imaging where there was a change in metabolic oxygen consumption when subjects viewed a flashing checkerboard [Belliveau et al., 1991], and has since become a mainstay experimental technique for cognitive neuroscientists.

2.2 General Linear Model

The basic analytical technique most commonly used in fMRI experiments is to generate a Statistical Parametric Map (SPM) which illustrates which voxels, or regions of interest, are affected by an experimental technique and to what degree. Most commonly, this comes in the form of a General Linear Model (GLM) which is used to describe how the voxel intensity (as a proxy for neural activation levels) is linearly dependent on an experimental condition. For example, if the experimenter is interested in localizing primary visual cortex, the role of the GLM would be to illustrate how significantly each voxel is modulated by a visual input when it is turned on and off.

Mathematically, this is accomplished by a linear regression of the voxel intensities against the experimental and other confounding variables. The GLM takes the matrix form:

$$\begin{aligned} Y &= X\beta + \epsilon \\ \epsilon &\sim N(0, \sigma I) \end{aligned}$$

for data measurements of a single voxel in the rows of Y and experimental design vari-

ables in the columns of X [Friston, 2007]. That is, each column of X will have a separate value in each row describing whether a variable was present (and to what degree) during that sample acquisition. Here, β is a vector of coefficients for that voxel, describing how each variable in X affects the measurements. The residual errors (the difference between the data and how much is predicted by X and β) are captured in the rows of ϵ. When the system is solved for the best estimate of $\hat{\beta}$ such that the errors are normally distributed about 0 with the smallest possible variance (that is, the Ordinary Least Squares method), the result and significance of $\hat{\beta}$ under the null-hypothesis of a T-distribution can be used to infer whether the system was significantly modulated by the variables in X.

More complicated analyses can be formed by contrasts across the β vector: for instance, the result of viewing face stimuli is best compared not to rest, but to viewing a similarly complex image of a house. In this case, the contrast vector "Face – House" would require the subtraction $\beta_F - \beta_H$ and statistical inference can be performed on this result.

The solutions, given any contrast vector c, n measurements, and p dimensions in the model, are given by:

$$\begin{aligned}
\hat{\beta} &= (X^T X)^{-1} X^T Y \\
\hat{\beta} &\sim N(\beta, \sigma^2 (X^T X)^{-1}) \\
\hat{\sigma}^2 &= \frac{\epsilon^T \epsilon}{n - p} \\
t &= \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}
\end{aligned}$$

The critical point to note about the GLM is that it performs its analysis in a "mass-univariate" manner. That is, each voxel or region is treated independently of the rest when the statistics are calculated. The result is that if there is no effect visible in a single voxel, or single region, as a whole, any subtleties will be averaged out and lost. Because the spatial and temporal scale of fMRI is drastically different from most techniques used in other species, it is exceptionally difficult to compare studies in humans and animals – what

happens in large scales in the brain is not necessarily what is seen when looking at single cells.

2.3 Feature Domains

A useful example of data gleaned from high-resolution studies in animals is the concept of a feature domain. One of the first discoveries in visual cortex was that neurons respond preferably to lines of a specific orientation [Hubel and Wiesel, 1959], and that the preferred orientation is similar for neurons close to each other in cortex [Hubel and Wiesel, 1962]. Furthermore, the actual topography of the preferred orientation has been shown to be highly detailed across the surface of cortex [Hubel and Wiesel, 1963]; 'feature maps' of orientations, and other salient features in areas beyond primary visual cortex, illustrate an intricate layout of representations across many cortical domains.

This type of cortical organization has been difficult to observe in humans. The primary difficulty is that the scale of feature domains are on the order of 0.5mm in diameter; typically, fMRI experiments are carried out with a resolution of 3mm-per-side voxels. Given the GLM analysis approach which would form a contrast vector over averaged examples of responses, voxels containing mixed amounts of different features would not show up as statistically selective for either – even though they contain very selective domains at a higher resolution (Figure 2-1). Even if some voxels by chance contained more of one feature than another, neighboring voxels likely would not share this preference and the subtle per-voxel results would be averaged out in any region of interest.

Researchers have used two alternate approaches to show feature selectivity in fMRI. One approach hypothesizes that highly active neurons fatigue and later show reduced activity to a repeated stimulus, a fact that is well-established in many electrical and imaging studies [Grill-Spector and Malach, 2001, Ganel et al., 2006]. Reduced fMRI activation to repeated stimuli thus suggests the presence of fatigued neurons in a given region of interest, and is evidence of fine-grain selectivity when the same region is not fatigued to similar stimuli. The major drawback of this technique, however, is that it typically requires an event-related experiment to measure a brief reduction in the hemodynamic timecourse of

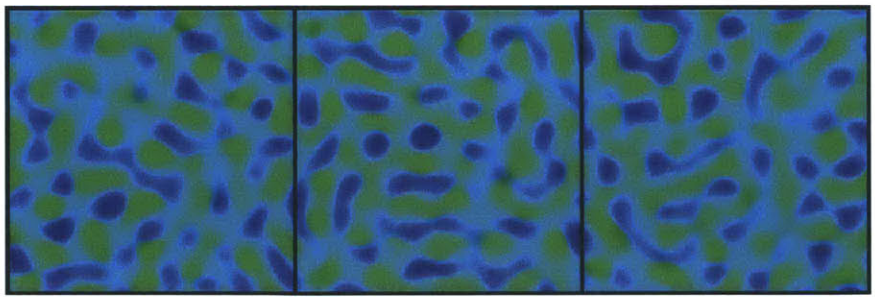


Figure 2-1: Hypothetical feature domains for 'green' and 'blue' features, sized at approximately 0.5mm diameter. The black borders represent the 3mm voxel size for a typical fMRI experiment. Under an experiment comparing 'green' vs 'blue' responses, no statistical difference would be expected of these voxels given that they contain highly-mixed sets of both domains.

the fMRI signal, averaged over many latencies, and typically for a large region as a whole. Experiments often can only report whether a large region of brain is or is not selective for a single feature. The second alternative is to simply increase the resolution of the fMRI scan. In fact, recent studies have successfully shown orientation [Yacoub et al., 2008] and ocular-dominance [Cheng et al., 2001] feature maps in human primary visual cortex through high-resolution fMRI, but these come at the cost of smaller field-of-view in the images, lower contrast-to-noise ratios, and increased time constraints on the experiment due to longer image acquisitions. Additionally, to date these techniques have only been successful in high-strength magnetic fields, such as 4T and 7T – much stronger and more expensive than the standard 1.5T and 3T MRI machines available to most researchers.

2.4 Multivariate Techniques

Over the last ten years, a series of studies have pushed beyond the univariate GLM framework in fMRI to examine how voxels work together. The goal of this approach is to extract data that is too subtle, or does not exist, when looking at each voxel independently.

The first example of this technique examined how large regions of ventral cortex, where most higher-level vision is processed, correlate voxel-wise when a subject views different images. When subjects viewed images within object categories, like 'chairs', voxels in ventral cortex correlated more between trails than when subjects viewed a different cate-

gory, like 'shoes' [Haxby et al., 2001]. Previously, researchers believed that ventral cortex was composed of a series of 'modules' dedicated to processing fairly specific visual categories; the most well-studies of these are the Fusiform Facial Area (FFA) and Parahippocampal Place Area (PPA) which respond preferably to the stimuli types they are named for [Kanwisher et al., 1997, Epstein and Kanwisher, 1998]. Because these areas are fairly large, on the order of a cubic centimeter, their discovery was well-suited for fMRI resolution and the univariate GLM analysis. However, the Haxby experiment showed that there are numerous correlated patterns of activity across these same regions that encode many different visual stimuli, even though none of these categories elicited a large-region, univariate response detectable via GLM.

2.4.1 Support Vector Machines

The distributed encoding experiments paved the way for more formal techniques that attempted to 'learn' the patterns of activity in the brain. The first of these used a Support Vector Machine (SVM), a type of linear classifier, to predict which stimulus caused a given pattern of activity across the cortex.

In its purest form, a linear classifier is any function which takes the form:

$$
\begin{aligned}
w, x \in & \mathbb{R}^D \\
f_{w,b}(x) = & \sum_{d=1}^{D} w^d x^d + b \\
L(x) = & \text{sign}(f_{w,b}(x))
\end{aligned}
$$

In other words, the classifier computes the dot product between a data point x and a second point w representing a hyperplane in D dimensional space (hence, projecting the data onto the hyperplane), additionally given a bias term b. The sign of this classification function can then be used to designate the output into one of two classes, or labels, here +1 and -1.

It is critically important to understand how the hyperplane is picked. The first important concept is that of a training set – the collection of data from which the classification function is learned. For simplicity, imagine two neurons: one which responds preferably to 'green' stimulation, and a second neuron which prefers 'blue' stimulation (Figure 2-2). A trial set containing stimulations of both types should cluster nicely by eye in 2-dimensional space. The goal of a SVM is to form the decision boundary that maximally separates these clusters. That is:

$$
\begin{aligned}
x, w \in & \mathbb{R}^D \\
y \in & \{+1, -1\} \\
\min_{w,b} C & \sum_{n=1}^{N} (1 - y_n f_{w,b}(x_n))_+ + \|f_{w,b}\|^2
\end{aligned}
$$

In other words, the SVM must minimize a 'cost' function based on the penalty of mislabeled points ($(1 - y_n f_{w,b}(x_n))_+$, where $()_+$ indicates a hinge function: $(x)_+ = \max(0, x)$), the amount the optimization is penalized for mislabeling (C), and a regularization function ($\|f_{w,b}\|^2$) based on the hyperplane itself which prefers wider separating boundaries [Cortes and Vapnik, 1995, Vapnik, 1998, Vapnik, 1999]. The result is clearly understandable in 2-dimensions, where the SVM selects a line that maximally separates the clusters of data, with the ability to encompass points within the $\{+1, -1\}$ borders depending on the penalty C.

Given more dimensions, it is straightforward to imagine a SVM using more neurons, or perhaps voxels encompassing thousands of neurons acting together, to build this hyperplane in D-dimensional space. In fact, the coefficient of the hyperplane in any given dimension will determine how much that dimension, and the neurons it is proxy for, have to say about the classification problem in question. In this manner, the analysis becomes truly multivariate as it depends on the coordinated activity of all voxels in a region of interest, and not just one at a time.

It is worthwhile to note that there are several other families of algorithms that can be

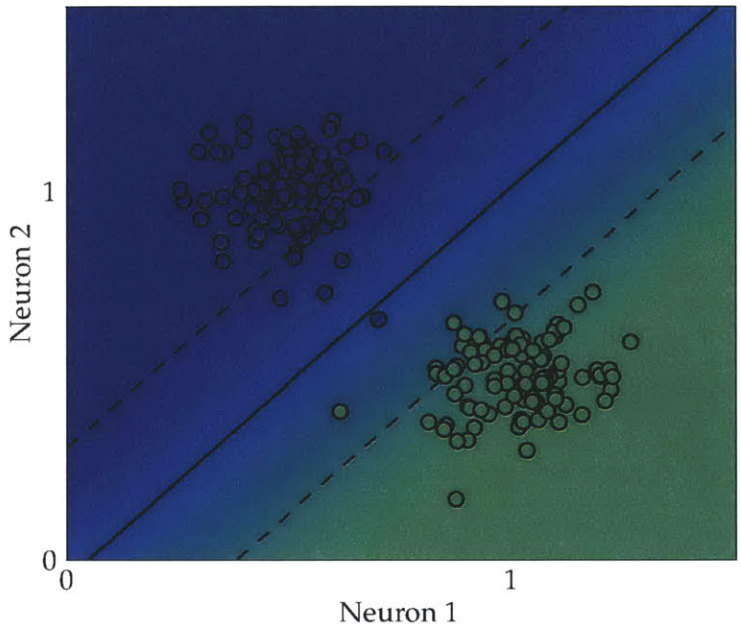


Figure 2-2: Example SVM solution. Neurons 1 and 2 are partially selective for green and blue features, respectively. All points represent stimulation trials and show how much each neuron responds to a green or blue stimulus. A SVM finds a hyperplane that maximally separates the clusters of data. Here, the solid line represents the decision boundary, and the dotted lines are the +1 and -1 boundaries that solve the problem within a given amount of tolerance. Now, any new trial points that fall in a green or blue region of the graph can be predicted to be the result of a green or blue stimuli, depending on which side of the decision boundary they fall on. Note that there is one mislabeled point in this example.

used as linear (or nonlinear) classifiers; any algorithm that solves the above minimization for a hyperplane, regardless of the assumptions used to find the 'optimal' decision boundary, is sufficient; as is any algorithm that can map a feature vector to a discrete class label. Examples of other classifiers used in neuroimaging studies include Gaussian Naive Bayes [Mitchell et al., 2004], Linear Discriminant Analysis [Carlson et al., 2003], and Sparse Multinomial Logistic Regression [Yamashita et al., 2008], to name a few. Support Vector Machines have become common in part due to multiple free software implementations and the intuitive mechanics behind them, but they are by no means the only solution to multivariate analyses.

2.4.2 Sparse Multinomial Logistic Regression

A popular contender to SVM's in neuroimaging is Sparse Multinomial Logistic Regression (SMLR), because the assumptions this algorithm makes are particularly suited to neuroimaging. Because neuroimaging data tend to have many more voxels than they do samples, SVM's are susceptible to over-fitting (assigning weight to voxels where the pattern is strictly due to noise). To overcome this problem, fMRI data is often subjected to feature-selection – running an analysis on a restricted subset of voxels, or region of interest.

SMLR [Yamashita et al., 2008] presents an alternative solution by automatically selecting relevant features in a Bayesian framework, and additionally is natively expandable into multiclass problems. It expresses the probability of sample data belonging to a given class, using the linear hyperplane w as before, by:

$$
\begin{aligned}
y \in & -1, 1 \\
x \in & \mathbb{R}^D \\
P(y_1, y_2, \ldots, y_N | x_1, x_2, \ldots x_N) = & \prod_{n=1}^{N} P(y_n|x_n) = \prod p_n^{y_n} (1 - p_n)^{1-y_n} \\
p_n \equiv & P(y_n = 1|x_n) = \frac{1}{1 + exp(-f_w(x_n))}
\end{aligned}
$$

Similar to the SVM, SMLR must solve a cost function, via maximizing the likelihood function:

$$
\max_w L(w) = \prod_{n=1}^{N} p_n^{y_n} (1 - p_n)^{1-y_n}
$$

In true Bayesian fashion, the likelihood of a given hyperplane w can be weighted by a prior distribution:

$$L^*(w) \propto L(w)P(w|\alpha) = L(w)\prod_{d=1}^{D} N(0, \alpha^{-1})|_{w_d}$$

The prior states that coefficients near 0 are more likely for every dimension of the hyperplane w (the prior is the product, over all feature dimensions, of a 0-mean Gaussian evaluated at the coefficients of that dimension). Given a hyperparameter α which controls the width of the Gaussian priors, thereby allowing more or less weight in multiple dimensions, the prior will effectively create a sparse hyperplane and limit the exposure of the classifier to noisy features. If the prior were flat ($\alpha = 0$), SMLR would assign weight to every voxel, and behave similarly to a SVM.

Despite the advantages on SMLR on paper, its popularity has not spread much beyond the original authors. In part this is likely due to the advanced mechanics of the optimization function, and the fact that it tends to perform similarly, and much slower computationally, in actual implementations.

2.4.3 Relation between GLM and Classifiers

It should be noted that once $\hat{\beta}$ is calculated from the GLM, there is nothing to prevent a multivariate analysis across voxels in that framework. For instance, a contrast across the beta vectors is equivalent to a hyperplane (that is, the data can be projected onto it) that can be used in a linear classifier – so the results can be used as a multivariate classifier as well! The only formal difference between this method and a SVM is how the w vector coefficients are calculated: via ordinary least squares in the GLM, and via the hinge loss in the SVM. These two methods will yield different classification functions, but qualitatively the results will be the same.

The major difference, then, between the GLM and other classification frameworks are the standard practices that researchers employ. First, in order to get the best mean activation map through ordinary least squares, the GLM averages over all trials to yield a single beta vector. Classifiers typically keep all trial data separate, using the variance structure across all the trials per feature in order to form an optimal hyperplane. Additionally, when looking

for regions of activation in a GLM analysis, researchers seek beta vectors with one or more clusters of high coefficients that are easily distinguishable by eye. To reduce the variance of the beta vector (thereby increasing the significance), spatial and temporal smoothing of the fMRI signal is standard. Thus, if there is a strong univariate response of a group of voxels as selective for one feature versus another, using a GLM framework (including using the betas for classification) or using a SVM will work fine; if, however, the signal is weakly encoded across a wide region of voxels, in a truly multivariate manner, a standard GLM analysis would be insufficient to detect feature encoding.

2.5 Feature domains in fMRI

After it was shown that patterns of activity were indeed predictable in fMRI through machine learning algorithms [Cox and Savoy, 2003, Carlson et al., 2003], the next logical steps were to push the boundaries of what was learnable and to try to figure out exactly what signal was underlying this process.

Using SVM's, several researchers have been able to predict what orientation a subject is viewing at any given moment [Kamitani and Tong, 2005, Haynes and Rees, 2005]. As discussed above, orientation domains are typically not visible through conventional fMRI, so to be able to predict the orientation of a stimulus was a huge leap forward in terms of analytical techniques. At the time, the running hypothesis was that orientation domains were distributed randomly and non-uniformly throughout the regular voxel grid, in effect 'aliasing' the feature domains in the recorded image [Boynton, 2005]. Although there has been considerable debate whether this is true, or that, instead, the signal is created by larger inhomogeneities in feature-selectivities and distributions [de Beeck, 2009, Kamitani and Sawahata, 2009], the fact that researchers could now base analyses on previously undetectable feature sensitivities remains a major breakthrough in functional imaging.

The ability to decode brain images rapidly spread throughout the research community. A number of publications showed the variety of what was learnable, including perceived motion directions [Kamitani and Tong, 2006], semantic categories of sentences [Mitchell et al., 2004],

what a subject was attending to [Haynes and Rees, 2005, Kamitani and Tong, 2005, Kamitani and Tong, 200 the contents of a subject's working memory [Polyn et al., 2005], and many other tasks. Not all of these studies, however, truly pushed the boundaries of machine learning or neuroimaging; as mentioned above, classification can be a trivial extension of the GLM, and indeed, some researchers found that they were able to decode mental states using conventional techniques [O'Craven and Kanwisher, 2000]. For a more detailed review of the history of the emergence of multivariate analysis in MRI, see [Norman et al., 2006, Haynes and Rees, 2006].

2.6 Reconstruction

The fact that neural patterns in fMRI can be used to predict which of a set of stimuli a subject is experiencing has revolutionized the field, but classification is only one side of the coin. The brain itself is not always subject to 'classification' problems, where vast patterns across thousands of neurons must fit into a small handful of previously learned categories. Rather, every pattern is subtly different because the scene it encodes might not fit into a rigid categorical structure. In fact, the brain must represent – and to some degree, be able to reconstruct – the entirety of its relevant input.

Attempts to reconstruct visual input to the brain have used a different mathematical structure from the classification problems above, but the concepts can be understood in similar forms. Basically, a reconstruction technique will first map out the 'receptive field' of a given region of interest through a combination of GLM and other univariate techniques. So, for example, a group of voxels in primary visual cortex might represent a few degrees of the visual field. Because the visual system is understood to decompose input into various basis functions like oriented line segments and edges, a classification paradigm can be used to predict how much of each basis function a region of cortex is representing – and then sum together all of the basis functions to wholly reconstruct the visual image.

In this manner, researchers have completely reconstructed visual scenes using black and white image patches from V1 [Miyawaki et al., 2008] and even full-color images based on the entire visual cortex[Naselaris et al., 2009]. Both of these studies first decomposed im-

ages into a linear combination of local image bases ϕ_n [Olshausen and Field, 1996], similar to the orientation- and feature-selective bases thought to underly processing in visual cortex. Once each image was decomposed into a sum of multiple bases, the viewed image was reconstructed as

$$
\begin{aligned}
I &= \sum_{n=1}^{N} C_n \phi_n \\
\hat{I}(x) &= \sum_{n=1}^{N} \lambda_n \hat{C}_n(x) \phi_n \\
\hat{C}_n(x) &= P_{SMLR}(\phi_n | x)
\end{aligned}
$$

Or, the reconstructed image $\hat{I}$ of the fMRI data x is the sum over all bases n of the *a-priori* weighted basis function $\lambda_n \phi_n$ times the amount of contrast $\hat{C}(x)$ predicted from the fMRI data. The amount of contrast was predicted using the SMLR algorithm, outlined above, to determine the likelihood of the MRI data being in response to a particular image base.

2.7 Future Directions

Machine learning techniques have transformed the field of neuroimaging in no small way. Research has spilled beyond cognitive neuroscientists and into the medical imaging community as well, and now SVM's have been used to detect psychiatric illness such as schizophrenia [Sun et al., 2009], quantify and track tumor volume [Shubhangi and Hiremath, 2009], and detecting liver fibrosis [Freiman et al., 2010], to name a few applications.

The heyday of classification-for-classification's sake, however, has begun to wane. Instead, the variability of the results from classification are of interest themselves; for instance, how the population code is affected by attention [Serences et al., 2009], or whether recognition modulates the amount of correlation amongst voxels encoding visual objects [Hsieh et al., 2010]. That is, classification has truly become a tool by which to measure

processes such as attention in the brain.

Part II

How is recognition shaped by prior information?

Chapter 3

Imaging prior information in the brain

Abstract

In making sense of the visual world, the brain's processing is driven by two factors: the physical information provided by the eyes ('bottom-up' data) and the expectancies driven by past experience ('top-down' influences). We use degraded stimuli to tease apart the effects of bottom-up and top-down processes, since they are easier to recognize with prior knowledge of the fully coherent images. Using machine learning algorithms, we quantify the amount of information brain regions contain about stimuli as the subject learns the coherent images. Our results show that several distinct regions, including high-level visual areas and retinotopic cortex, contain more information about degraded stimuli with prior knowledge. Critically, these regions are separate from those which exhibit classical priming, indicating that top-down influences are more than feature-based attention. Together, our results show how the neural processing of complex imagery is rapidly influenced by fleeting experiences.

3.1 Introduction

At what stage of visual processing does bottom-up information combine with top-down expectations to yield the eventual percept? This question lies at the heart of a mechanistic understanding of feed-forward/feed-back interactions, as they are implemented in the brain and as they might be instantiated by computational visual systems. Furthermore, this question is of central significance not only for vision, but for all sensory processing since the combination of current and prior data is ubiquitous as a processing principle.

A compelling demonstration of the role of prior experience is obtained with images so degraded that they are initially perceived as devoid of anything meaningful. However, after being shown their coherent versions, observers are readily able to parse the previously uninterpretable image. The well-known Dalmatian dog picture [Gregory, 1970] – a black-and-white thresholded photograph, or Mooney image [Mooney, 1957] – is a classic example of this phenomenon. Other examples of top-down knowledge facilitating sensory processing include phonemic restoration [Kashino, 2006] and the interaction between depth perception and object recognition [Bulthoff et al., 1998].

The approach of comparing neural responses to degraded images before and after exposure to the fully coherent image has been used by several research groups to identify the correlates of top-down processing. For instance, PET scans of brain activity elicited by Mooney images before and after disambiguation show that inferior temporal regions of cortex, as well as medial and lateral parietal regions, exhibit greater activity in response to recognized images [Dolan et al., 1997]. Progressive revealing paradigms, where an image gradually increases in coherence, elicit increased and accelerated fMRI activation in several regions including the fusiform gyrus and peri-striate cortex, when subjects have prior experience with the images [James et al., 2000]. Additionally, EEG correlates show that distorted or schematic line-drawings elicit face-specific N170 ERP components, which are believed to reflect activity in the fusiform face area, only after a subject learns to interpret them as a face [Bentin et al., 2002, Bentin and Golland, 2002]. All of these results indicate that prior experience rapidly modifies brain activity and the final response to a given stimulus. Not surprisingly, much of this modification has been localized in the higher areas

of the visual pathway, which are more driven by the object percept that an image induces rather than its low level characteristics.

While these results certainly demonstrate the existence of top-down influences, we believe that they may be just the tip of the proverbial iceberg at best and misleading at worst. A critical issue left unaddressed in previous studies is the distinction between traditional priming, and an actual increase of image-specific information encoded in a given brain region. We assert that the true criterion for declaring the presence of learning in a region is that it alters or amplifies stimulus-specific information as encoded in the pattern of activity across the region, rather than merely changing the level of activation as a whole. The critical question is thus: is a downstream neuron (or a machine learning algorithm) better able to decode which stimulus was seen after exposure to the coherent image, compared to before? If there is no enhancement of such decodability, then that region's activity is likely driven entirely by the physical characteristics of the stimulus, and no learning has occurred – regardless of whether the total activity level in that region has changed.

It is thus critical to explore top-down influences from an information-theoretic perspective. To this end, we adopted a multivariate analytical technique to measure the amount of decodable information in various cortical regions. This involves examining a fine-grained, but potentially weak, pattern of activity across all voxels in a region of cortex, rather than a univariate approach which detects which voxels, independently, are significantly more active in one condition versus another.

An algorithm such as a Support Vector Machine (SVM) [Vapnik, 1998] is a linear classifier that detects the ensemble activation patterns that distinguish two or more categories of data; formally, it projects data onto a high-dimensional hyperplane (where each dimension corresponds to a single voxel's activity in a given region of interest) that maximally separates two or more data clusters in that space. The ability of the SVM to accurately predict which cluster new data belong to corresponds to the amount of information those voxels convey about the cluster labels. That is, SVMs are accurate when voxel activity in their underlying region of interest encodes something about the problem at hand.

Multivariate techniques have been used previously to show how different object categories are represented across visual cortex [Carlson et al., 2003, Haxby et al., 2001] and

can be used to predict which stimuli from a fixed set a subject is viewing at any given time [Cox and Savoy, 2003, Haynes and Rees, 2005, Kamitani and Tong, 2005]. There is some debate in the literature as to the origin of the cortical activation patterns underlying such decodability, and whether they represent the aliasing of feature-selective neurons inside a grid of voxels [Boynton, 2005, Haynes and Rees, 2005, Kamitani and Tong, 2005], or rather a larger, non-uniform distribution of feature selectivity [de Beeck, 2009, Kamitani and Sawahata, 2009, Swisher et al., 2010]. In both interpretations, however, the ability to decode cortical patterns reflects how much information the underlying neurons contain about a stimulus in question. In the present context, multivariate analysis gives us a useful tool to investigate whether the amount of information contained in various cortical regions is modulated by a subject's prior knowledge of the stimuli.

3.2 Results

The overall methodology we followed is illustrated in Figure 3-1. Black and white 'Mooney' images were shown to participants during functional magnetic resonance imaging. We examined whether priming the subjects with coherent, full-color versions of the Mooney images before each run led to better recognition, and then, using machine learning techniques, investigated whether neural activity patterns in different regions of the brain showed corresponding increases in information content. This allowed us to identify the neural correlates of prior information processing in the cortex.

Our experiments yielded three key results. First, in surprising contrast to inferences from past studies, classically-primed voxels were not "information rich." In other words, decoding accuracy did not improve post-priming across these voxels. Second, a whole-brain analysis revealed that increased recognition accuracy, due to priming, was correlated with an increase in object-related information across many regions of the visual system. Finally, the facilitation of behavioral and neural information was dependent on complex image features (or objects) rather than simple features like oriented lines.

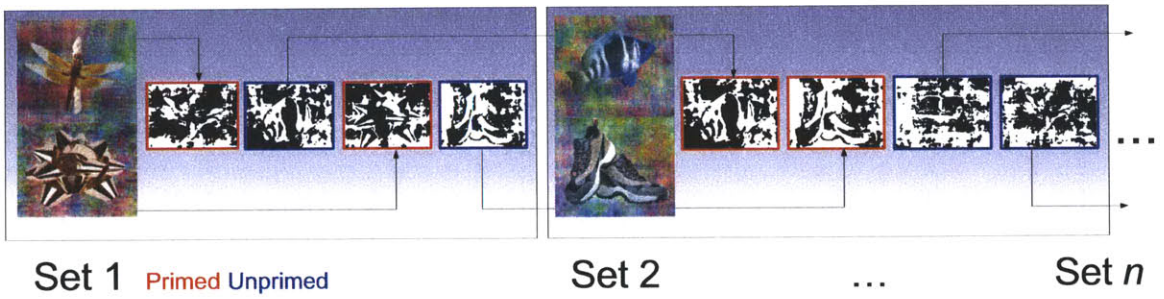


Figure 3-1: Overall methodology. Before each fMRI scan run, two images – one containing a natural object, and one containing an artificial object – were shown to the subject. During the run, the subject was shown Mooney images of four different objects – the two primed images, and two novel, unprimed images – in a block paradigm, and was asked to identify the contents of each image as either natural or artificial. For the next set of images, undegraded versions of the two previously unprimed images were used to prime the subject, and two additional novel images were introduced. In this manner, every image (except the first primed and last unprimed images) was shown to the subject in both primed and unprimed conditions.

3.2.1 Priming increases recognition accuracy

During each scan, subjects were asked in a 2-AFC task whether each Mooney image they saw was of a 'natural' or 'artificial' object, as a way of measuring recognition in the scanner. All subjects consistently recognized the primed Mooney images with higher accuracy than the unprimed images in each run (Figure 3-2; 1-tailed McNemar test, $p < 10^{-10}$).

If any subject did not show a significant increase in recognition for one of the images (which would occur if the subject recognized the Mooney image in the unprimed condition, or if they did not recognize it in the primed condition), that image was excluded from further analysis for that subject. From a total of 38 images shown collectively to subjects, 10 were excluded in this manner, leaving 28 images included in the analysis. For these remaining images, subjects could not recognize the images at greater than chance accuracy in the unprimed condition. Additionally, they did not learn to recognize the images even after multiple presentations before priming.

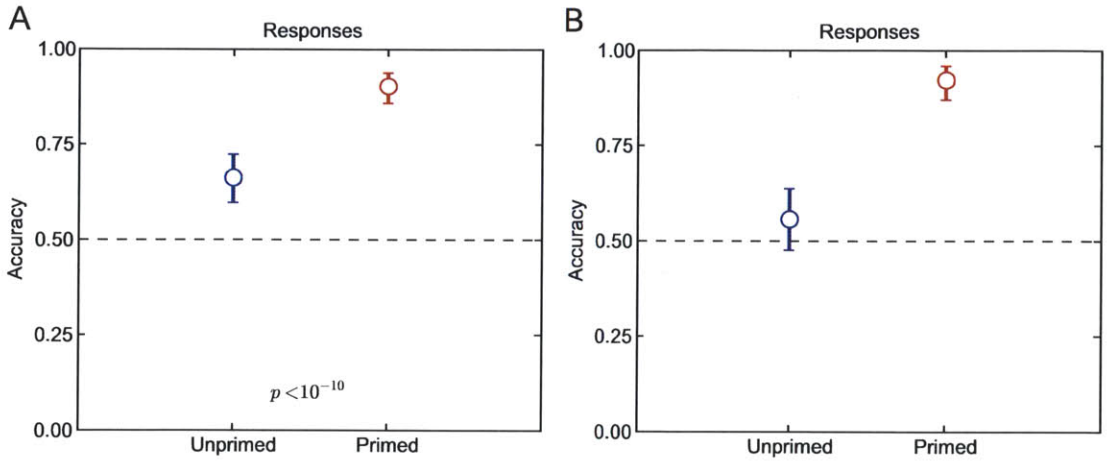


Figure 3-2: Behavioral performance recognizing the primed and unprimed Mooney images during the fMRI scans. A) Across all images, subjects performed significantly better for primed images in a 2-AFC task indicating whether the image was 'natural' or 'artificial.' Shown is the expected value and 95% confidence intervals for the proportion correct over all images and subjects; p value calculated from a 1-tailed paired McNemar test corrected for multiple comparisons over all subjects and image sets. B) When images were selected to show a per-image priming effect (resulting in discarding 10 out of 38 images), subjects were at chance at recognizing the unprimed images.

3.2.2 Classically 'primed' voxels carry no information

Past studies [Dolan et al., 1997, James et al., 2000] have localized cortical areas where activity is elevated post-priming using Mooney images. Can a simple increase in activation levels in the primed versus unprimed conditions explain an increase in recognition ability? To address this issue, we used a General Linear Model to find voxels which showed overall priming. Figure 3-3 shows the group average primed response over all subjects; the location of these voxels is consistent with past reports [Dolan et al., 1997]. Individual subject maps, which were extensively larger but more varied in location, were used to define the ROI's on a per-subject basis. A liberal statistical threshold ($p < 0.01$ and a clustering method to correct for multiple comparisons to $p < 0.05$) was chosen to ensure that the GLM-selected ROI was not so small as to negatively bias classification. The resulting average size across subjects was just over 2000 voxels across both hemispheres (for comparison, functionally-defined V1 ranged from 100-200 voxels per hemisphere, and fusiform gyrus was about 600 voxels per hemisphere). No significant activation was found for the control contrast,

Natural > Artificial.

To examine whether priming affected the amount of information encoded in this brain region, the scan data were subjected to a multivariate analysis where SVMs attempted to decode which Mooney image (from all sets) a subject was viewing during any given trial.

Remarkably, restricting our multivariate classification analysis to the 'primed' ROI resulted in chance performance of the SVMs, as seen in Figure 3-3. Knocking out the GLM-selected voxels (and using the remainder of cortex as an ROI), however, resulted in a significant increase in decoding accuracy for both primed and unprimed conditions (McNemar test, $p < 0.01$ for both conditions). Therefore, the voxels that were classically 'primed' could not be the same set of voxels that carry information about the stimuli. This is an important distinction, as it suggests that the sites of priming-related activity enhancements identified by previous studies are not necessarily the loci of top-down influences. An agnostic whole-brain analysis reveals the loci of increased information, as described in the next section.

Figure 3-3: Difference between SVM and GLM results. A) Group average response showing 'primed' voxels, which show greater activation for primed versus unprimed images in a GLM analysis, here illustrated in red and yellow. B) When these voxels are used as a region of interest for the SVM analysis, no increase in information decoding is seen – and in fact, SVM performance drops to chance. Note that although we show the group average response here in A, each ROI was determined from a subject's individual data.

3.2.3 Priming increases decoding accuracy throughout the visual system

Our results show overall higher SVM classification accuracy for primed versus unprimed images in multiple regions across the visual system, mirroring our behavioral findings. Figure 3-4 shows a map of the regions that exhibit changes in the information of their neural activity patterns, and Figure 3-5 shows examples of the actual decoding accuracy of several notable regions. All regions exhibiting significant differences in decoding are reported in Table 3.1. We find it especially interesting to note that experience dependent information increases in early visual areas such as pericalcarine cortex, in addition to higher level areas such as the fusiform gyrus and lateral occipital cortex. No regions exhibited lower decoding accuracy for the primed images.

Cortical region	LH	RH
Anatomically defined		
Cuneus	$n.s.$	$p = 0.046$
Fusiform gyrus	$n.s.$	$p = 0.028$
Inferior parietal	$p = 0.034$	$p = 0.047$
Inferior temporal	$p = 0.043$	$n.s.$
Lateral occipital	$n.s.$	$p = 0.051$
Pericalcarine*	$p = 0.010$	$p = 0.069$
Precentral gyrus	$p = 0.074$	$n.s.$
Rostral middle frontal	$p = 0.054$	$n.s.$
Superior temporal gyrus	$n.s.$	$p = 0.029$
Functionally defined		
V1	$n.s.$	$n.s.$
V2	$n.s.$	$p = 0.062$
V3	$p = 0.068$	$p = 0.066$
V4v	$n.s.$	$n.s.$

Table 3.1: Significance of increased SVM-decodability in various cortical regions. Results from a 1-tailed McNemar test are presented as p values for both left- and right-hemispheres for every cortical region exhibiting a significant, or borderline significant, increase in information. *Pericalcarine cortex results exhibit stronger significance, though borderline in the right hemisphere, when the analysis was restricted to highly visually-responsive voxels.

Several regions showed a significant effect in one hemisphere, but not the other. The most dramatic of these include lateral-occipital cortex, superior temporal gyrus, and fusiform

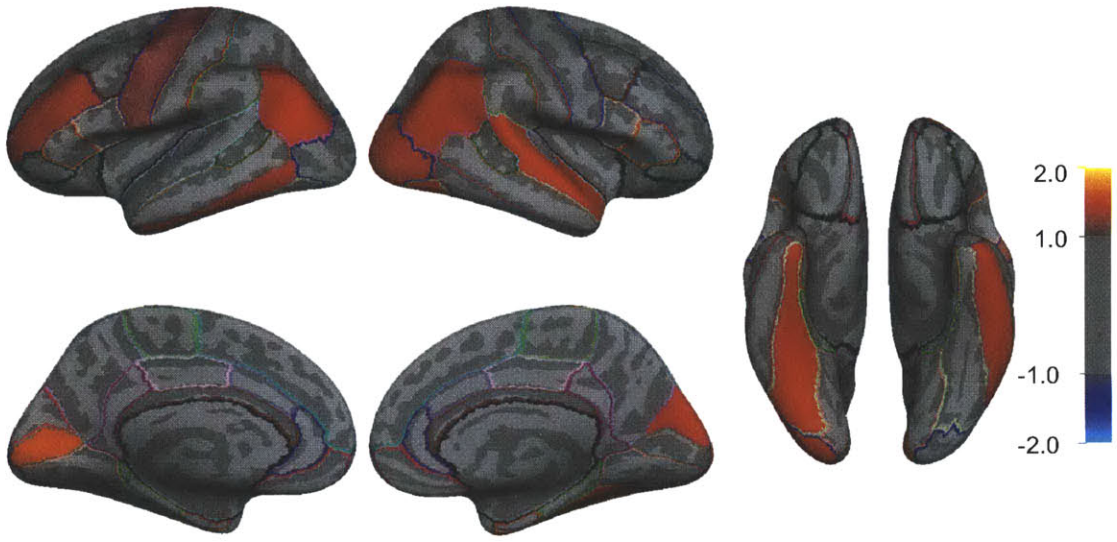


Figure 3-4: Object priming increases the information found in many regions of cortex. Using atlas-based anatomical regions of interest for the SVM analysis, several notable regions including peri-calcarine cortex, inferior parietal cortex, lateral occipital cortex, and fusiform gyrus show significantly increased decoding accuracy for primed versus unprimed images. No areas showed higher decoding accuracy for unprimed images. The magnitude of the color scale indicates the level of significance in log power (e.g., +2 and -2 both indicate $p < 10^{-2}$), whereas the sign of the scale indicates whether primed images yielded higher accuracy (positive scale) or vice-versa. Colored lines indicate the borders of anatomically-based ROI's.

gyrus, which showed significant effects in the right hemisphere but not the left, and precentral gyrus and rostral-middle-frontal cortex, which showed borderline significant effects in the left hemisphere but not the right.

Overall, our data reveal that prior experience leads to an increase in information content across multiple cortical areas. Interestingly, these results are distinct from earlier findings of the neural correlates of priming.

3.2.4 Simple stimuli do not elicit the same results

Since each region of the brain encodes stimuli with different levels of complexity, we investigated whether the complexity of the stimuli carrying prior information would affect the regions of the brain that show an increase in information. To this end, we conducted a similar experiment where subjects were shown degraded 'simple' stimuli containing a field

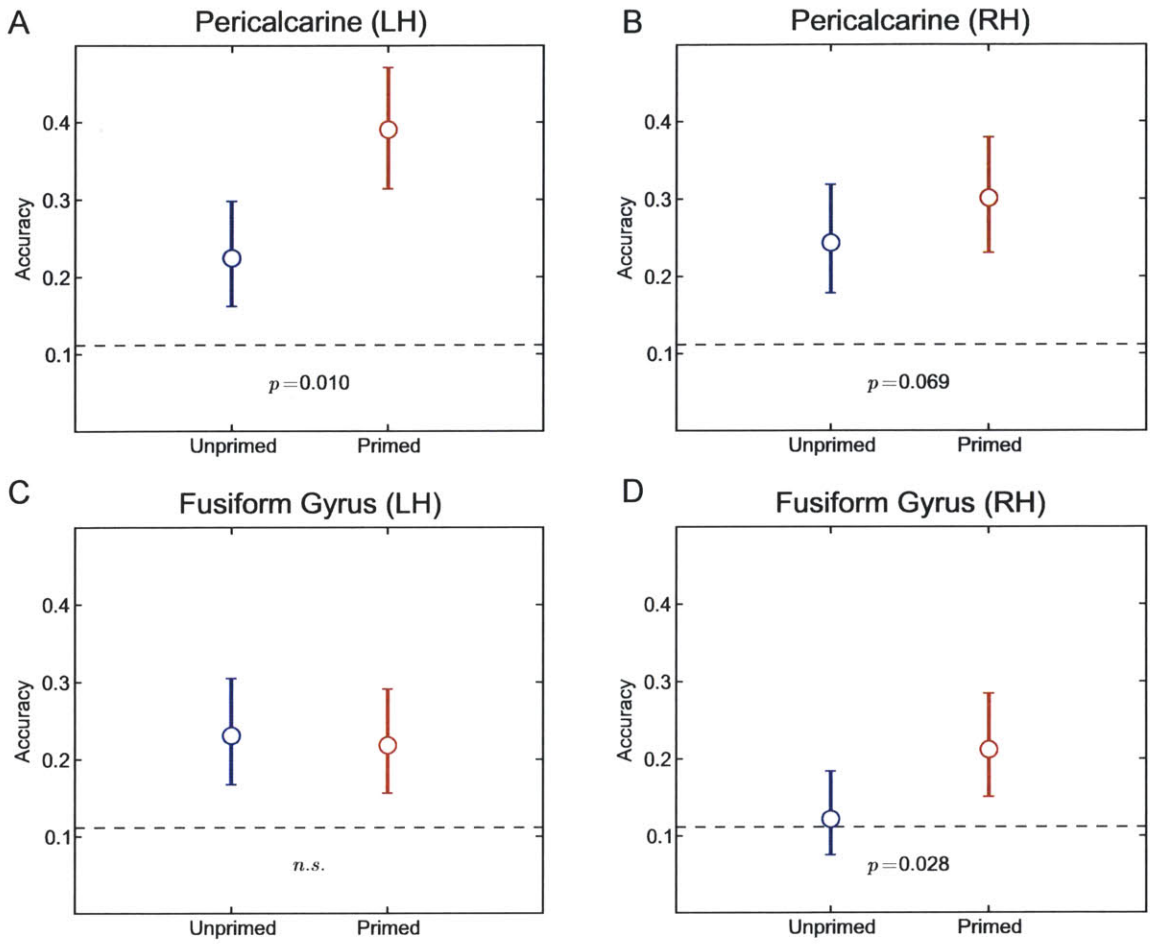


Figure 3-5: Example results of the decoding accuracy in a few notable cortical regions. A, B) Pericalcarine (masked to visually-responsive voxels) regions showed relatively high decoding accuracy, and were significantly higher in the primed condition. C, D) Classification accuracy was lower overall in the fusiform gyrus, but significantly elevated in the primed condition in the right hemisphere only. Dashed lines indicate chance performance levels over the entire group of subjects.

of oriented line segments (see Figure 3-6), instead of complex Mooney images.

We found no evidence of an increase in recognition for these simple stimuli (Figure 3-6). Correspondingly, no increase in SVM decoding ability was seen in this experiment, in any ROI tested. Additionally, no GLM-based effects were found indicating increased activation in the primed condition.

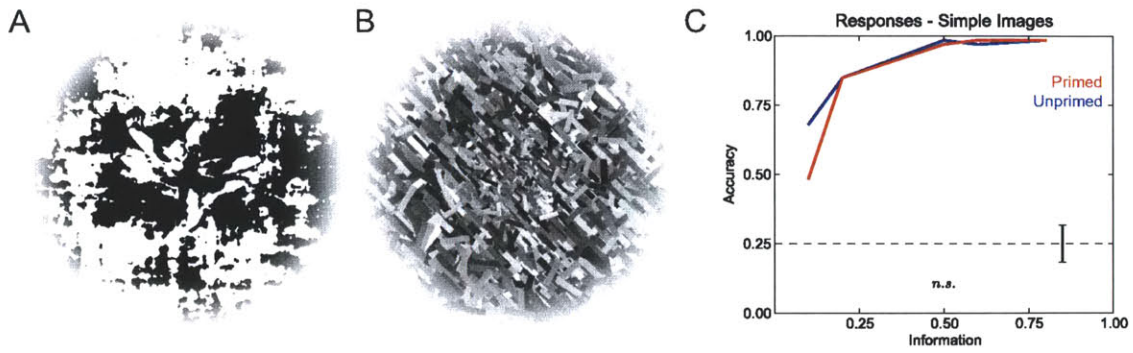


Figure 3-6: Difference between the 'simple' and 'complex' experiments. A) Mooney image of a 'complex' object (a dragonfly). Subjects showed an increased ability to recognize Mooney images when primed with the coherent images, as shown in Figure 3-2. B) 'Simple' stimuli created by compositing many oriented line segments with an average orientation, here 45 degrees. C) Unlike complex stimuli, subjects' ability to recognize simple stimuli was not increased when primed with the true orientations. Information is the mean percent alignment of all oriented stimuli to the true orientation. Blue lines indicate performance prior to viewing the coherent stimuli, while red lines indicate performance after exposure. Black dotted line is chance, and the black scale bars indicate the average 95% confidence interval of the true percent correct under a Bernoulli distribution.

3.3 Discussion

The increased decoding accuracy of primed versus unprimed images suggests that there is an increase in image-specific information in regions of the brain exhibiting this form of priming. In general, the regions in which we find an increased decoding ability occur across multiple levels of the ventral visual processing system. This includes higher-level areas such as fusiform gyrus and lateral occipital cortex, and lower levels including pericalcarine cortex, and in particular V2 and V3. Surprisingly, no effects were found in V1 and V4v. A recent study [Hsieh et al., 2010] has also reported an increase in information in the foveal confluence and object-selective cortex (area LOC).

More critically, we show that these effects are distinct from traditional priming, as we were unable to find any evidence of information encoding in the GLM-selected regions, as discussed below. Additionally, we show that prior information is carried over complex, and not simple, image features.

It is important to note the differences between the results reported here and those obtained in the GLM analysis. Prior studies have shown how the overall activity in regions of

cortex in response to Mooney images is modulated before and after exposure to the unambiguous versions [Dolan et al., 1997, James et al., 2000], and our localization in the GLM analysis generally agrees with their results. However, this overall modulation of activity must be interpreted separately from the SVM analysis results. For instance, consider the answer to the question, "Which regions of the brain are involved in priming?" To answer this, a GLM analysis is appropriate because it contrasts the overall activity in the brain in response to the Mooney images before and after priming with the coherent images. The result is a subtraction of the average activity, over all images, in one condition versus the other. In contrast, our experiment asks the question, "Which regions of the brain encode more image-specific information, for primed and unprimed images?" The SVM analysis thus looks at the differential pattern of activity between all images in a single condition. The overall level of activation in no way changes the pattern of differences between any two images. The SVM analysis thus reports on how much information is contained in a region pertaining to the stimuli, and not the overall activity level.

Therefore, it is best to consider the GLM and SVM methods as answering two distinct questions. The fact that decoding ability falls to chance using only the GLM-selected voxels highlights this distinction. In simple terms, the GLM analysis selects voxels that are highly activated in response to all images as an average; the SVM analysis performs well when voxels respond more to one stimulus than another. As a result, prior studies showing 'priming' effects via this GLM-type of analysis, in reality, show parts of the brain which do not necessarily encode the stimuli at all. Rather, they show increased activity in regions more likely involved in attention and arousal rather than increased information encoding of a specific stimulus. Adaptation-based results are an exception to this, where a lower level of activity is presumably the result of fatigued feature-specific neurons in a given area, but the point remains that pattern information changes represent a qualitatively different result than that obtained by the standard GLM analysis.

It is interesting to note the discrepancy between the simple and complex stimuli. Although we show increased pattern decoding for primed complex objects, we did not find similar results for simple stimuli – either in subjects' behavioral responses, via pattern decoding, or via GLM. It is unlikely that no priming occurs for simple stimuli, as even V1 is

modulated (albeit weakly) by attention [Kastner et al., 1998, Pinsk et al., 2004], but it does appear that prior information, in this context at least, does not help the brain to decode the orientation of line segments. This raises the interesting question of whether this lack of priming is true only for the most basic of visual features, or if progressively more complex stimuli show a graded amount of information increase with prior knowledge. On a similar note, it may be the case that wholly different types of stimuli might enhance information in different regions entirely – for instance, prior information about motion stimuli may increase the amount of information in the dorsal visual stream, rather than the ventral areas shown in this study.

These results also suggest that information priming is more than simply feature-based attention. It is well established that neurons with multiple stimuli in their receptive fields will respond as if only one stimulus is present when that stimulus is attended to [Reynolds et al., 1999], and that attending specific features of an image will sharpen or increase the cortical response to those features [Haenny et al., 1988, McAdams and Maunsell, 2000, Motter, 1994b, Motter, 1994a, Saenz et al., 2002]. In this interpretation, the increase in information could be due to increased attention to visual features that comprise the true object hidden in the degraded image, and decreased attention to the 'noise' features. Several factors, however, distinguish our results from classical feature-based attention experiments. First, attention was never explicitly directed to any image features – any specific feature-based attention was entirely endogenous as a result of increased prior knowledge about the stimuli. Second, the experiment we conducted with 'simple' stimuli was more similar to a classical feature-based attention experiment, as the subjects were cued with oriented stimuli before each scan. Our inability to detect a greater amount of information in this case stands in contrast to the results that visual cortex, even on the single-unit level in V4, exhibits sharpened responses to oriented stimuli when cued [Haenny et al., 1988, Motter, 1994b, Motter, 1994a]. Our results indicate that information priming is object-based, whereas feature-based attention is known to work on lower levels. Finally, feature-based attention calls for an increased response in many neurons through multiplicative gain mechanisms [Martinez-Trujillo and Treue, 2004] when comparing attended- vs. non-attended stimuli. The fact that the most significantly activated voxels, as selected by our GLM, did not ex-

hibit an increase in information suggests that there is more than a simple multiplicative mechanism at work. Therefore, feature-based attention cannot fully explain our results and differentiates this study from other Mooney-based experiments [Dolan et al., 1997, James et al., 2000].

In some areas, we have shown a significant effect in one hemisphere but not the other. For some regions, like the cuneus and pericalcarine areas, their physical proximity to one another could explain why the cuneus is primed in the right hemisphere and pericalcarine in the left, if the true foci of information priming in individual subjects happened to fall near the atlas boundaries. Higher-level regions, however, provide a more interesting story. For instance, the FFA is known to respond to faces differently in the left and right hemispheres, where the right hemisphere tends to respond more to a subject's overall percept and to whole faces rather than parts [Rossion et al., 2000]. Additionally, the two hemispheres of LOC are known to be modulated differentially by task demands: the right hemisphere activity is increased more than the left by visually matching objects [Large et al., 2007]. These types of lateralization effects agree nicely with the fact that we show information priming in the right fusiform gyrus and lateral-occipital regions, but not the left, as the right hemisphere seems to be more involved in processing the specific visual form information of objects [Koivisto and Revonsuo, 2003, Marsolek, 1995, Marsolek, 1999].

It is interesting to speculate whether this information-based priming simply amplifies response patterns across cortex, or if there is a substantial alteration of the neural code. Training the classifiers on the primed data and testing them on the unprimed data (rather than cross-validating within each condition) results in above-chance classification rates throughout visual cortex, indicating that there is significant conservation of the pattern between both conditions. However, our data cannot rule out the possibility of a true alteration of the neural code where the patterns differ; indeed, we would expect that both mechanisms – amplification and alteration – would occur. That is, we hypothesize that patterns in the neural code representing the stimulus are likely amplified, but additionally the brain might alter the neural code in areas representing missing information (for instance, using top-down information not found in the actual stimulus).

We have shown that an increase in knowledge about a stimulus translates into an in-

crease in pattern information across several regions in visual cortex. Although the time-course of the fMRI signal does not allow us to conclude these are feedback-phenomena, the fact that this increase is dependent on behavioral priming and complex images highly suggests that this may indeed be a case for top-down facilitation of processing in the visual system.

3.4 Acknowledgements

We are grateful for the generous sponsorship of the John Merck Foundation and the Jim and Marilyn Simons Foundation. Additionally, this work was supported by NIH EY07023 to MS and NEI R21EY015521-01 to PS.

3.5 Methods

3.5.1 Participants

Nine right-handed volunteers between the ages of 22-35 were recruited from the MIT community to participate in the fMRI experiment. All participants gave informed written consent. The study was approved by the MIT Committee On the Use of Humans as Experimental Subjects (COUHES). Four of these volunteers were excluded from analysis based on technical difficulties with the data or a failure to perform the behavioral task adequately, leaving 5 subjects included in the main analysis. Two of these subjects, plus two additional subjects, participated in the 'simple' experiment.

3.5.2 Visual stimuli

Various photographs of 3-D, real world objects were collected from the internet or commercial photography collections. Photos containing depictions of humans or human-like faces, lettering in English or another language, and those that had high emotional valence (such as images of snakes or spiders) were explicitly excluded from the stimulus set. Photos were first converted to gray-scale by using only the green channel as the luminance signal. Each

object was manually extracted from the background and placed on a neutral gray. To minimize the low-level differences between images, the magnitudes of all Fourier components were averaged across the entire image set, and the images were reconstructed via inverse Fourier transform based on their original phase and the averaged magnitude of each particular frequency. To additionally degrade the images, phase noise was introduced, partially 'scrambling' the images via shifting the phase angles randomly [Dakin et al., 2002]. These degraded images were then converted to Mooney-like images by thresholding at the per-image median gray level intensity. The amount of noise added during the scrambling step was titrated so that the final Mooney image was not recognizable without seeing the fully coherent version.

For the 'simple' experiment, noisy oriented stimuli were generated by overlaying hundreds of oriented line segments of random luminance on top of a gray background. The size and width of each segment was scaled exponentially to approximate the foveal magnification factor in cortex, so that segments near fixation were smallest. To degrade the stimuli, each oriented segment was rotated proportionately from the true stimulus orientation to a random orientation.

3.5.3 fMRI Experiment

Scanning was performed on a 3.0-Tesla Siemens scanner using a standard head coil at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research, MIT. A high-resolution T_1-weighted 3D-MPRAGE anatomical scan was acquired for each participant (FOV 256×256, 1 mm^3 resolution). To measure BOLD contrast, 33 slices parallel to the AC/PC line were acquired using standard T_2^*-weighted gradient-echo echoplanar imaging (TR 2000 ms, TE 30 ms, flip angle 90°, slice thickness 3 mm, in-plane resolution 3 × 3 mm). Stimuli were rear-projected onto a screen in the scanner bore.

Scan runs began and ended with 16 seconds of fixation-rest and included 8 image presentation trials in a block paradigm. Each trial consisted of 16 seconds of visual stimulation, where a given image was flickered on and off at 6 Hz to reduce habituation. Trials were followed by 2 seconds of rest and subsequently a 4 second response window, where

the subject was required to perform a 2-AFC task regarding the content of the image seen in the previous block by pressing a button. Another 10 seconds of rest followed the response phase before the next trial began. Participants were additionally instructed to report via button press the occurrence of a low-contrast flicker in the fixation cue as a means of monitoring fixation and alertness. This fixation cue was present during the image presentations and rest, but not during the response phases.

Each scan run consisted of two arbitrarily-paired sets of images. Each set contained one image of an 'artificial' object, such as a shoe, and one image of a 'natural' object such as a butterfly. Before the run began, one image set was designated as 'primed' and the undegraded versions of those images were shown to the subject. Each image was presented twice per run, for a total of eight image presentations (2 sets $\times$ 2 images $\times$ 2 trials) presented in pseudorandom order. During each response phase, the subject was required to press a button indicating whether they recognized the previously viewed image as either natural or artificial. The same image sets were repeated for a total of 3 runs. The non-primed image set was then designated as primed and a new image set introduced for the next three runs. Sessions continued in this manner and typically included 3-6 total image sets (meaning classification was between 6-12 total images, depending on the subject) for a maximum session duration of 2 hours. Therefore, every image was presented to the subject a total of 12 times – 6 times before they saw the coherent version, followed by 6 times after priming. The very first primed images and the last unprimed images, which were not shown in both conditions, were excluded from analysis.

For the 'simple' experiment, conducted in separate scan sessions, each scan run contained repeated presentations of two degraded orientation stimuli, out of four total (horizontal, vertical, and two diagonals). During each rest period, subjects were asked to indicate which out of all four orientations they saw in the previous block, thereby performing a 4-AFC task similar to the above experiment. The same two stimuli were repeated for 2 additional runs before the subjects were shown the coherent oriented stimuli. This was followed by three runs with the same stimuli, after exposure to the coherent stimuli. After 6 total runs, two new orientations were chosen and the process repeated for a maximum duration of 2 hours.

3.5.4 Analysis

Cortical reconstruction and volumetric segmentation was performed with the Freesurfer image analysis suite, which is documented and freely available for download online. The technical details of these procedures are described in prior publications [Dale et al., 1999, Fischl et al., 1999a]. Functional scans were motion corrected in AFNI [Cox, 1996] and coregistered to individual subjects' anatomical scans using a combination of manual alignment, custom scripts and SPM8. No spatial or temporal smoothing was applied. The aligned functional images were then processed in a General Linear Model (GLM) using a block-related analysis with a double-gamma HRF in FEAT (FMRI Expert Analysis Tool) Version 5.98, part of FSL (FMRIB's Software Library) [Beckmann et al., 2003]. Time-series analysis was carried out using FILM with local autocorrelation correction [Woolrich et al., 2001] and regression against motion correction parameters. Individual blocks were treated independently; that is, eight $\hat{\beta}$ images, or one per block, were output from the GLM per run. The resulting $\hat{\beta}$ images were converted to z-scored images and finally extracted per region of interest for multivariate pattern analysis as described below.

Regions of interest were defined by automatic surface parcellation [Desikan et al., 2006, Fischl et al., 1999b, Fischl et al., 2004] in Freesurfer and by retinotopic field sign boundaries for V1 through V4v determined from separate scan sessions according to previously described methods [Dougherty et al., 2003, Sereno et al., 1995]. Additionally, a 'primed' region of interest was determined using a general linear model in FEAT for the contrast Primed > Unprimed with $p < 0.01$ and a clustering method of $p < 0.05$ to correct for multiple comparisons.

MVPA

Multivariate pattern analysis (MVPA) was performed using custom Python routines based on the PyMVPA package [Hanke et al., 2009b, Hanke et al., 2009a] with the Shogun 0.9 backend [Sonnenburg et al., 2006]. Briefly, linear Support Vector Machines [Vapnik, 1998] were trained and used to predict which image, from all a given subject had seen, was presented during a given block, using 1-vs-Rest multi-class classifiers. As there were between

6 and 12 different images depending on the subject, chance classification rates were between 17% and 8%. Data were trained and classified independently using leave-one-run-out cross validation. Statistics were calculated by comparing paired classifications of each image before and after priming in a 1-tailed McNemar test, summed over all image pairs per subject; a final group McNemar statistic was formed using a correction for non-independent clusters across subjects [Durkalski et al., 2003].

Results for any region of interest were excluded if neither the primed nor unprimed images had decoding accuracy significantly higher than chance, even if the differences between the two results were statistically significant. Results from individual images were also excluded for any subject who did not show a significant improvement in 2-AFC discriminability for that particular image set, thereby restricting the SVM analysis to images which elicited a behavioral priming effect. Of a total of 38 images, 10 were excluded in this manner.

Part III

What changes in the brain when an object is recognized?

Chapter 4

Decoding the visual eureka

Abstract

The visual analysis of every image we encounter involves a process of perceptual organization, one that transforms the input from an amorphous collection of regions into meaningful entities. This 'visual eureka' phenomenon is uniquely useful as it separates the initial representation of a stimulus in the brain from the higher-order processes of true recognition, and interpretation, of the same stimulus. What neural changes accompany this perceptual state transition? To address this question, we recorded high-density EEG signals while observers viewed sequences of images wherein an object progressively emerges out of noise. For each sequence there was a point that marked the onset of recognition. Analyzing neural activity straddling this point, we find biomarkers for the onset of visual recognition that are not tied to any particular object class. Additionally, using multivariate techniques, we show how feedback is involved in the moment of recognition.

4.1 Introduction

Figure 4-1a shows three well-known images that present significant recognition challenges to first-time viewers. After scanning these images for several seconds, observers typically report a 'eureka!' moment when the seemingly meaningless collection of black and white blobs organizes itself into a coherent object [Gregory, 1970, Cavanagh, 1991, Rubin et al., 1997]. We are interested in identifying the neural correlates of this perceptual state change.

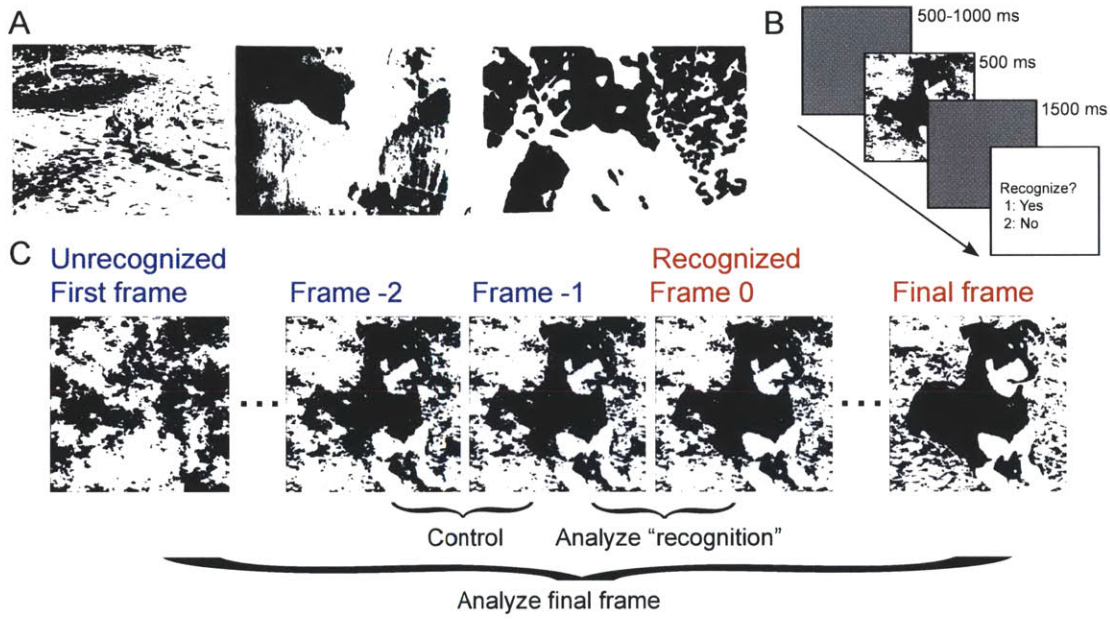


Figure 4-1: A) Images that are hard to recognize for first-time viewers, but become highly interpretable after a 'visual eureka' moment. B) Experimental setup. Individual frames of RISE sequences were presented sequentially to a subject at a rate of approximately one every 5 seconds, and after each presentation the subjects were asked to report whether or not they recognized the stimulus via button press. C) Example frames from a sequence (a dog), and segmentation of the data for analysis. Each RISE sequence was recognized in a stepwise manner (once the sequence was recognized, on a given frame, all subsequent frames were recognized as well). The effects of recognition were determined by analyzing the differences between the first recognized frame (Frame 0) and the frame immediately preceding it (Frame -1), which was not recognized. To control for any effects caused simply by an increase in information, and not by a perceptual recognition event, the two frames prior to recognition (Frames -1 and -2) were also analyzed for differences. Finally, the very first and very last frames of the entire sequence were analyzed as well.

This goal is motivated by both basic and applied considerations.

From the basic science perspective, the identification of such a marker would enable novel studies of the development of recognition (tracking how the marker changes as a function of age or visual experience) and also allow an investigation of the process by which the brain organizes complex visual arrays into meaningful percepts. Indeed, the results we present here have already begun to provide some insights into not just the markers but also the mechanisms of perceptual organization.

From the applied perspective, having a neural marker for the onset of recognition will be invaluable for assessing high-level vision in patients with certain kinds of neurological disorders. Conditions like autism, visual agnosia and Alzheimer's disease are associated with impaired recognition abilities [Farah, 1990, Hof and Bouras, 1991, Behrmann et al., 2006a, Behrmann et al., 2006b], but a quantification of these impairments has been hard to come by because we currently lack any reliable biomarkers of recognition.

Since we are interested in studying a temporally evanescent moment of recognition onset, the brain activity assessment modality has to possess high temporal resolution. We have therefore chosen to use electro-encephalography (EEG), which involves directly recording electrical potentials on the surface of the scalp.

Interestingly, there is very limited work so far on identifying the EEG correlates of the visual eureka moment. This is perhaps due to the methodological challenges involved. Most EEG analyses require a precise knowledge of event onset times in order to temporally coregister signals recorded across different trials. The visual eureka, however, is unpredictable in its time of occurrence. One individual might see the dalmatian dog in Figure 4-1a after 5 seconds of viewing, while another might take two minutes. Trying to mark the perceptual change by having subjects press a button (or take some other action) is difficult because variabilities in reaction time are too great to allow precise EEG signal registration. Additionally, the preparation for response [Vaughan et al., 1968] might contaminate any markers for recognition.

We have sought to overcome these challenges through the use of a methodology we have developed called Random Image Structure Evolution, or RISE [Sadr and Sinha, 2004]. Briefly, images can be degraded procedurally, to enable the experimenter to control how much visual information is presented to the subject at any given time. In this manner,

the onset of visual recognition can be triggered by a temporally precise increase in visual information, rather than by unpredictable endogenous processes. Critically, the method of image degradation can be chosen so as to preserve low-level image statistics effecting attributes such as luminance and contrast.

In our analyses of the EEG data recorded in response to several RISE sequences, we looked not only for neural activity changes associated with recognition onset, but also their temporal staging across different brain regions. The hypothesis that guides this analysis is that perceptual organization and recognition of potentially ambiguous images, such as those in Figure 4-1a and in RISE sequences, involves significant top-down analysis [Cavanagh, 1991, Bar et al., 2006, Craft et al., 2007, Ahissar et al., 2009]. The very fact that the time of recognition of such images is highly variable indicates that a straightforward bottom-up strategy cannot be the underlying process. Instead, the brain likely adopts an active 'hypothesize and test', with hypotheses being generated at a higher level and then being verified against bottom-up data [Hupe et al., 1998, Kersten et al., 2004]. In our analyses we seek to determine if the temporal ordering of neural activations across the brain is consistent with this notion of top-down guidance.

4.2 Results

To identify the changes that occur in the brain coincident with the moment of recognition, we adopted a modified version of a RISE sequence and recorded EEG signals from subjects as they viewed the stimuli. Briefly, images of real-world objects were scrambled in Fourier space to contain a specific amount of phase coherence, and then thresholded to contain only two luminance levels (black and white), to yield images akin to those produced by Mooney [Mooney, 1957]. The resulting images were presented to subjects sequentially so that every new image contained more information than the last, until a fully coherent two-tone image was presented; after that, a fully scrambled version of a new image was presented to the subject. See Figures 4-1b-c for an example.

4.2.1 Stepwise recognition and data segmentation

In each RISE sequence, subjects' recognition responses exhibited a step-like profile – once an image was recognized, all versions of the same image containing more information were recognized as well. We analyzed three sets of data, defined relative to the first frame in which the object was recognized; this frame was denoted as Frame 0 (see Figure 4-1c). To analyze the effects of recognition, the first recognized frame was compared to the immediately prior frame that had not been recognized; that is, we compared Frame 0 to Frame -1. To control for effects induced primarily from an increase in information and low level changes in the images, a second data set was analyzed containing the responses to the prior two image frames, Frames -1 and -2. A final dataset included the very first and very last frames of the entire RISE sequence, to compare our results with those obtained using large increases in information (and viewing coherent versus non-coherent objects). These dataset segmentations are illustrated in Figure 4-1c.

4.2.2 Waveform analysis

In order to elucidate the effects of recognition in the brain, we examined the evoked response potentials. Figure 4-2a shows the average potential response for all stimuli immediately after recognition, for a sparsely sampled set of sensors across the brain. This is contrasted with the responses to stimuli one frame earlier, which were not recognized.

As evidenced by Figure 4-2a, the polarity of change in ERP waveforms corresponding to recognized and unrecognized frames partitioned the set of sensors into two rough profiles, in frontal and occipital regions. We examined this further by averaging over these broad regions of interest for further analysis. To establish whether the responses were significantly different, successive paired t-tests (comparing identical stimuli per subject) were calculated across the entire waveform. Time points that were significant at a level of $p < 0.05$ and remained significant for greater than 10 ms (to correct for multiple comparisons) are shaded in Figure 4-2c.

Several differences are immediately noticeable in Figure 4-2c. In occipital-parietal regions, recognition of the stimulus resulted in a decreased positive inflection around 200

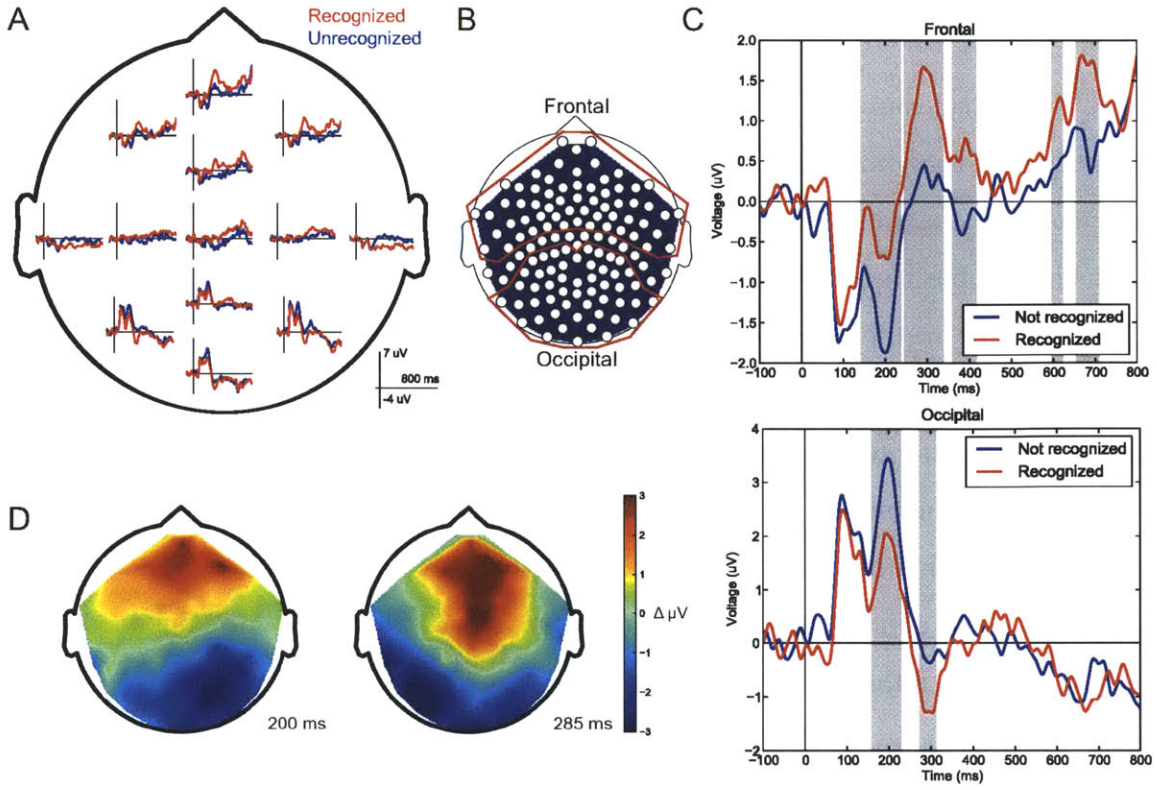


Figure 4-2: ERP analysis. A) Selected sensors showing the difference between recognized and unrecognized stimulus ERP's. B) Schematic illustrating the two main ROI's. C) ERP results averaged over these two ROI's. Shaded timepoints indicate where the two ERP's are significantly different at $p < 0.05$ for 10 ms or longer. Occipital sensors exhibited a decreased positive inflection around 200 ms post stimulus, and an increased negative inflection 300 ms post stimulus. Frontal sensors exhibited a decreased negative inflection 200 ms post stimulus, and an increased positive inflection around 300 ms post stimulus. Time 0 represents when the stimulus was presented; the stimulus duration was 500 ms. D) Full-head ERP differences between recognized and non-recognized images, shown at the two timepoints where occipital regions showed the greatest difference. The differential patterns of activation are clearly bilateral and involve both frontal and occipital regions.

ms after the stimulus onset, and an increased negative inflection 300 ms post stimulus onset. Frontal responses were more complicated, with significant differences extending from 150-400 ms post stimulus, including a decreased negative inflection at 200 ms, and an increased positive inflection at 300 ms. These differences can not be accounted for by a simple increase in image saliency or contrast, as the ERP responses for the last unrecognized image and the image immediately prior to that one (Frames -1 and -2) did not show any significant differences.

In order to clearly visualize the evoked potentials, the differential response (Recognized – Unrecognized ERP's) was plotted over the entire scalp. Timepoints 200 ms and 285 ms, which roughly correspond to the peak differences in the occipital regions, are shown in Figure 4-2d. These results clearly indicate that the ERP differences were substantially bilateral, and involve both frontal and occipital regions simultaneously.

4.2.3 Recognition can be decoded

To extend the results of the traditional waveform analysis, we subjected the data to a multivariate decoding experiment to see if we could predict whether the EEG signal at any given timepoint was caused by a 'recognized' image or not. Rather than looking for a difference in the overall average potential response, this technique is sensitive to more subtle differences that are potentially distributed across multiple sensors. The classifier is trained to detect these differences in a subset of the data and uses this information to accurately predict which of two categories the remaining data fall into. Greater-than-chance accuracy of the classifier would thus indicate that there is a learnable signal present in the data, whether or not it is visible by eye or statistically significant via univariate measures.

Briefly, linear Sparse Multinomial Logistic Regression (SMLR) [Yamashita et al., 2008] classifiers were trained and used to predict data based on a single timepoint, using N-fold cross-validation across each image set. This allowed us to plot classification accuracy at each time point. Figure 4-3a shows the overall accuracy for the same two previous regions of interest, frontal and occipital, in addition to the entire head as an ROI. The averaged classifier output resulted in a gradual increase in accuracy that peaked around 200-300 ms post stimulus. Classification was significantly above chance throughout the stimulus duration, and dropped between 500-600 ms, just after the stimulus was turned off, in all three ROI's.

In order to establish that the classifiers were picking up a true perceptual recognition signal, rather than one driven by low-level differences caused by the recognized frame containing more information than the last unrecognized frame (due to the nature of the RISE sequences), the analysis was repeated on a different slice of data containing only

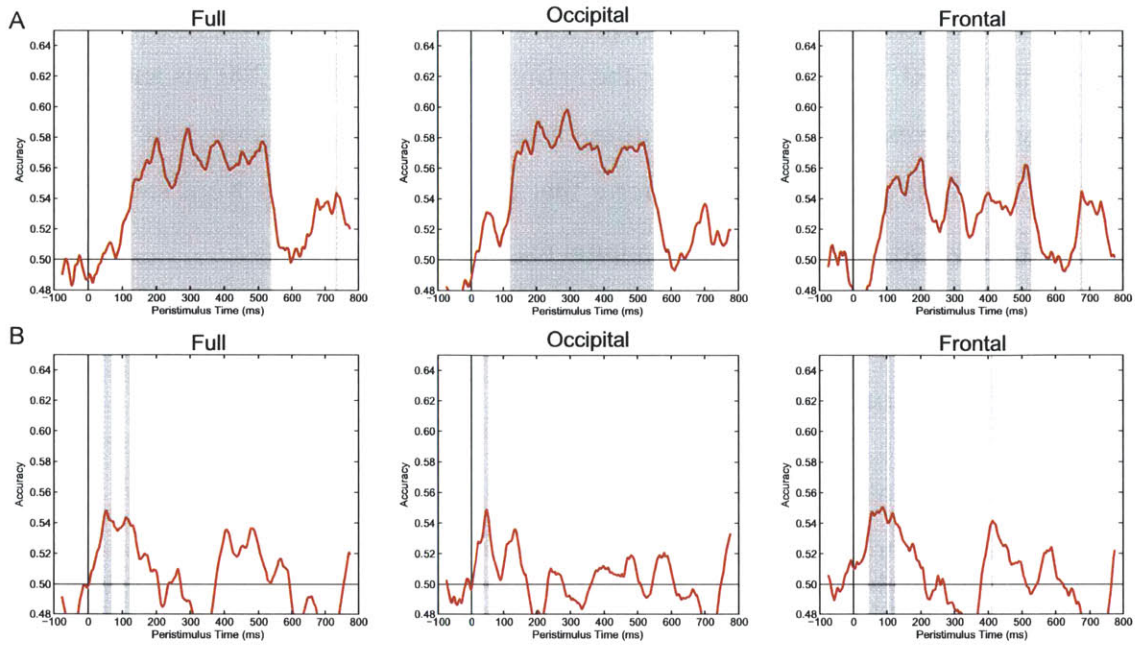


Figure 4-3: Classification by time, averaged over all subjects, for the full head, occipital regions only, and frontal regions only. A) Classification of recognized vs unrecognized images. Overall classification is similar for all three, peaking around 200-300 ms post stimulus and showing significant periods of above-chance accuracy. All three decrease by 600 ms, just after stimulus offset. B) The same analysis run on the last two unrecognized images. Classification accuracy is significantly lower, and peaks around 100 ms post stimulus. Error bands are the 99% confidence intervals of a Bernoulli process. Chance classification rate is 50%.

the last two unrecognized images. Figure 4-3b shows the same accuracy plots using the unrecognized data only. In this case, accuracy was significantly lower and peaked around 100 ms post stimulus; it did not persist throughout the stimulus duration as the recognized data set did.

Surprisingly, the accuracy plots do not simply reflect the differential waveform analysis. That is, rather than showing peaks in accuracy when the ROI's showed significant ERP differences (as seen in Figure 4-2c), the classification accuracy plots were elevated throughout the stimulus duration. In other words, classification analyses revealed significant multivariate differences in the dataset which were not visible in the averaged ERP plots.

4.2.4 Feedback occurs upon recognition

In addition to running the classification analysis on three broad regions of interest, we adopted a searchlight ROI approach which ran the classification on clusters of 6 neighboring sensors. In this manner, we were able to restrict the results of our analysis both spatially and temporally. This analysis revealed interesting dynamics on the flow of information across the cortex.

For the first recognized and last unrecognized images, Figure 4-4a shows frames from the searchlight classifier analysis showing performance across all sensors from 80 ms to 140 ms post stimulus onset. From approximately 80 ms to 140 ms post stimulus, classification accuracy was high in frontal areas and progressed backwards into occipital regions, indicative of a feedback-like process. After 140 ms, classification accuracy was high throughout the brain for the duration of the stimulus, similar to the accuracy reported in Figure 4-3a.

In the control dataset of the last two unrecognized frames, however, the searchlight analysis yielded qualitatively different results. Figure 4-4b shows this control data: there are similar early frontal differences beginning around 80 ms post stimulus, but there is a lack of feedback. After 140 ms, searchlight classification accuracy remained low throughout the brain, as in Figure 4-3b.

Additionally, we ran the searchlight classification analysis on the very first frame in the RISE sequence and the very last frame. Rather than classifying recognition states between similar images, this data segmentation includes large image-level differences, and captures the differential response between coherent and incoherent objects. The result of this classification showed only feedforward-like classification accuracy, as seen in Figure 4-4c. Therefore, image-level differences inherent in the RISE sequences could not have resulted in the feedback pattern we observe.

4.3 Discussion

We have presented results of our experiments to identify the neural correlates of the visual eureka moment in the brain.

We have found multiple signatures of visual recognition. The clearest of these occur

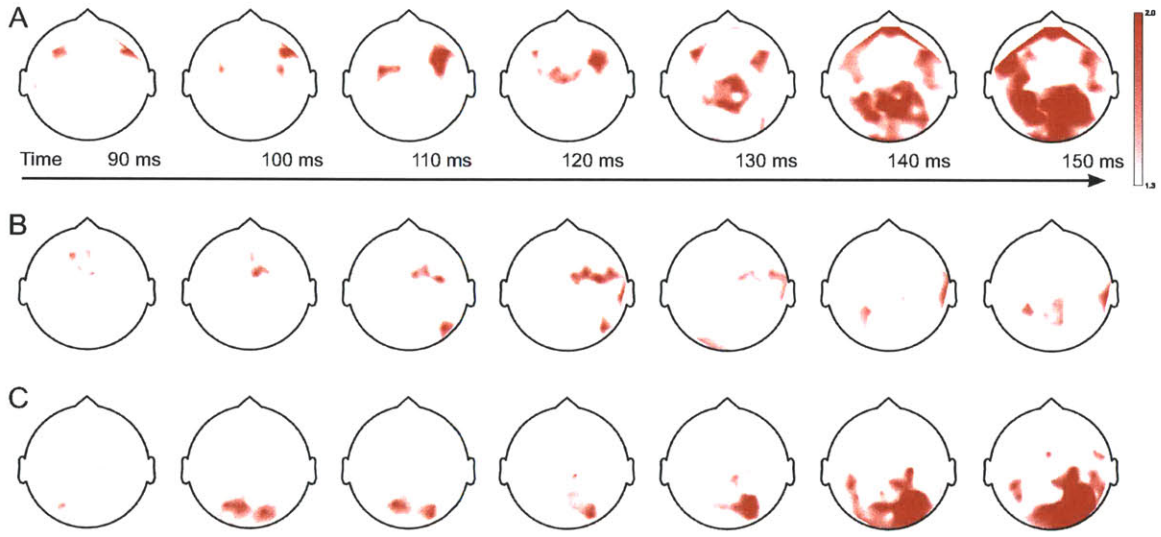


Figure 4-4: Selected frames from the searchlight classifier analysis. A) From 80 to 140 ms post stimulus, there was feedback-like activity seen as a frontal-to-occipital sweep of classification accuracy. Colors indicate the negative log10 p-score of classification performance occurring by chance throughout the brain, given a 6-sensor cluster centered on each point in space and time. B) The control dataset, which consisted of the last two unrecognized frames, showed similar early frontal activity but did not result in any significant feedback pattern. C) The same analysis run on the very first and very last RISE frames, where there are large low-level differences, revealed only feed-forward activity.

in occipital regions; they present as a decreased positive inflection around 200 ms post stimulus, and an increased negative inflection at 300 ms. We found additional and extended differences in the frontal lobe as well, and these components can serve as biomarkers for visual recognition studies.

Previous studies that have used a progressive revealing paradigm have had to contend with a significant limitation. The stimuli used gradually increase in image contrast, as well as in the amount of complex imagery presented. These changes potentially obfuscate the distinct moment of recognition in the sequence. Indeed, EEG components revealed by such sequences [Doniger et al., 2000] are found to scale progressively with the amount of the stimulus revealed. That is, for every frame of the image that was revealed in their progressive paradigm, the component was stronger until it peaked at the frame of recognition. A similar study using fMRI also showed that activity in high-level visual cortex (areas LOC and the fusiform face area) increased during a progressive revealing task, but was critically not related to the moment of recognition [Carlson et al., 2006], instead responding before

and during recognition. This is distinctly different from our component, which shows no change in the two frames prior to the recognition moment, and then a qualitative change coincident with the recognized frame. Our scrambled and thresholded images hold low-level properties steady across the set and thus result in a marked step-wise recognition behavior. Our findings of a marker for perceptual recognition are therefore not confounded by, and do not represent, areas responding to low-level aspects of the image.

In addition to the direct signatures of visual recognition, we have examined the aspects of the eureka moment using a classification paradigm. By doing so, we have discovered that previously undetected, spatially distributed, signatures of recognition exist throughout the stimulus duration as evidenced by the high accuracy rate of classification. The signature of classification was qualitatively different from that of the ERP analysis – rather than peaking at specific time points, classification accuracy was relatively high and steady throughout the stimulus duration. This may be a more generally valid distinction, and opens the door for new studies contrasting the results between traditional ERP analysis and multi-sensor classification.

Even more interesting, we believe, is the fact that we could trace the directional flow of information in the brain, and have discovered that feedback-like activity occurs during the moment of recognition in the case of degraded imagery. It has long been suspected that recognition involves significant feedback activity in the form of cortico-cortical loops [Mumford, 1992, Ullman, 1996], especially in the case of degraded or low-salience imagery; our experimental results provide direct evidence for these hypotheses. Of particular interest, though, is that we did not see feedback activity when the image-level differences were large, in analyzing the very first and last frames of the sequence. Rather, all directional information flow in this case was feed-forward. We hypothesize that a form of rapid learning occurs that changes the directional flow of information, so that future presentations of the same (previously uninterpretable) image are processed in a feed-forward manner.

4.4 Acknowledgements

We are grateful for the generous sponsorship of the James McDonnell Foundation and the Jim and Marilyn Simons Foundation.

4.5 Methods

4.5.1 Subjects

Ten adults between the ages of 20 and 30 years participated in this experiment. All adults had normal or corrected-to-normal vision, and none had a history of neurological abnormalities. Three of these subjects were excluded from analysis due to excessive artifacts, leaving seven subjects included in the analyses reported here.

4.5.2 Stimuli

A variation of a RISE technique [Sadr and Sinha, 2004] was used to create the stimuli for this experiment. Images of 32 real-world objects (for instance, a dog, school bus, shoes) were used to create 32 RISE sequences containing 20 images each. To create each sequence, the original image was decomposed into its Fourier components and recomposed with a decreasing amount of added phase noise per frame, to create 19 'scrambled' image variations. This preserved the low-level properties of the image (i.e., the amount, or magnitude, of all component frequencies) but rendered the images progressively more recognizable with decreasing phase noise. Each image (including the original) was then thresholded to become a two-tone, black and white Mooney image [Mooney, 1957]. Thus, in each sequence, a recognizable image evolved out of noise.

4.5.3 Procedure

Participants saw 32 RISE sequences in random order, with each sequence containing 20 frames. The pictures were presented against a uniform gray background. Each trial consisted of stimulus presentation (500 ms) and a post-stimulus recording period (1500 ms)

during which the screen was a uniform gray. The inter-trial interval, during which a black fixation cross was presented on a gray background, varied randomly between 500-1000 ms. Participants were required to determine whether or not they recognized each stimulus, and responded using two buttons on a button box. Participants were instructed to make their responses as quickly and accurately as possible.

4.5.4 Electrophysiological Recording and Processing

While participants were performing the above task, continuous EEG was recorded using a 128-channel Hydrocel Geodesic Sensor Net (Electrical Geodesics, Inc.), referenced online to vertex (Cz). The electrical signal was amplified with 0.1 to 100Hz band-pass filtering, digitized at a sampling rate of 1 KHz, and stored on a computer disk. Data were analyzed offline using NetStation 4.3 analysis software (Electrical Geodesics, Inc.). The continuous EEG signal was segmented into 900 ms epochs, starting 100 ms prior to stimulus onset. Data were filtered with a 1-50 Hz band-pass elliptical filter and baseline-corrected to the mean of the 100 ms period before stimulus onset. NetStation's automated artifact detection tools combed the data for eye blinks, eye movements, and bad channels. Segments were excluded from further analysis if they contained an eye blink (threshold +/- 70 μV) or eye movement (threshold +/- 50 μV). In the remaining segments, individual channels were marked bad if the difference between the maximum and minimum amplitudes across the entire segment exceeded 80 μV. If more than 10% of the 128 channels (i.e., 13 or more channels) were marked bad in a segment, the whole segment was excluded from further analysis. If fewer than 10% of the channels were marked bad in a segment, they were replaced using spherical spline interpolation. Waveforms were then re-referenced to the average reference configuration.

4.5.5 Waveform analysis

Data from the first recognized frame and the last unrecognized RISE frame (frames 0 and -1) were directly compared by performing a paired t-test on the ERP waveforms, per image, per subject, over all subjects and images. To correct for multiple comparisons, results were

only reported where $p < 0.05$ for 10 ms consecutively. Prior to performing the t-test, the ERP's were averaged across either occipital or frontal sensors, as shown in Figure 4-2b. In addition to this data segmentation, a control analysis was performed with the data from the last two unrecognized frames.

4.5.6 Multivariate pattern analysis

Data were trained and classified using the Sparse Multinomial Logistic Regression[Yamashita et al., 2008] algorithm, and N-fold cross-validated by excluding one entire image set from training at a time. Both of these algorithms were implemented in Python using the PyMVPA toolbox[Hanke et al., 2009b, Hanke et al., 2009a]. Classification accuracy was calculated from the sum of confusion matrices over all subjects, calculated independently by sensor, ROI, and/or millisecond sample depending on the exact experiment. Additionally, the results were smoothed over time by summing the confusion matrices in a given 25 ms time window (independently by ROI or sensor). Confidence intervals of classification accuracy were calculated as the 99% CI of a Bernoulli process with a true percent success rate equal to that of the classifier's.

Part IV

What changes in the brain as we learn to recognize the world?

Chapter 5

Visual cortex reorganization following late onset of sight in congenitally blind humans

Abstract

What kinds of neural changes underlie the very initial stages of human visual development? This question is operationally difficult to answer given the severe challenges of conducting brain imaging studies with newborn infants. We report here our results from an unusual opportunity to address this question by studying congenitally blind individuals whose sight we were able to restore post adolescence. Contrary to our prior understanding of critical periods for visual development, we find strong evidence of brain plasticity in these individuals. More specifically, using functional connectivity analyses, we find that there is a significant enhancement of cortical decorrelation as a function of time following the onset of sight. These findings have important implications for our understanding of brain plasticity and the heretofore largely theoretical notions of the development of efficient coding in the brain.

5.1 Introduction

Critical to our understanding of how the visual system works is how it develops. Humans acquire significant visual abilities within the first few months of life; however, the neural changes accompanying these early stages of development have been hard to elucidate due to the operational difficulties of conducting neuroimaging studies in very young children. By the time the children are old enough to be cooperative subjects in such studies, they have already passed major developmental milestones.

From the perspective of methodological feasibility, therefore, the ideal research subject to study visual learning is a developmentally mature individual who has been deprived of normal visual input. Such subjects are very rare because cases of curable congenital blindness are almost always treated within a few months after birth. However, in developing countries the advances of modern medicine are not as accessible – India, for example, has a relatively large population of children who are born with cataracts due to prenatal infections or heritable vulnerabilities. Modern eye clinics can easily restore vision to many of these children, but for various cultural and economic reasons, many do not receive treatment until well into their teens. This has disastrous consequences for their prospects for receiving education and general societal integration. In addressing the humanitarian need of providing sight to congenitally blind individuals, we are also presented with an unusual scientific opportunity to study their visual development from the moment sight is restored. An effort that we launched recently, named Project Prakash [Mandavilli, 2006], has pursued these twin objectives.

Past work from our laboratory has behaviorally documented the significant gains in visual proficiency the congenitally blind individuals exhibit as a function of time after sight restoring surgery [Ostrovsky et al., 2006, Bouvrie and Sinha, 2007, Ostrovsky et al., 2009, Held et al., 2011]. These results motivate an examination of any changes in brain organization after sight onset. In this regard, it is especially interesting to investigate the changes brought about within cortical areas that in the normal brain are visually driven. Would the influx of patterned visual information after the onset of sight result in organizational changes within these areas, or are these regions incapable of significant reorga-

nization late in life, consistent with the notion of a critical period in visual development [Wiesel and Hubel, 1965b]?

Motivated by this general question, in the present study we examine how groups of neurons interact together using resting-state functional magnetic resonance imaging (fMRI). The co-activation of groups of voxels in fMRI in the absence of explicit stimulation (i.e., 'resting state') is used to infer an underlying connectivity amongst those regions [Biswal et al., 1995]. Past work has shown that resting-state functional connectivity is very stable across time in normally-sighted individuals [Damoiseaux et al., 2006, Chen et al., 2008, Shehzad et al., 2009]. This longitudinal stability in typical individuals serves as a good reference against which to examine how, if at all, resting-state functional connectivity varies over time in the newly sighted individuals.

It is difficult to make a strong prediction regarding longitudinal trends in data from the Prakash subjects post sight-onset. On the one hand, the classical 'critical period' notion [Wiesel and Hubel, 1965b] suggests that visual cortex would have little, if any, plasticity to change in response to sight restoration after years of congenital blindness. On the other hand, our past evidence [Ostrovsky et al., 2006, Bouvrie and Sinha, 2007, Ostrovsky et al., 2009, Held et al., 2011] of behavioral visual skill improvements suggests that that the underlying neural circuitry is likely to exhibit corresponding changes as well.

5.2 Results

In order to examine whether there was a change in connectivity amongst various brain regions during visual development, we performed a resting state connectivity experiment with six subjects who ranged in age from 14 to 24 years. All subjects had been blind from birth due to bilateral congenital cataracts; prior to surgery, all subjects had the ability to detect the presence and direction of projected light, but were functionally blind to the extent that they could not count fingers at a distance of one meter. Our experiment consisted of imaging subjects during rest (that is, no explicit task was performed) using functional magnetic resonance imaging (fMRI) and examining the correlation of voxels across multiple brain regions. For each of the six subjects, the experiment was conducted longitudinally at

multiple post-surgery dates. The period of follow-up averaged 5 months.

In order see which areas, if any, showed a change in resting-state functional connectivity, we ran a searchlight analysis where the average correlation coefficient amongst neighboring voxels was calculated. This analysis quantified the local 'smoothness' of the dataset. That is, the average correlation of neighboring voxels is a measure of how closely related the activity in that region was to itself. After running the searchlight analysis, the coefficients were projected into a spherical atlas space [Fischl et al., 1999b] and a General Linear Model (GLM regression test with days post surgery and subjects as regressors) was run on each point in the atlas, to determine the dependence of the correlation coefficients on the elapsed time since each patient's surgery.

The results of the correlation searchlight GLM are shown in Figure 5-1. Contrary to expectations from the critical period hypothesis, we found evidence of dramatic reorganization within the first few months after sight onset. The data revealed a decrease in the correlation amongst neighbors occurring in occipital areas commonly associated with peripheral visual activity, based on general patterns of retinotopic mapping [Sereno et al., 1995]; no change was seen near the occipital pole. Additionally, several areas showed an increase in local correlation over time, including the frontal pole, the insular gyrus, and the medial wall.

To more rigorously test this pattern of results, we subjected the data to an independent component analysis [Beckmann and Smith, 2004, Beckmann et al., 2005]. Briefly, this type of analysis extracts spatial patterns of correlated activity that are maximally independent from one another over time.

All imaging sessions, including three conducted prior to surgery and in two normal control subjects, revealed two strong independent components of co-activated voxels localized in areas of putative visual cortex. These components are readily distinguishable from others and have been seen consistently in normal subjects as well [Beckmann et al., 2005]. Figure 5-2a shows examples of these two components for various post-surgery dates in one subject. Generally, one component appeared to be consistently located near the occipital pole, an area associated with foveal stimulation based on patterns of human retinotopic mapping [Sereno et al., 1995]; the second component encircled the first in areas associated

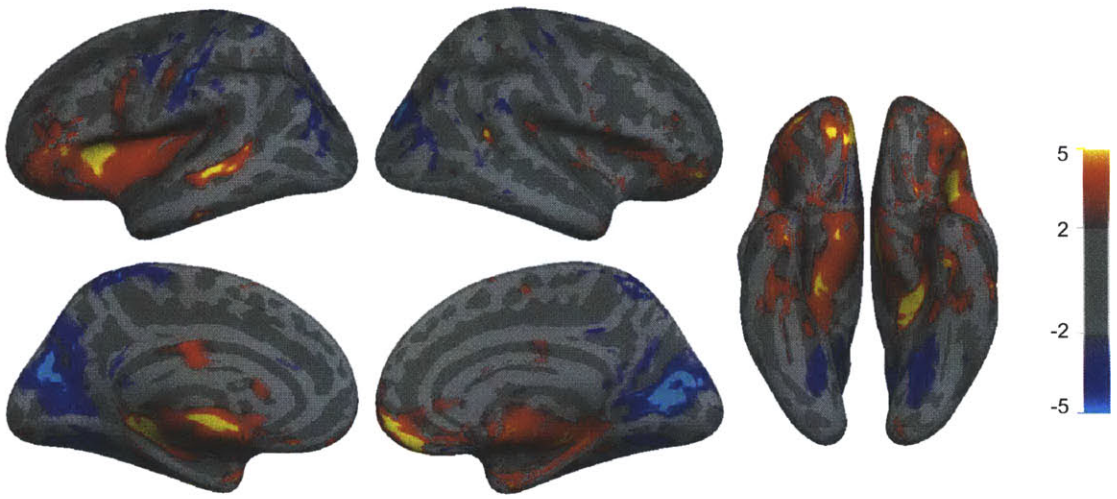


Figure 5-1: GLM results illustrating the dependence of local correlation coefficients on the number of days post surgery. Peripheral visual areas showed a decrease in correlation over time, indicating more independent activity. Frontal, insular, and medial regions, however, showed an increase in local correlation over time. The z-score of the GLM results (correlation coefficients versus time and subjects as regressors) is plotted, with the sign of the score indicating whether the correlation increased with time (positive z) or decreased (negative z).

with peripheral activation.

The data revealed marked changes in the appearance of these components over time. In order to systematically characterize these changes, we examined how the power, in terms of the percent of explained variance, of the foveal and peripheral components varied as a function of time post sight-onset. Although the components were easily recognizable in each repeat of the experiment, the amount of variance the peripheral component explained significantly decreased over time after surgery. Figure 5-2b shows that this decrease occurred consistently across all subjects examined, for the peripheral but not the foveal component (GLM regression test with days post surgery and subjects as regressors; $p = 2.2 * 10^{-5}$). Since each component represents the co-activation of voxels, a decrease in power indicates that the voxels were acting more independently, and thus, arguably, encoding more information about the visual world.

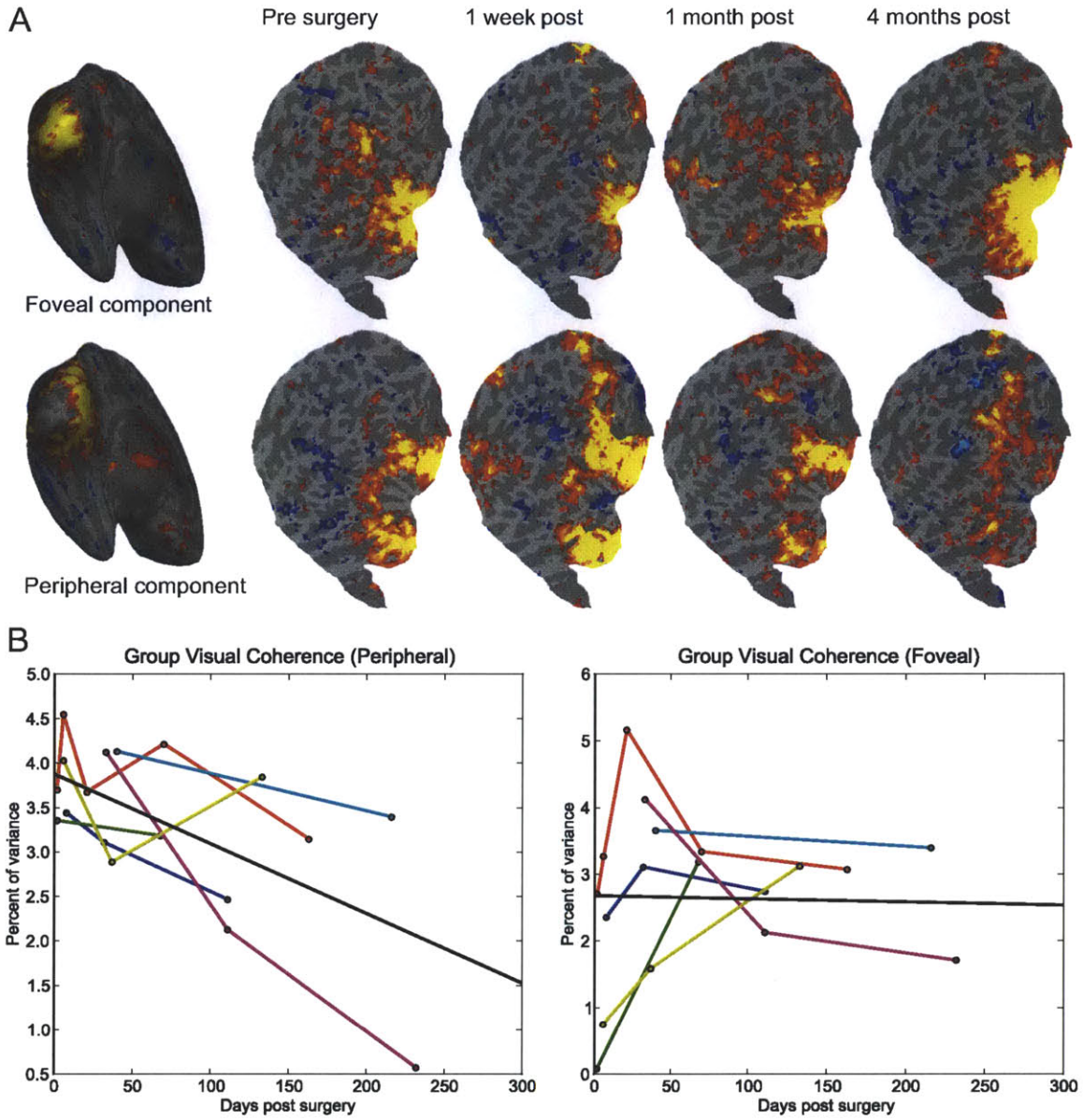


Figure 5-2: Resting-state connectivity reveals two strong independent components localized within putative visual cortex. A) Foveal and peripheral components shown in the left hemisphere, inflated and then flattened, for one example subject over an extended period of time. Both components elicited similar patterns of activity in the right hemisphere. Titles are days relative to the surgery date on which the experiment was repeated. B) The power of the peripheral component, in terms of the percent of total explained variance, significantly decreased over time for all subjects. The foveal component did not show any time dependency. Each colored line represents an individual subject; the solid black line is the GLM regression with days post surgery and subjects as regressors ($p = 2.2 * 10^{-5}$).

5.3 Discussion

We believe that these results are significant from two perspectives. First, the very fact that resting state functional connectivity changes suggests that the brain maintains a signifi-

cant measure of plasticity into late childhood or even early adulthood, contrary to classical notions of critical periods in the visual system [Wiesel and Hubel, 1965b]. This is consistent with the behavioral improvements we have observed in previous studies of congenitally blind individuals treated late in life [Ostrovsky et al., 2006, Bouvrie and Sinha, 2007, Ostrovsky et al., 2009, Held et al., 2011].

Second, and more specifically, these results provide direct empirical support for an important theoretical hypothesis about information encoding in the brain. The 'efficient coding' hypothesis [Barlow, 1989, Barlow and Foldiak, 1989, Dimitrov and Cowan, 1998] maintains that the cortex achieves efficiency in information encoding by reducing redundancy across units, or equivalently, enhancing their decorrelation. Our data showing that large regions of visual cortex undergo progressive decorrelation is consistent with this theoretical prediction of efficient coding. Decorrelation allows for factorial coding wherein responses of different units are statistically independent from each other. Such coding, also referred to as minimum entropy coding [Barlow, 1989, Atick and Redlich, 1990, Atick and Redlich, 1992], facilitates downstream neural processing by simplifying 'readout' [Barlow, 1989, Redlich, 1993]. The fact that we observe increased decorrelation after the onset of sight also suggests that efficiency of encoding is enhanced via visual experience rather than simple physiological maturation independent of experience.

The neural underpinnings of the observed decorrelation are unclear, but one plausible candidate is reduction in receptive field sizes of visual cortical cells post sight-onset. Past research has demonstrated that visual deprivation results in an increase in receptive field sizes [Ganz et al., 1968] which would lead to greater overlap and correlation across neurons. Reduction of this deprivation-induced overgrowth following onset of sight would yield reduced correlations.

The localization of these changes in the visual periphery, rather than the fovea, is a surprising finding. One potential explanation is that decorrelation also occurs in foveal regions but at a granularity finer than that detectable in our scanner, possibly due to differences between foveal and peripheral regions based on cortical magnification [Tootell et al., 1988]. Another possibility is that peripheral regions, as they encircle foveal regions, have experienced encroachment from other cortical areas and must be 'reclaimed' by visual cortex.

This is a particularly enticing hypothesis as auditory cortex is known to have strong connectivity with peripheral, but not foveal, retinotopic cortex [Falchier et al., 2002, Eckert et al., 2008, Cate et al., 2009].

This evidence of marked reorganization of functional connectivity in visual cortex brings up several interesting questions that can feasibly be addressed by Project Prakash. Notably, are there any age-related limits to the extent of such plasticity? Also, how do changes in functional connectivity relate to the inception of regional specialization in the visual cortex? Answers to these questions hold the promise of new insights into the mechanisms underlying the development of functional visual skills.

5.4 Methods

5.4.1 Patients

Subjects were selected through an outreach and ophthalmic screening program under joint supervision of Project Prakash Center for Vision and Shroff Charity Eye Hospital. The goal of this outreach program was to identify patients in need of clinical eye care across several villages in northern India; a subset of these patients who presented with treatable congenital bilateral cataracts were presented the option to volunteer for this study. The volunteer status of a patient did not affect the quality of any medical services rendered. Only subjects over the age of 14 and in good physical health were further selected for imaging. Typically, subjects were functionally blind, but maintained the ability to detect ambient light and to orient towards the direction of projected rays.

Each patient underwent cataract surgery with an intra-ocular lens implant in one or both eyes, replacing the opacified natural lens and allowing for functionally useful acuity, averaging 20/80. One patient did not exhibit an increase in acuity post surgery, likely due to residual optic nerve damage; this patient was excluded from analysis. Six remaining subjects were used for analysis.

5.4.2 Imaging

Functional imaging was conducted at the Mahajan Imaging Center, New Delhi, India, in accordance with the rules and regulations set forth by Institutional Review Boards at MIT and Schroff Charity Eye Hospital. Scanning was conducted on a 1.5 T General Electric MRI machine. A high-resolution T_1-weighted FSPGR anatomical scan was acquired for each participant (FOV 256×256, 1 mm^3 resolution). To measure BOLD contrast, between 19 and 33 slices (depending on the size of the subject's head) parallel to the AC/PC line were acquired using standard T_2^*-weighted gradient-echo echoplanar imaging (TR 2000 ms, TE 50 ms, flip angle 90°, slice thickness 5 mm, in-plane resolution 3.4 × 3.4 mm). Resting-state data was captured in two or three runs of 77 TR's (approximately 2.5 minutes per run) each, depending on the time available per subject on that day. Subjects returned to the scanning center periodically to repeat the experimental session depending on their availability.

5.4.3 Analysis

Cortical reconstruction and volumetric segmentation was performed with the Freesurfer image analysis suite, which is documented and freely available for download online. The technical details of these procedures are described in prior publications [Dale et al., 1999, Fischl et al., 1999a]. Functional scans were motion corrected and despiked in AFNI [Cox, 1996] and coregistered to individual subjects' anatomical scans using a combination of manual alignment, custom scripts and SPM8. No slice timing correction was performed. Data were projected into a spherical surface atlas space using automatic cortical surface parcellation [Fischl et al., 1999b, Fischl et al., 2004, Desikan et al., 2006] in Freesurfer.

5.4.4 Correlation searchlight

To determine the local correlation amongst neighboring voxels, a searchlight routine from PyMVPA [Hanke et al., 2009b, Hanke et al., 2009a] was run on the motion-corrected, despiked functional data. The pairwise temporal correlation amongst all neighboring voxels in a sphere with a radius of 11 mm was averaged and projected into the spherical atlas

space. A general linear model was used to determine the dependence of local correlation for each point in the atlas space on time, by regressing against time and individual subjects.

5.4.5 ICA mapping

Independent component maps were calculated separately for each session, using multi-run temporal concatenation. Analysis was carried out using Probabilistic Independent Component Analysis [Beckmann and Smith, 2004] as implemented in MELODIC (Multivariate Exploratory Linear Decomposition into Independent Components) Version 3.10, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). The peripheral and foveal independent components were identified by eye and projected onto the flattened cortical surface using Freesurfer for display. To determine the dependence of each component on time post surgery, the percent of variance each component explained was fed into a general linear model with regressors against time and individual subjects (to account for a difference of mean variance across subjects).

Chapter 6

Development of face-selective regions after the onset of sight in congenitally blind individuals

Abstract

What kinds of neural changes underly the initial stages of visual development? This question is operationally difficult to answer in humans given the severe challenges of conducting experiments with newborn infants. We report here our results from an unusual opportunity to address this question by studying congenitally blind adults whose sight we were able to restore. Contrary to our prior understanding of critical periods for visual development, we find strong evidence of brain plasticity in these individuals. In this study, we show how regions of the brain selective for faces and objects develop after the onset of sight. We find that face-selective responses develop non-monotonically over a period of several months, but object-selective responses appear to be stable immediately after sight restoration.

6.1 Introduction

Instances of sight onset after several years of congenital blindness are extremely rare. It is estimated that over the entire recorded history of neuroscience, fewer than 20 such cases have been reported and studied in any detail [Gregory and Wallace, 1963, Sacks, 1995, Valvo, 1971]. This is because an infant born with a curable form of blindness is typically treated within a few months after birth.

Although rare, cases of late sight onset offer a unique opportunity to address several fundamental questions regarding brain plasticity and early stages of visual learning. Project Prakash, a humanitarian and scientific initiative from our laboratory [Mandavilli, 2006], has enabled us to identify, treat and study such individuals in India. Our work thus far has documented the significant gains in visual proficiency that congenitally blind individuals exhibit as a function of time after sight-restorative surgery [Ostrovsky et al., 2006, Ostrovsky et al., 2009, Bouvrie and Sinha, 2007, Held et al., 2011]. These results motivate an examination of changes in brain organization after sight onset. In this regard, it is especially interesting to investigate the changes brought about within cortical areas that in the normal brain are visually driven. Would the influx of patterned visual information after the onset of sight result in organizational changes within these areas?

Motivated by this general question, in the present study we examine how regions of the ventral visual system develop selective responses to image categories. It is well established that the ventral stream contains several regions that can be distinguished by their selective responses to certain categories of objects; the most well studied of these regions is the fusiform facial area, or FFA [Kanwisher et al., 1997], which responds maximally to visual stimuli containing faces – a stimulus type of great ecological significance. Very little, however, is known about how the ventral stream develops. Imaging results in younger children offer conflicting evidence of the existence of a face-selective region in the fusiform gyrus, and there is general consensus that such a region continues to develop into adolescence (see [Grill-Spector et al., 2008] for review). The prolonged development of the FFA stands in contrast to developmental patterns seen for generic object-specific responses (in the lateral occipital complex, or LOC [Grill-Spector et al., 1998]), which appear adult-like by the age

of seven in normal children (also reviewed in [Grill-Spector et al., 2008]). However, no prior experiments can tease apart the effects of visual learning and overall system development. In this context, longitudinal studies of our congenitally blind, sight-restored subjects provide us the opportunity to address several fundamental questions:

1. Does the human brain remain adequately plastic to allow the development of functionally specialized cortical regions, despite a lifetime of visual deprivation?
2. If localized functional specialization does develop late in life, is it spatially and topographically similar in organization to that found in the normally developed brain?
3. How rapidly do functional modules develop after the onset of sight?
4. Do different kinds of functional specialization have different timelines of development after the onset of sight?
5. How does the timeline of neural development correlate with the timeline of development of perceptual skills? Is there a systematic temporal lag relating one timeline to the other, and what inferences about causality can be derived from such a lag?

These are the questions that motivate the experiments we describe here.

6.2 Results

In order to examine the development of category-specific response regions in the ventral stream, an fMRI experiment was conducted on blind subjects whose sight had been recently restored. In this experiment, images of several distinct categories – faces, objects, places, and scrambled images – were shown to the subjects in a block paradigm. The experiment was conducted as close to the surgery date as possible and repeated several times in the following weeks and months, depending on each subject's availability.

6.2.1 Rapid development of face-selective regions

To examine the neural response of these subjects to faces, a statistical parametric map was created from the BOLD activation contrast of Faces – Objects.

Figure 6-1 shows the activation in response to faces for all three subjects, projected onto the ventral surface of the right hemisphere only. Activation in the left hemisphere was not as as reliably detected. All three subjects showed patterns of activity that significantly changed over time relative to the surgery date. Notably, all three subjects showed an early FFA-like precursor. Two subjects, BP and SC, exhibited small early face-selective responses that grew significantly over time; a third subject, CP, exhibited a larger FFA-like precursor which grew and then was refined by 3 months post surgery. Subject BP exhibited a slightly later pattern of growth-refinement, with the size of his FFA diminishing from 1-8 months before growing large again by 10 months post surgery. It is possible, but not definitive, that subject SC also showed a reduction in his FFA-like response from 2 days to 1 month post surgery.

6.2.2 Face responses are related to behavioral changes in face detection

In order to establish a correlation between the changes in the FFA-like response and each subject's ability to recognize faces, a behavioral experiment was conducted outside of the scanner. In this experiment, subjects were presented images of faces and non-faces and asked to indicate whether or not they perceived a face. Their d-prime was calculated, and the experiment was repeated several times over the subsequent weeks to establish a similar time-course to that of the scanning experiment.

Figure 6-2 shows each subject's face detection d-prime overlaid with the calculated size of their right-hemisphere FFA-like response, over time post surgery. For our purposes, the size of the FFA response was calculated by counting the number of superthreshold voxels ($z > 2$) located within the right hemisphere fusiform gyrus (as determined by automatic surface parcellation in Freesurfer, see Methods).

Shortly after surgery, subject BP was able to distinguish faces from non-faces (d-prime greater than 2), but subjects CP and SC were not (d-prime less than 1). All three, however, had a significant FFA-like precursor, as seen here and in Figure 6-1, indicating that the FFA precursor was in fact present before the subjects were reliably able to detect faces. As

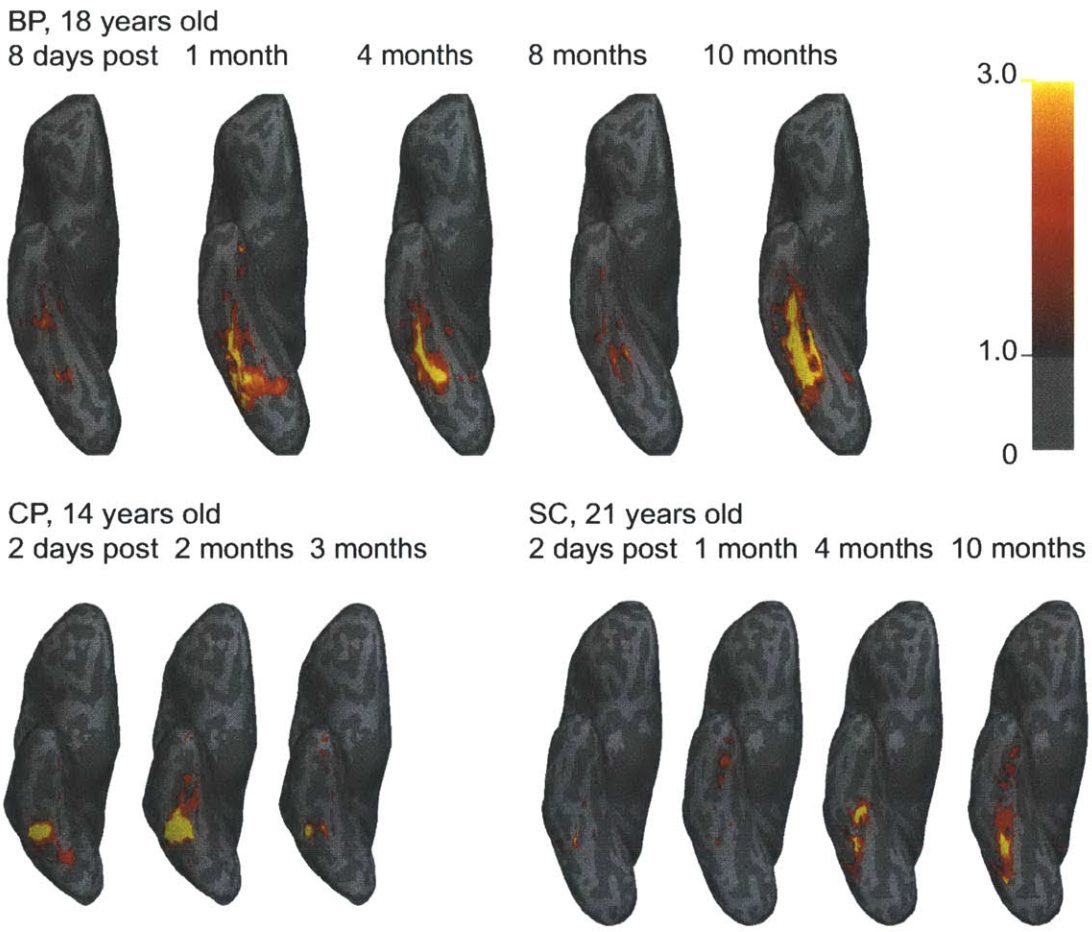


Figure 6-1: Face-selective responses for all subjects following sight-restorative surgery. Shown is the z-score response for the contrast Faces – Objects on the ventral surface of the right hemisphere. Left hemispheres did not exhibit reliable face-selective activations in all subjects. Notably, all three subjects had early precursors of an FFA-like response and exhibited significant and rapid change over time.

time progressed, all three subjects improved at the face-detection task as the size of their FFA-like response changed. Notably, the behavioral responses plateaued before the size of the FFA response became stable in all three subjects, indicating that the FFA response is plastic long after the establishment of a simple face-detection task.

Additionally, for the two subjects which exhibited a growth-refinement face response (subjects BP and CP), the peak of their FFA-like responses occurred simultaneously with the beginning of the behavioral plateau.

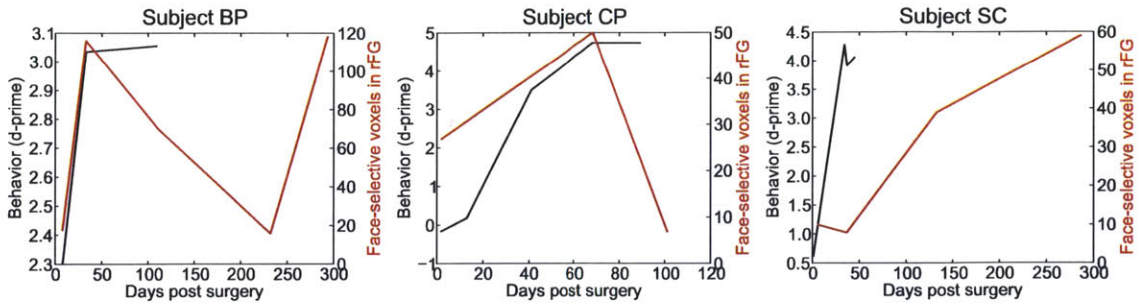


Figure 6-2: Face-detection behavior and how it relates to the FFA-like response in all three subjects. All three subjects showed a significant FFA-like response before they improved at a face-detection task (subjects SC and CP performed at chance at face detection but had significant FFA-like responses within 20 days post surgery). Additionally, the FFA response continued to change after the behavioral results plateaued in all three subjects. Finally, the FFA-like response peak coincided with the beginning of the behavioral plateau in the two subjects who exhibited a growth-refinement pattern in their neural responses.

6.2.3 Object-selective responses do not noticeably change

Because the literature suggests that face- and object-selective regions develop differentially [Grill-Spector et al., 2008], we were very interested to see whether object-selective regions developed differentially to the face-selective regions in the recently sighted patients. To this end, we analyzed the object-selective responses using the contrast Objects – Scrambled Images.

Figure 6-3 shows lateral views of object-selective responses in all three subjects. Similar to the face-selective responses shown above, all three subjects exhibited an LOC-like precursor shortly after surgery. However, this precursor did not noticeably change: unlike the face-selective responses shown above, there was no clear pattern of growth or plasticity that emerged. In fact, although subject CP arguably showed modest growth of the object-selective response in putative LOC, subjects BP and SC, if anything, showed a decrease in LOC-like activation.

6.3 Discussion

We have conducted longitudinal neuroimaging experiments in blind patients whose sight had been surgically restored, and have found that the onset of patterned visual information

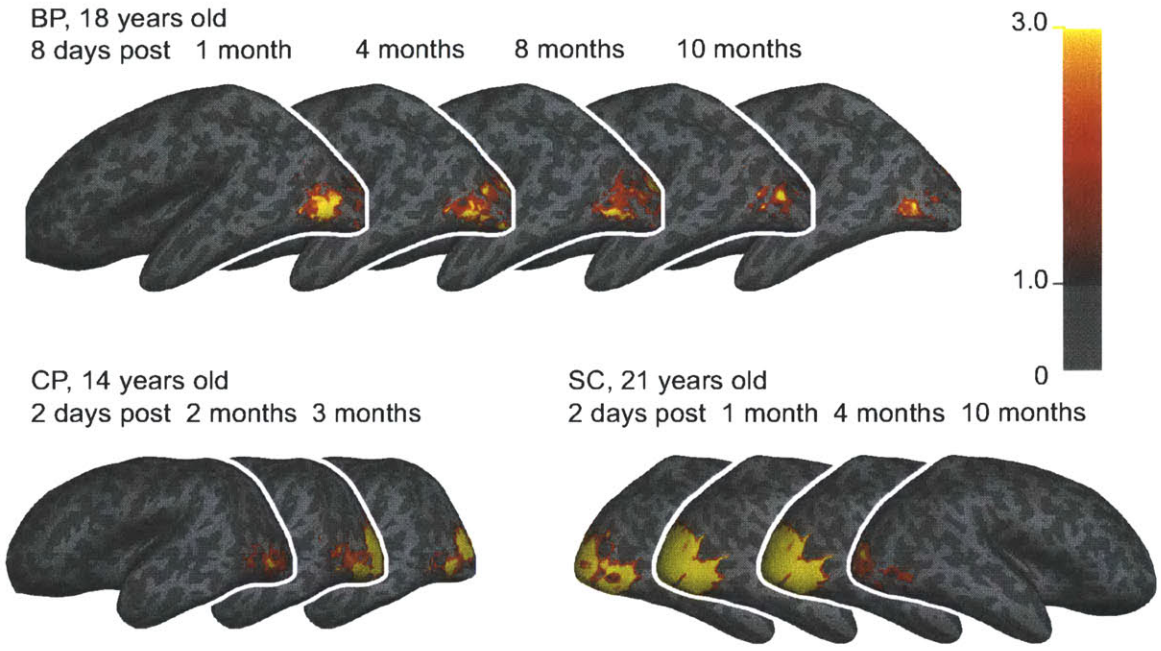


Figure 6-3: Object-specific responses for all subjects. As with the face-selective responses, all three subjects showed object-specific responses immediately after surgery in putative LOC. Although some change is evident between sessions, there is no clear pattern of overall growth or variability across all subjects.

results in rapid and extensive modification of the ventral visual stream. Specifically, we report here that face-selective responses, specifically putative FFA, emerge rapidly over the course of the first few months after the onset of sight.

Several aspects of these results are notable. First, the data reveal that a precursor to an FFA-like region exists in subjects within days after surgery. Second, this precursor undergoes rapid change over the first few months of visual experience, with evidence pointing towards an overgrowth and refinement pattern, until a relatively normal FFA region is formed. This rapid onset and growth of face-selectivity stands in contrast to the literature, which indicates a prolonged development of face selectivity [Grill-Spector et al., 2008] over the course of years, if not decades. However, we also note that the FFA-like response remained variable after behavior in a face-detection task plateaued, suggesting that the FFA response may continue to be refined for months or years after the onset of sight, in agreement with the developmental literature.

It is also interesting to note that object-selective responses do not exhibit such notice-

able change as face-selective regions. This roughly agrees with the literature in that LOC regions develop much more rapidly than face-selective regions [Grill-Spector et al., 2008]; it could be the case that LOC develops in our subjects within one or two days, before our imaging began, or that it relies more heavily on pre-developed pathways and less on visual learning.

These results address the broad questions we listed in the introduction. First, and perhaps most significantly, the data show a remarkable degree of plasticity in the visual cortex of the newly sighted despite several years of profound visual deprivation. These results run counter to the notion of a 'critical period' for visual plasticity [Wiesel and Hubel, 1965b] and have important scientific and clinical implications. Scientifically, they motivate a re-examination of the issue of how, and how much, cortical plasticity is maintained until late adolescence or early adulthood. From the applied perspective, they suggest optimistic prognoses for sight restoring surgeries late in life, and perhaps recovery from other kinds of extended sensory deprivations. Second, the specific case of the emergence of face-related responses suggests that the cortical localization of such functions might be similar in the brains of the newly sighted versus controls, despite their dramatically shifted developmental timelines. It will be interesting to probe the underlying reasons for this spatial overlap. Third, the functional specialization becomes evident very rapidly after the onset of sight suggesting that a very modest amount of visual experience is sufficient to set up this organization. Fourth, different kinds of functional specialization exhibit different developmental timelines, as we have observed for general object responses on the one hand and face-specific responses on the other. Finally, in the domain of face perception, the timelines of development of neural specificity and behavioral skill exhibit an interesting relationship. The presence of an FFA-like precursor before subjects improved at the face-detection task suggests that face-selective cortex may exist before face-detection skills become behaviorally manifest.

6.4 Methods

6.4.1 Patients

Subjects were selected through an outreach and ophthalmic screening program under joint supervision of Project Prakash Center for Vision and Shroff Charity Eye Hospital. The goal of this outreach program was to identify patients in need of clinical eye care across several villages in north India; a subset of these patients who presented with treatable congenital, bilateral, cataracts were presented the option to volunteer for this study. The volunteer status of a patient did not affect the quality of any medical services rendered. Only subjects over the age of 13 and in good physical health were further selected for imaging. Typically, subjects were functionally blind, but maintained the ability to detect ambient light and the direction of projected rays.

Each patient underwent cataract surgery with an intra-ocular lens implant in one or both eyes, replacing the opacified natural lens and allowing for functionally useful acuity, averaging 20/80. One patient did not exhibit an increase in acuity post surgery, likely due to residual optic nerve damage; this patient was excluded from analysis.

In addition to the three subjects reported here, an additional seven patients were imaged but excluded from analysis for failure to elicit a consistent control activation to scrambled images. The high rate of failure is likely due to a multitude of factors, including: high motion artifacts due to the young age of the patients, falling asleep due to the young age or fatigue from an extended stay in the hospital, inability to focus on the presented images due to nystagmus (which was present in most of our patients, as it has a high co-morbidity rate with congenital cataracts) or misunderstanding of the task due to age or language barriers, among other reasons.

6.4.2 Imaging

Functional imaging was conducted at the Mahajan Imaging Center, New Delhi, India, in accordance with the rules and regulations set forth by Institutional Review Boards at MIT and Schroff Charity Eye Hospital. Scanning was conducted on a 1.5 T General Electric

MRI machine. A high-resolution T_1-weighted FSPGR anatomical scan was acquired for each participant (FOV 256×256, 1 mm^3 resolution). To measure BOLD contrast, 19-33 slices (depending on the size of the subject's head) parallel to the AC/PC line were acquired using standard T_2^*-weighted gradient-echo echoplanar imaging (TR 2000 ms, TE 50 ms, flip angle 90°, slice thickness 5 mm, in-plane resolution 3.4 × 3.4 mm). Subjects were shown various images falling into one of several categories – faces, places, objects, and scrambled images – at a rate of one image per second in a block paradigm, so that each category of images was displayed for 14 seconds before the next category was shown. Images were presented in full color at approximately 20 degrees of visual angle. The scrambled images were produced by phase-scrambling the Fourier components of images of objects, individually by RGB channel. Once all four categories were displayed, a 14 second resting period occurred before all four categories were repeated again in a pseudorandom fashion. In this manner, each run consisted of 4 repeats of each category block, and including a 14 second rest at the beginning of the scan run, lasted 292 seconds. Depending on the available time, 4-6 of these scan runs were completed per subject, per scan session; the session was repeated as many times as possible over the months following the patient's surgery. The same images were used in each scan session.

6.4.3 Analysis

Cortical reconstruction and volumetric segmentation was performed with the Freesurfer image analysis suite, which is documented and freely available for download online. The technical details of these procedures are described in prior publications [Dale et al., 1999, Fischl et al., 1999a]. Functional scans were motion corrected and despiked in AFNI [Cox, 1996] and coregistered to individual subjects' anatomical scans using a combination of manual alignment, custom scripts and SPM8. Data were projected into a spherical surface space using automatic cortical surface parcellation [Fischl et al., 1999b, Fischl et al., 2004, Desikan et al., 2006] in Freesurfer. Data were additionally smoothed using a 3d Gaussian kernel of width 5mm. To analyze the BOLD effect specific to each stimulus type, a GLM contrast was analyzed in FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl)

per scan run and the results combined using fixed-effects modeling per individual subject [Woolrich et al., 2001], independently for each session. Results were masked to anatomically defined regions of interest, determined via automatic surface parcellation in Freesurfer, before displaying: the fusiform gyrus for face-selective responses, and lateral-occipital cortex for object-specific responses.

6.4.4 Behavioral assessment

Subjects were shown a set of 300 images and asked to categorize the images, one at a time, as faces or non-faces. This set of images is consisted of 60 non-face images collected from natural scenes devoid of faces, 180 false alarm images selected from the Pittsburgh Pattern Recognition face-detection algorithm [Schneiderman, 2004], and 60 genuine faces. Each image was made monochrome and normalized for scale, luminance and contrast. Stimuli were presented on a computer screen one by one for unlimited time until the subject made a response. Subjects were allowed to take as many breaks as they would like at anytime during the experiment.

Part V

Conclusions

Chapter 7

Conclusions

7.1 Conclusions

I have explored three major themes in this thesis, all related to how the brain changes as we learn to see. Each finding highlights how plastic the visual system is, both in rapid learning – in the developed brain – and when we learn to see during development.

7.2 Prior information facilitates object recognition

In this chapter, we explored the role of prior knowledge and expectations shaping everyday perception. Specifically, we wanted to find signatures in the brain where knowledge of a coherent stimulus shaped the representation of the same image in a degraded form.

We addressed this question using Mooney images, which were unrecognizable to subjects upon their first viewing. After exposure to the coherent stimulus, the same images, which were previously meaningless, became interpretable. By using a multivariate decoding technique we were able to find where information about these images is represented in the brain, and show how that information changed as the coherent stimulus was learned.

The increased decoding accuracy of primed versus unprimed images suggests that there is an increase in image-specific information in regions of the brain exhibiting this form of priming. In general, the regions in which we found an increased decoding ability coincide with all levels of the ventral visual processing system. This included higher-level areas such as fusiform gyrus and lateral occipital cortex, and lower levels including peri-calcarine cortex, and, in particular, V2 and V3.

Interestingly, we did not find any evidence of information encoding in classically primed regions. This distinction is notable because it suggests that some previously reported effects of priming are not, in fact, involved in representing visual stimuli at all.

Additionally, we showed that prior information is carried over complex object features, as we were unable to recreate a similar effect using simple orientation-based stimuli.

7.3 The moment of recognition involves feedback

In this experiment, we presented evidence describing the effects of recognition – and in particular, the visual eureka moment – in the brain.

We found multiple signatures of visual recognition. The clearest of these occur in occipital regions; they present as a decreased positive inflection around 200 ms post stimulus, and an increased negative inflection at 300 ms. We found additional and extended differences in the frontal lobe as well, but the occipital components are sufficient to serve as a biomarker for visual recognition studies.

In addition to the direct signatures of visual recognition, we examined the aspects of the eureka moment using a classification paradigm. By doing so, we discovered that smaller signatures of recognition exist throughout the stimulus duration. Even more interesting, however, is the fact that we could trace out the directional flow of information in the brain, and discovered that feedback-like activity occurs during the moment of recognition. This same technique also revealed that viewing more coherent images utilizes exclusively feedforward mechanisms.

7.4 The brain remains plastic when deprived of sight

We have conducted neuroimaging experiments in blind patients whose sight had been surgically restored, and discovered that the onset of patterned visual information results in rapid modification of the ventral visual stream. Specifically, we found two important changes in the visual system: that resting-state connectivity is modulated with learning, and that key modules, like the FFA, develop in tandem with behavioral learning.

We believe that these results are significant from several perspectives. First, the fact that resting state functional connectivity changes suggests that the brain maintains a significant measure of plasticity into early adulthood. This is consistent with the behavioral improvements we have observed in previous studies of congenitally blind individuals treated late in life [Bouvrie and Sinha, 2007, Ostrovsky et al., 2006, Ostrovsky et al., 2009].

Second, we showed that a precursor to FFA-like regions exist in subjects within days

of surgery. This precursor undergoes rapid change over the first few months of visual experience, with evidence pointing towards an overgrowth and refinement pattern, until a relatively normal FFA region is formed. This rapid onset and growth of face-selectivity stands in contrast to the literature, which indicates a prolonged development of face selectivity [Grill-Spector et al., 2008].

It is also interesting to note that object-selective responses did not exhibit such noticeable change as face-selective regions. This roughly agrees with the literature in that LOC regions develop much more rapidly than face-selective regions [Grill-Spector et al., 2008]; it could be the case that LOC develops in our subjects within one or two days, before our imaging began, or that it relies more heavily on pre-developed pathways and less on visual learning.

7.5 Final thoughts

The issue of the genesis of visual meaning is central to our exploration of vision. Working with developmentally mature and developmentally nascent visual systems, I have uncovered the neural correlates of such genesis over timescales that range from milliseconds to months. While these results are gratifying, they are merely the beginning of a more comprehensive investigation into the precise mechanisms by which we achieve our robust pattern recognition abilities.

Part VI

Appendix

Appendix A

Single-unit recording efforts

A.1 Supplementing our findings with single-unit recording approaches

As is evident, I have used a diversity of techniques in my work, including behavioral assessments, EEG and fMRI. One technique that combines excellent spatial localization with fine temporal resolution is single-unit electrophysiology. Indeed, I devoted significant effort towards incorporating this methodology in my work in order to illuminate issues of rapid learning at the single neuron level. As I detail below, so far the results have been mixed, but this is an approach that holds great promise and should be pursued in further work related to the questions we have considered here.

A.2 Introduction

In order to expand on our results showing information priming in visual cortex through fMRI, we were very interested to see if we could show a similar effect in single unit recordings in macaque monkeys. Previous work had shown that neurons in macaque V4 learn to encode more information about degraded stimuli over time: over several training sessions, V4 neurons show increased selectivity for trained stimuli in a delayed match-to-sample task [Rainer et al., 2004]. We were specifically interested to see if these results could be extended using a single-trial learning paradigm similar to our behavioral results in fMRI, and additionally if the results could be extended into V1.

One possibility of how such an effect could be instantiated in cortex is via competitive attentional modulation [Reynolds et al., 1999], where neurons are biased to represent one feature in their receptive field over another when multiple stimuli are present. In the case of a degraded image, prior knowledge of a stimulus could bias a neuron to represent the features in its receptive fields that encode the true stimulus, rather than extraneous features that also appear in its receptive field.

A.3 Approach

We conducted several experiments in macaques to try and elicit a similar effect, with varying levels of success. In all experiments, the macaques were trained to fixate on a small visual cue and rewarded with a drop of juice for a successful trial. Some experiments additionally required the monkey to perform an active task based on visual stimuli, as described below.

While the monkeys were viewing the various stimuli, we recorded from one to four electrodes implanted in V1 and/or V4 chambers, as targeted by cortical anatomy and verified by retinotopic mapping of cells at the beginning of each session. Single units were isolated on each electrode when possible, and the receptive field and preferred orientation were mapped out according to standard procedures. This receptive field was used to present the experimental stimuli in each recording session.

RISE (Random Image Structure Evolution) sequences were used to present progressively more or less visual information to the monkey over time. In this implementation, the images were phase scrambled in Fourier space [Dakin et al., 2002], and the amount of coherence to the original, non-scrambled image defined the amount of visual information presented. It is known that people's ability to recognize individual images from a RISE sequence is greater when the amount of information decreases over time [Sadr and Sinha, 2004]; that is, the 'descending' pathway gives the viewer substantial prior information about the stimulus, which aids recognition. Therefore, we hypothesized that individual neurons should show a similar form of hysteresis.

A.4 Results

A.4.1 Orientation discrimination

Our first attempt in isolating a RISE effect consisted of training the monkey to discriminate between multiple orientations in a delayed match-to-sample task. However, the monkey subject showed considerable difficulty learning this task; his performance plateaued at substandard accuracy after an extended training period (Figure A-1). For this reason and the

fact that we discovered that oriented stimuli do not elicit any information priming in fMRI, we abandoned this project to develop a more sophisticated approach.

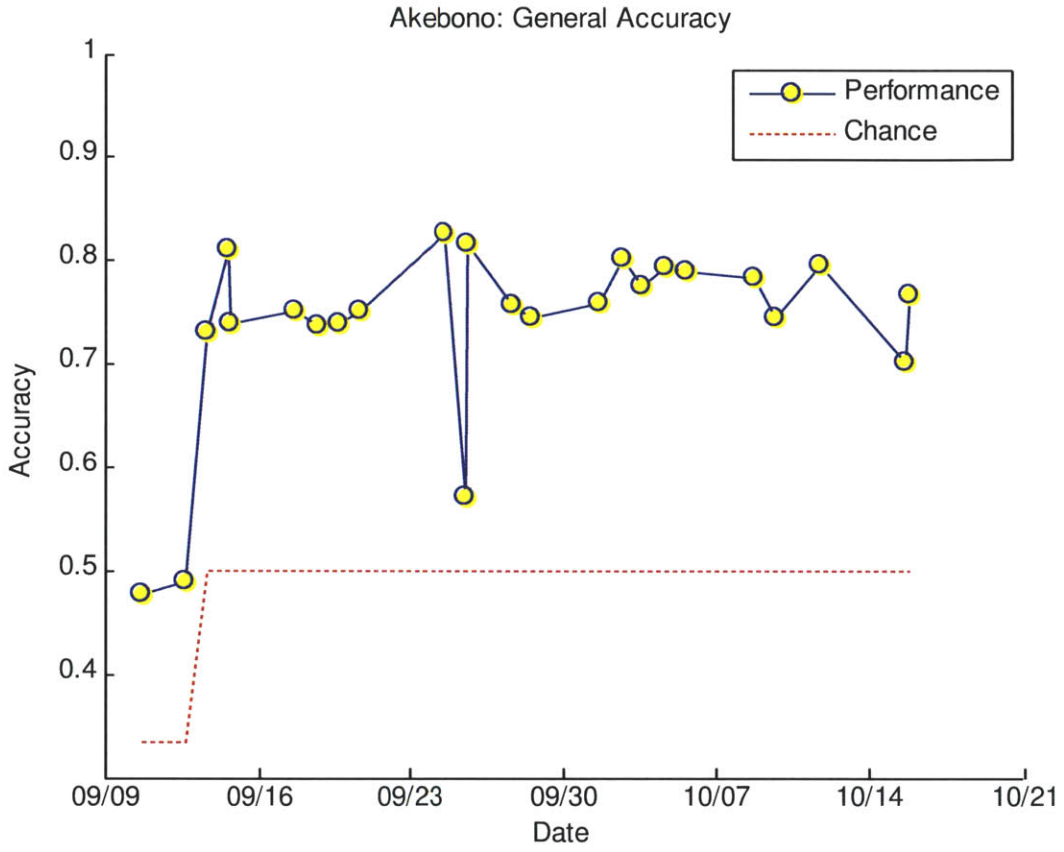


Figure A-1: Accuracy at discriminating between multiple orientations for a single monkey subject over time. The monkey was presented with a forced choice of 2 or 4 orientations in a delayed match-to-sample task, but accuracy never reached a satisfactory level (90%).

A.4.2 Passive RISE yields no information priming

In this experiment, the monkey was only required to fixate while a RISE sequence was presented in a neuron's receptive field. As no stimulus-related task was required of the monkey, we referred to this experiment as "passive RISE." In order to prevent simple effects like repetition suppression, two RISE sequences were presented in an interchanging fashion; see Figure A-2.

We hypothesized that one of two effects might occur: individual neurons might 1) have greater activation when the RISE sequences descended, as more information was available

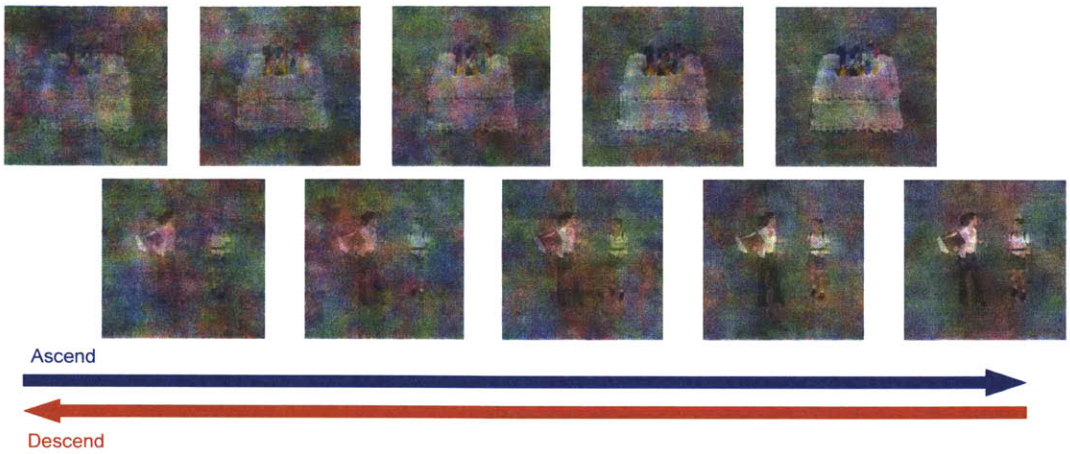


Figure A-2: Overall stimulus methodology for the passive RISE experiment. Before each stimulus presentation, the monkey was required to maintain fixation on a centrally-presented cue. Each stimulus was presented within the receptive field of a recorded neuron. Each block was designated as either an "ascending" or "descending" block, which determined the order in which the stimuli were presented. Each trial consisted of a single stimulus from one of two RISE sequences; the overall coherence of the stimulus either increased or decreased depending on the designation of the block. The two RISE sequences were interchanged pseudo-randomly, but each stimulus within a sequence was progressively more or less coherent than the previous (and all coherence levels were shown for both sequences).

to the neuron and presumably the monkey was attending more to the stimulus than the noise features; or 2) neurons might show greater selectivity bias for one stimulus over another in a competitive manner.

Overall, neurons did show a contrast-based gain effect. Figure A-3 shows this effect clearly, for a representative V4 cell, where increased coherence in both the ascending and descending conditions increased the response of the cell. However, there was no overall modulation of any cell's firing rate based on the experimental conditions of ascending or descending.

Our second hypothesis was that neurons might show an increased selectivity for their preferred stimuli, rather than an overall modulation of firing rate. Even though V1 and V4 neurons are not tuned for the complex objects we used in this experiment, each object does have a disparate set of low-level features that the neurons might show selectivity for. Selectivity was defined as the preference for the first RISE sequence over the second in every block, and a selectivity index was calculated to reflect this. Figure A-4 shows the

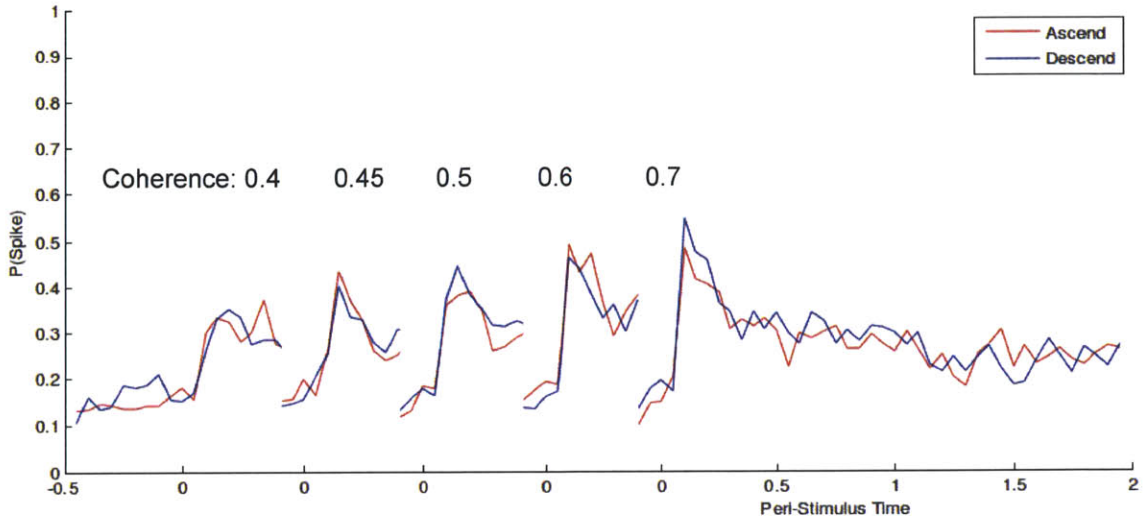


Figure A-3: Peri-stimulus time histogram for a representative V4 cell in the passive RISE experiment. This cell shows a marked increase in response based on higher coherence (and effective contrast) levels, but shows no overall modulation by the ascending or descending experimental conditions.

selectivity for a representative V1 neuron for all blocks, plotted as descending selectivity versus ascending selectivity. If selectivity was consistent for this neuron and there was a strong competitive bias, the points should form a strong regression line with a slope greater than 1 (because selectivity would be increased in the descending vs ascending condition). However, our results do not show a consistent result – selectivity, as defined here, often changes from one stimulus to the other (where points fall in the odd quadrants, with positive x and negative y, and vice-versa). This could be simply due to technical reasons – for instance, it is possible that single units were not isolated well enough to record consistent selectivity.

Given that we could find no evidence for hysteresis in the form of selectivity or activation modulation, a simple explanation is that a passive RISE effect does not exist. That is, it is entirely possible that the behavioral effect, including information priming, requires that the subject attend to the object of interest.

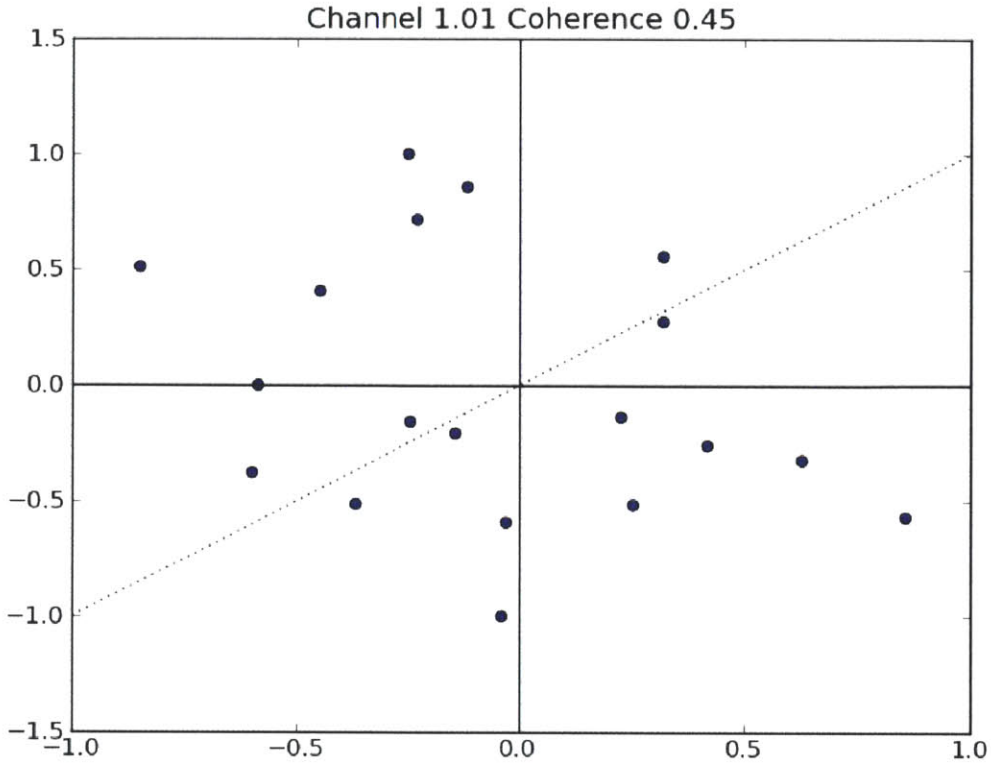


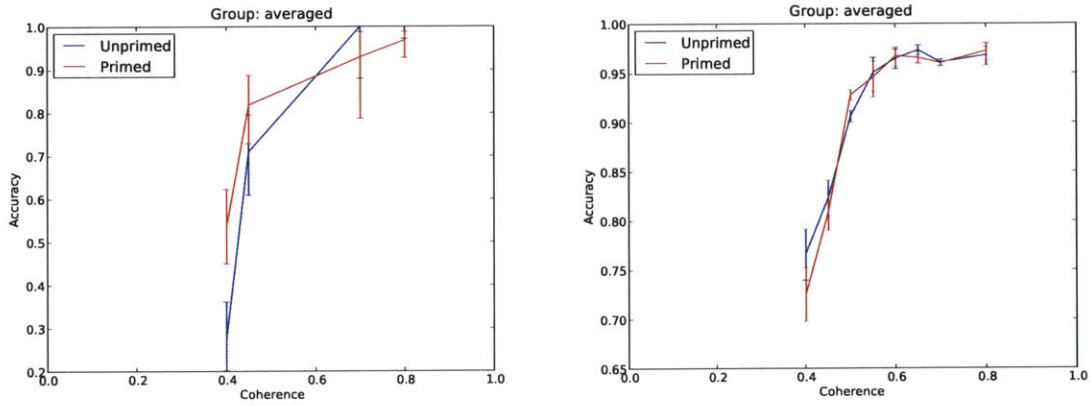
Figure A-4: Selectivity of a representative V1 neuron for two targets in ascending (x-axis) and descending (y-axis) conditions. Selectivity is defined via the index $\frac{resp_1 - resp_2}{resp_1 + resp_2}$. For a neuron with consistent selectivity, points should fall in the even quadrants ((+, +) and (-,-)), and increased selectivity in descending vs ascending conditions would be seen as a regression line with slope greater than 1. These results show neither consistent nor increased selectivity.

A.4.3 Detection task

In order to require the monkeys to attend the RISE stimuli and increase our chances of discovering a single-unit effect, a new experiment was devised. In the active RISE task, the monkey was trained to fixate while a series of noise images were presented the neuron's receptive field. After a number of these distracter noise images were presented, a degraded target image was presented in the same location, and the monkey was required to release a lever indicating detection of the target. At the beginning of the trial, a cue image was presented centrally (outside of the neuron's receptive field) at full coherence; if the cue image was identical to the target image, the trial was considered to be primed (analogous

to the descending method in the previous experiment). The amount of coherence presented in the target image was changed daily to try to hone in on any possible effect.

Although the monkey learned the task, he did not show nearly as strong behavioral evidence for information priming as did human subjects at the same task. Figure A-5 shows a comparison of human and monkey results for this experiment. The reason for this discrepancy is unclear, but it is possible the monkey was simply over-trained at detecting these images – human subjects were only exposed to the experiment once, whereas the monkey was rewarded for high accuracy over a period of several months of training. Indeed, the monkey's overall accuracy was higher, even for low-coherence targets.



(a) (b)

Figure A-5: Behavioral results for the active RISE task. A) In preliminary testing, human subjects showed a significantly increased ability to detect targets in noise when primed with an identical image. B) Averaged over a period of 4 months, one monkey subject did not elicit such strong behavioral evidence.

Although we did record single units from this monkey during the course of this experiment, the effective unit yield was incredibly low. Out of 10 successful recording sessions over a period of 1 month of recording time (ignoring sessions where units were isolated but the monkey failed to work, or where no usable units were isolated), only about 5 neurons proved to be stable and isolated enough for analysis. Other neurons died early into recording, or proved not to respond well to the stimuli. Like the passive RISE experiment, the neurons that were analyzed showed no significant preference for primed or unprimed conditions. This is not surprising given that the monkey showed no behavioral response difference for primed vs. unprimed stimuli.

A.5 Discussion

We attempted to measure the single-unit responses in macaque monkeys to RISE sequences, to see if priming increased information in low-level visual cortex.

Unfortunately, the work proved more challenging than fruitful. Out of several attempts, we were unable to reproduce a RISE-like behavioral response in a macaque subject – in no instance did we find evidence of the monkey utilizing prior knowledge of a stimulus to enhance his perception. In the successful recordings we were able to make, especially in the passive task, we found no difference in neural processing between the primed and unprimed conditions.

It is difficult to draw a strong conclusion from these results, as many of the causes for the failure of these experiments are technical – on top of experimental – reasons. In addition to better technical recordings, it is clear that a stronger behavioral component will be required of any future work.

Acronyms

FFA Fusiform Facial Area. 35

fMRI Functional Magnetic Resonance Imaging. 31

GLM General Linear Model. 31

PPA Parahippocampal Place Area. 35

SMLR Sparse Multinomial Logistic Regression. 38

SPM Statistical Parametric Map. 31

SVM Support Vector Machine. 35

Bibliography

[Ahissar et al., 2009] Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philos Trans R Soc Lond B Biol Sci*, 364(1515):285–299.

[Atick and Redlich, 1990] Atick, J. J. and Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Computation*, 2(3):308–320.

[Atick and Redlich, 1992] Atick, J. J. and Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation*, 4(2):196–210.

[Bar et al., 2006] Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Schmidt, A. M., Dale, A. M., Hamalainen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., and Halgren, E. (2006). Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A*, 103(2):449–454.

[Barlow, 1989] Barlow, H. (1989). Unsupervised learning. *Neural Computation*, 1(3):295–311.

[Barlow and Foldiak, 1989] Barlow, H. B. and Foldiak, P. (1989). *Adaptation and decorrelation in the cortex*, pages 54–72. Addison-Wesley, England.

[Beckmann et al., 2005] Beckmann, C. F., DeLuca, M., Devlin, J. T., and Smith, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philos Trans R Soc Lond B Biol Sci*, 360(1457):1001–1013.

[Beckmann et al., 2003] Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2003). General multilevel linear modeling for group analysis in fmri. *NeuroImage*, 20(2):1052 – 1063.

[Beckmann and Smith, 2004] Beckmann, C. F. and Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging*, 23(2):137–152.

[Behrmann et al., 2006a] Behrmann, M., Avidan, G., Leonard, G. L., Kimchi, R., Luna, B., Humphreys, K., and Minshew, N. (2006a). Configural processing in autism and its relationship to face processing. *Neuropsychologia*, 44(1):110–129.

[Behrmann et al., 2006b] Behrmann, M., Thomas, C., and Humphreys, K. (2006b). Seeing it differently: visual processing in autism. *Trends Cogn Sci*, 10(6):258–264.

[Belliveau et al., 1991] Belliveau, J. W., Kennedy, D. N., McKinstry, R. C., Buchbinder, B. R., Weisskoff, R. M., Cohen, M. S., Vevea, J. M., Brady, T. J., and Rosen, B. R. (1991). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, 254(5032):716–719.

[Bentin and Golland, 2002] Bentin, S. and Golland, Y. (2002). Meaningful processing of meaningless stimuli: the influence of perceptual experience on early visual processing of faces. *Cognition*, 86(1):B1–14.

[Bentin et al., 2002] Bentin, S., Sagiv, N., Mecklinger, A., Friederici, A., and von Cramon, Y. D. (2002). Priming visual face-processing mechanisms: electrophysiological evidence. *Psychol Sci*, 13(2):190–193.

[Biswal et al., 1995] Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magn Reson Med*, 34(4):537–541.

[Bouvrie and Sinha, 2007] Bouvrie, J. V. and Sinha, P. (2007). Visual object concept discovery: Observations in congenitally blind children, and a computational approach. *Neurocomputing*, 70(13-15):2218 – 2233. Selected papers from the 3rd International Conference on Development and Learning (ICDL 2004), 3rd International Conference on Development and Learning; Time series prediction competition: the CATS benchmark.

[Boynton, 2005] Boynton, G. (2005). Imaging orientation selectivity: decoding conscious perception in v 1. *Nature Neuroscience*, 8(5):541–542.

[Bulthoff et al., 1998] Bulthoff, I., Bulthoff, H., and Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nat Neurosci*, 1(3):254–257.

[Carlson et al., 2006] Carlson, T., Grol, M. J., and Verstraten, F. A. J. (2006). Dynamics of visual recognition revealed by fmri. *Neuroimage*, 32(2):892–905.

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

[Cate et al., 2009] Cate, A. D., Herron, T. J., Yund, E. W., Stecker, G. C., Rinne, T., Kang, X., Petkov, C. I., Disbrow, E. A., and Woods, D. L. (2009). Auditory attention activates peripheral visual cortex. *PLoS One*, 4(2):e4645.

[Cavanagh, 1991] Cavanagh, P. (1991). What's up in top-down processing? In Gorea, A., editor, *Representations of Vision: Trends and Tacit Assumptions in Vision Research*, pages 295–304. Cambridge University Press.

[Chen et al., 2008] Chen, S., Ross, T. J., Zhan, W., Myers, C. S., Chuang, K.-S., Heishman, S. J., Stein, E. A., and Yang, Y. (2008). Group independent component analysis reveals consistent resting-state networks across multiple sessions. *Brain Res*, 1239:141–151.

[Cheng et al., 2001] Cheng, K., Waggoner, R. A., and Tanaka, K. (2001). Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron*, 32(2):359–374.

[Cortes and Vapnik, 1995] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

[Cox and Savoy, 2003] Cox, D. D. and 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):261–270.

[Cox, 1996] Cox, R. W. (1996). Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*, 29(3):162–173.

[Craft et al., 2007] Craft, E., Schutze, H., Niebur, E., and von der Heydt, R. (2007). A neural model of figure-ground organization. *J Neurophysiol*, 97(6):4310–4326.

[Dakin et al., 2002] Dakin, S. C., Hess, R. F., Ledgeway, T., and Achtman, R. L. (2002). What causes non-monotonic tuning of fmri response to noisy images? *Curr Biol*, 12(14):R476–7; author reply R478.

[Dale et al., 1999] Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. i. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194.

[Damoiseaux et al., 2006] Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., and Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proc Natl Acad Sci U S A*, 103(37):13848–13853.

[de Beeck, 2009] de Beeck, H. P. O. (2009). Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fmri analyses? *Neuroimage*.

[Desikan et al., 2006] Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., and Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980.

[Dimitrov and Cowan, 1998] Dimitrov, A. and Cowan, J. D. (1998). Spatial decorrelation in orientation-selective cortical cells. *Neural Comput*, 10(7):1779–1795.

[Dolan et al., 1997] Dolan, R. J., Fink, G. R., Rolls, E., Booth, M., Holmes, A., Frackowiak, R. S., and Friston, K. J. (1997). How the brain learns to see objects and faces in an impoverished context. *Nature*, 389(6651):596–599.

[Doniger et al., 2000] Doniger, G. M., Foxe, J. J., Murray, M. M., Higgins, B. A., Snodgrass, J. G., Schroeder, C. E., and Javitt, D. C. (2000). Activation timecourse of ventral visual stream object-recognition areas: High density electrical mapping of perceptual closure processes. *Journal of Cognitive Neuroscience*, 12(4):615–621.

[Dougherty et al., 2003] Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J., and Wandell, B. A. (2003). Visual field representations and locations of visual areas v1/2/3 in human visual cortex. *J Vis*, 3(10):586–598.

[Durkalski et al., 2003] Durkalski, V. L., Palesch, Y. Y., Lipsitz, S. R., and Rust, P. F. (2003). Analysis of clustered matched-pair data. *Stat Med*, 22(15):2417–2428.

[Eckert et al., 2008] Eckert, M. A., Kamdar, N. V., Chang, C. E., Beckmann, C. F., Greicius, M. D., and Menon, V. (2008). A cross-modal system linking primary auditory and visual cortices: evidence from intrinsic fmri connectivity analysis. *Hum Brain Mapp*, 29(7):848–857.

[Epstein and Kanwisher, 1998] Epstein, R. and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676):598–601.

[Falchier et al., 2002] Falchier, A., Clavagnier, S., Barone, P., and Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci*, 22(13):5749–5759.

[Farah, 1990] Farah, M. (1990). *Visual Agnosia: Disorders of Object Recognition and What They Tell Us about Normal Vision*. MIT Press.

[Fine et al., 2003] Fine, I., Wade, A. R., Brewer, A. A., May, M. G., Goodman, D. F., Boynton, G. M., Wandell, B. A., and MacLeod, D. I. A. (2003). Long-term deprivation affects visual perception and cortex. *Nat Neurosci*, 6(9):915–916.

[Fischl et al., 1999a] Fischl, B., Sereno, M. I., and Dale, A. M. (1999a). Cortical surface-based analysis. ii: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2):195–207.

[Fischl et al., 1999b] Fischl, B., Sereno, M. I., Tootell, R. B., and Dale, A. M. (1999b). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp*, 8(4):272–284.

[Fischl et al., 2004] Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., and Dale, A. M. (2004). Automatically parcellating the human cerebral cortex. *Cereb Cortex*, 14(1):11–22.

[Freiman et al., 2010] Freiman, M., Sela, Y., Edrei, Y., Pappo, O., Joskowicz, L., and Abramovitch, R. (2010). Multi-class svm model for fmri-based classification and grading of liver fibrosis. volume 7624, page 76240S. SPIE.

[Friston, 2007] Friston, K. J. (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press.

[Ganel et al., 2006] Ganel, T., Gonzalez, C. L. R., Valyear, K. F., Culham, J. C., Goodale, M. A., and Kohler, S. (2006). The relationship between fmri adaptation and repetition priming. *Neuroimage*, 32(3):1432–1440.

[Ganz et al., 1968] Ganz, L., Fitch, M., and Satterberg, J. A. (1968). The selective effect of visual deprivation on receptive field shape determined neurophysiologically. *Exp Neurol*, 22(4):614–637.

[Gregory, 1970] Gregory, R. L. (1970). *The intelligent eye*. McGraw-Hill, New York.

[Gregory and Wallace, 1963] Gregory, R. L. and Wallace, J. G. (1963). Recovery from early blindness: A case study. In *Experimental Psychology Monograph No. 2*. Heffer.

[Grill-Spector et al., 2008] Grill-Spector, K., Golarai, G., and Gabrieli, J. (2008). Developmental neuroimaging of the human ventral visual cortex. *Trends Cogn Sci*, 12(4):152–162.

[Grill-Spector et al., 1998] Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., and Malach, R. (1998). A sequence of object-processing stages revealed by fmri in the human occipital lobe. *Hum Brain Mapp*, 6(4):316–328.

[Grill-Spector and Malach, 2001] Grill-Spector, K. and Malach, R. (2001). fmr-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)*, 107(1-3):293–321.

[Haenny et al., 1988] Haenny, P. E., Maunsell, J. H., and Schiller, P. H. (1988). State dependent activity in monkey visual cortex. ii. retinal and extraretinal factors in v4. *Exp Brain Res*, 69(2):245–259.

[Hanke et al., 2009a] Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., and Pollmann, S. (2009a). Pymvpa: A python toolbox for multivariate pattern analysis of fmri data. *Neuroinformatics*, 7(1):37–53.

[Hanke et al., 2009b] Hanke, M., Halchenko, Y. O., Sederberg, P. B., Olivetti, E., Frund, I., Rieger, J. W., Herrmann, C. S., Haxby, J. V., Hanson, S. J., and Pollmann, S. (2009b). Pymvpa: A unifying approach to the analysis of neuroscientific data. *Frontiers in Neuroinformatics*, 3:3.

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

[Haynes and Rees, 2005] Haynes, J.-D. and Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci*, 8(5):686–691.

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

[Held et al., 2011] Held, R., Ostrovsky, Y., Gandhi, T., de Gelder, B., S. Ganesh, U. M., and Sinha, P. (2011). The newly sighted fail to match seen with felt (in press). *Nat Neurosci*.

[Hof and Bouras, 1991] Hof, P. R. and Bouras, C. (1991). Object recognition deficit in alzheimer's disease: possible disconnection of the occipito-temporal component of the visual system. *Neurosci Lett*, 122(1):53–56.

[Hsieh et al., 2010] Hsieh, P.-J., Vul, E., and Kanwisher, N. (2010). Recognition alters the spatial pattern of fmri activation in early retinotopic cortex. *J Neurophysiol*, 103(3):1501–1507.

[Hubel and Wiesel, 1959] Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J Physiol*, 148:574–591.

[Hubel and Wiesel, 1962] Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, 160:106–154.

[Hubel and Wiesel, 1963] Hubel, D. H. and Wiesel, T. N. (1963). Shape and arrangement of columns in cat's striate cortex. *J Physiol*, 165:559–568.

[Hupe et al., 1998] Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by v1, v2 and v3 neurons. *Nature*, 394(6695):784–787.

[James et al., 2000] James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S., and Goodale, M. A. (2000). The effects of visual object priming on brain activation before and after recognition. *Curr Biol*, 10(17):1017–1024.

[Kamitani and Sawahata, 2009] Kamitani, Y. and Sawahata, Y. (2009). Spatial smoothing hurts localization but not information: Pitfalls for brain mappers. *Neuroimage*.

[Kamitani and Tong, 2005] Kamitani, Y. and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nat Neurosci*, 8(5):679–685.

[Kamitani and Tong, 2006] Kamitani, Y. and Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Curr Biol*, 16(11):1096–1102.

[Kanwisher et al., 1997] Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*, 17(11):4302–4311.

[Kashino, 2006] Kashino, M. (2006). Phonemic restoration: The brain creates missing speech sounds. *Acoustical Science and Technology*, 27(6):318–321.

[Kastner et al., 1998] Kastner, S., De Weerd, P., Desimone, R., and Ungerleider, L. G. (1998). Mechanisms of Directed Attention in the Human Extrastriate Cortex as Revealed by Functional MRI. *Science*, 282(5386):108–111.

[Kersten et al., 2004] Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as bayesian inference. *Annu Rev Psychol*, 55:271–304.

[Koivisto and Revonsuo, 2003] Koivisto, M. and Revonsuo, A. (2003). Object recognition in the cerebral hemispheres as revealed by visual field experiments. *Laterality*, 8(2):135–153.

[Kurson, 2008] Kurson, R. (2008). *Crashing Through: The extraordinary true story of the man who dared to see.* Random House Trade Paperbacks, New York.

[Large et al., 2007] Large, M.-E., Aldcroft, A., and Vilis, T. (2007). Task-related laterality effects in the lateral occipital complex. *Brain Res*, 1128(1):130–138.

[Mandavilli, 2006] Mandavilli, A. (2006). Visual neuroscience: look and learn. *Nature*, 441(7091):271–272.

[Marsolek, 1995] Marsolek, C. J. (1995). Abstract visual-form representations in the left cerebral hemisphere. *J Exp Psychol Hum Percept Perform*, 21(2):375–386.

[Marsolek, 1999] Marsolek, C. J. (1999). Dissociable Neural Subsystems Underlie Abstract and Specific Object Recognition. *Psychological Science*, 10(2):111–118.

[Martinez-Trujillo and Treue, 2004] Martinez-Trujillo, J. C. and Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol*, 14(9):744–751.

[McAdams and Maunsell, 2000] McAdams, C. J. and Maunsell, J. H. (2000). Attention to both space and feature modulates neuronal responses in macaque area v4. *J Neurophysiol*, 83(3):1751–1755.

[Mitchell et al., 2004] Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M., and Newman, S. (2004). Learning to decode cognitive states from brain images. *Machine Learning*, 57(1):145–175.

[Miyawaki et al., 2008] Miyawaki, Y., Uchida, H., Yamashita, O., aki Sato, M., Morito, Y., Tanabe, H. C., Sadato, N., and Kamitani, Y. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, 60(5):915–929.

[Mooney, 1957] Mooney, C. M. (1957). Age in the development of closure ability in children. *Can J Psychol*, 11(4):219–226.

[Motter, 1994a] Motter, B. C. (1994a). Neural correlates of attentive selection for color or luminance in extrastriate area v4. *J Neurosci*, 14(4):2178–2189.

[Motter, 1994b] Motter, B. C. (1994b). Neural correlates of feature selective memory and pop-out in extrastriate area v4. *J Neurosci*, 14(4):2190–2199.

[Mumford, 1992] Mumford, D. (1992). On the computational architecture of the neocortex. ii. the role of cortico-cortical loops. *Biol Cybern*, 66(3):241–251.

[Naselaris et al., 2009] Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., and Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, 63(6):902–915.

[Norman et al., 2006] Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fmri data. *Trends Cogn Sci*, 10(9):424–430.

[O'Craven and Kanwisher, 2000] O'Craven, K. M. and Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, 12(6):1013–1023.

[Olshausen and Field, 1996] Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609.

[Ostrovsky et al., 2006] Ostrovsky, Y., Andalman, A., and Sinha, P. (2006). Vision following extended congenital blindness. *Psychol Sci*, 17(12):1009–1014.

[Ostrovsky et al., 2009] Ostrovsky, Y., Meyers, E., Ganesh, S., Mathur, U., and Sinha, P. (2009). Visual parsing after recovery from blindness. *Psychol Sci*, 20(12):1484–1491.

[Pinsk et al., 2004] Pinsk, M. A., Doniger, G. M., and Kastner, S. (2004). Push-pull mechanism of selective attention in human extrastriate cortex. *J Neurophysiol*, 92(0022-3077 (Linking)):622–9.

[Polyn et al., 2005] Polyn, S. M., Natu, V. S., Cohen, J. D., and Norman, K. A. (2005). Category-Specific Cortical Activity Precedes Retrieval During Memory Search. *Science*, 310(5756):1963–1966.

[Rainer et al., 2004] Rainer, G., Lee, H., and Logothetis, N. K. (2004). The effect of learning on the function of monkey extrastriate visual cortex. *PLoS Biol*, 2(2):E44.

[Redlich, 1993] Redlich, A. N. (1993). Supervised factorial learning. *Neural Comput.*, 5(5):750–766.

[Reynolds et al., 1999] Reynolds, J. H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas v2 and v4. *J Neurosci*, 19(5):1736–1753.

[Rossion et al., 2000] Rossion, B., Dricot, L., Devolder, A., Bodart, J.-M., Crommelinck, M., De Gelder, B., and Zoontjes, R. (2000). Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *J. Cognitive Neuroscience*, 12(5):793–802.

[Rubin et al., 1997] Rubin, N., Nakayama, K., and Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory-contour perception. *Curr Biol*, 7(7):461–467.

[Sacks, 1995] Sacks, O. (1995). To see and not see. In *An anthropologist on Mars*. Vintage, New York.

[Sadr and Sinha, 2004] Sadr, J. and Sinha, P. (2004). Object recognition and random image structure evolution. *Cognitive Science: A Multidisciplinary Journal*, 28(2):259–287.

[Saenz et al., 2002] Saenz, M., Buracas, G. T., and Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nat Neurosci*, 5(7):631–632.

[Schneiderman, 2004] Schneiderman, H. (2004). Feature-centric evaluation for efficient cascaded object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.

[Serences et al., 2009] Serences, J. T., Saproo, S., Scolari, M., Ho, T., and Muftuler, L. T. (2009). Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *NeuroImage*, 44(1):223 – 231.

[Sereno et al., 1995] Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., and Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(5212):889–893.

[Serre et al., 2007] Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Trans Pattern Anal Mach Intell*, 29(3):411–426.

[Shehzad et al., 2009] Shehzad, Z., Kelly, A. M. C., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., Lee, S. H., Margulies, D. S., Roy, A. K., Biswal, B. B., Petkova, E., Castellanos, F. X., and Milham, M. P. (2009). The resting brain: unconstrained yet reliable. *Cereb Cortex*, 19(10):2209–2229.

[Shubhangi and Hiremath, 2009] Shubhangi, D. C. and Hiremath, P. S. (2009). Support vector machine (svm) classifier for brain tumor detection. pages 444–448.

[Smith et al., 1980] Smith, D. C., Lorber, R., Stanford, L. R., and Loop, M. S. (1980). Visual acuity following binocular deprivation in the cat. *Brain Res*, 183(1):1–11.

[Sonnenburg et al., 2006] Sonnenburg, S., Bernhard, B. S., Bennett, P., and Parrado-Hernandez, E. (2006). Large scale multiple kernel learning. *Journal of Machine Learning Research*, 7.

[Sun et al., 2009] Sun, D., van Erp, T. G., Thompson, P. M., Bearden, C. E., Daley, M., Kushan, L., Hardt, M. E., Nuechterlein, K. H., Toga, A. W., and Cannon, T. D. (2009). Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: Classification analysis using probabilistic brain atlas and machine learning algorithms. *Biol Psychiatry*, 66(11):1055–1060.

[Swisher et al., 2010] Swisher, J. D., Gatenby, J. C., Gore, J. C., Wolfe, B. A., Moon, C.-H., Kim, S.-G., and Tong, F. (2010). Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *J Neurosci*, 30(1):325–330.

[Timney et al., 1978] Timney, B., Mitchell, D. E., and Giffin, F. (1978). The development of vision in cats after extended periods of dark-rearing. *Exp Brain Res*, 31(4):547–560.

[Tootell et al., 1988] Tootell, R. B., Switkes, E., Silverman, M. S., and Hamilton, S. L. (1988). Functional anatomy of macaque striate cortex. ii. retinotopic organization. *J Neurosci*, 8(5):1531–1568.

[Ullman, 1996] Ullman, S. (1996). *High-level vision: Object recognition and Visual Cognition*. MIT Press.

[Valvo, 1971] Valvo, A. (1971). *Sight restoration after long-term blindness: The problems and behavior patterns of visual rehabilitation*. American Foundation for the Blind, New York.

[Vapnik, 1998] Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley-Interscience.

[Vapnik, 1999] Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Trans Neural Netw*, 10(5):988–999.

[Vaughan et al., 1968] Vaughan, H. G., Costa, L. D., and Ritter, W. (1968). Topography of the human motor potential. *Electroencephalogr Clin Neurophysiol*, 25(1):1–10.

[Wiesel and Hubel, 1965a] Wiesel, T. N. and Hubel, D. H. (1965a). Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J Neurophysiol*, 28(6):1029–1040.

[Wiesel and Hubel, 1965b] Wiesel, T. N. and Hubel, D. H. (1965b). Extent of recovery from the effects of visual deprivation in kittens. *J Neurophysiol*, 28(6):1060–1072.

[Woolrich et al., 2001] Woolrich, M. W., Ripley, B. D., Brady, M., and Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of fmri data. *NeuroImage*, 14(6):1370 – 1386.

[Yacoub et al., 2008] Yacoub, E., Harel, N., and Ugurbil, K. (2008). High-field fmri unveils orientation columns in humans. *Proc Natl Acad Sci U S A*, 105(30):10607–10612.

[Yamashita et al., 2008] Yamashita, O., aki Sato, M., Yoshioka, T., Tong, F., and Kamitani, Y. (2008). Sparse estimation automatically selects voxels relevant for the decoding of fmri activity patterns. *Neuroimage*, 42(4):1414–1429.

[Zernicki, 1991] Zernicki, B. (1991). Visual discrimination learning in binocularly deprived cats: 20 years of studies in the nencki institute. *Brain Res Brain Res Rev*, 16(1):1–13.