

ACQUISITION, STORAGE, AND RETRIEVAL
IN DIGITAL AND BIOLOGICAL BRAINS

Jeremy R. Manning

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania in Partial
Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2011

Supervisor of Dissertation
Michael J. Kahana
Professor of Psychology

Graduate Group Chairperson
Rita Balice-Gordon
Professor of Neuroscience

Dissertation Committee:

Geoffrey K. Aguirre
Assistant Professor of
Neurology

David H. Brainard
Professor of Psychology

Kenneth A. Norman
Associate Professor of
Psychology

Alan A. Stocker
Assistant Professor of
Psychology

COPYRIGHT

Jeremy R. Manning

2011

To Talia.

Acknowledgments

First and foremost, I would like to thank my advisor, Dr. Mike Kahana for the mentorship, encouragement, support, and intellectual stimulation he provided me during my time in his Computational Memory Laboratory. Mike was also a major force in attracting me to study neuroscience as an undergraduate — my first “real” research project grew out of a final paper for one of his courses.

I am also grateful to the current and former members of the lab for their knowledge, entertainment, and support: Dr. Josh Jacobs, Dr. Brad Lega, Dr. Sean Polyn, Dr. Per Sederberg, Dr. Misha Serruya, Dr. Marieke van Vugt, Dr. Christoph Weidemann, Dr. Kareem Zaghloul, John Burke, Patrick Crutchley, Tim Lew, Lynn Lohnas, Nicole Long, Neal Morton, Jonathan Miller, Ashwin Ramayya, Emily Rosenberg, and Alec Solway. In particular, I’d like to single out Josh Jacobs for his role in training me in the arts of human electrocorticographic data collection, analysis, and interpretation.

My dissertation committee chair, Dr. David Brainard, has been a continuing source of scientific inspiration — particularly in his ability to think clearly about complicated questions and explain them simply. In addition to his mentorship on Chapter 2 of this dissertation, I have found David’s suggestions and comments on all aspects of my work to be especially insightful and helpful. My other dissertation committee members, Dr. Geoff Aguirre, Dr. Ken Norman, and Dr. Alan Stocker,

have also been excellent sources of feedback on my work.

The support and training from the Neuroscience Graduate Group — notably Jane Hoshi, Fiona Cowan, Dr. Mikey Nusbaum, and Dr. Rita Balice-Gordon — has been invaluable and much appreciated.

I am indebted to the patients who donated their time and cooperation (and without whom I would have no data to analyze), as well as the neurologists, neurosurgeons, technicians, and funding agencies who made this work possible.

I would not have made it through my graduate training without the love, support, and proofreading abilities of my wife, Talia. Thanks for sticking by my side, and for making my life interesting and fun! Thanks also to my parents, Dr. Karen Rothman and Dr. Gordon Manning, and brother, Joshua Manning, for their encouragement, love, and scientific feedback. Finally, thanks to my cats, Macaroni and Mr. Tibbles, for keeping me company during the early hours of the morning when much of this work was performed.

ABSTRACT

ACQUISITION, STORAGE, AND RETRIEVAL IN DIGITAL AND BIOLOGICAL BRAINS

Jeremy R. Manning

Supervisor: Michael J. Kahana, Ph.D.

My work examines how the brain acquires, stores, and retrieves information. I first present a theoretical model of the retina, and use the model to explore how the design of sensory systems affects our ability to make inferences about the physical world. I next present three analyses of electrocorticographic recordings taken as human neurosurgical patients participated in experimental cognitive tasks. In the first analysis, I measure the relation between single-neuron spiking and local field potentials, which reflect the aggregate activity of large populations of neurons. In the second analysis, I ask how the brain represents the meanings of individual words as they are studied and remembered. In the third electrocorticographic analysis, I address the question of how our brains retrieve memories of past experiences.

Contents

Acknowledgments	iv
Abstract	vi
Contents	vii
List of tables	x
List of figures	xi
1 Introduction	1
1.1 Intracranial recordings from human epilepsy patients	2
1.2 Overview	5
2 Optimal design of photoreceptor mosaics: why we do not see in color at night	7
2.1 Abstract	7
2.2 Introduction	8
2.3 General formulation	13
2.4 Simplified formulation	17
2.5 Results	23
2.6 Discussion	37

3	Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans	51
3.1	Abstract	51
3.2	Introduction	52
3.3	Methods	54
3.4	Results	59
3.5	Discussion	68
3.6	Supplemental materials	73
4	Temporal and frontal networks reveal how conceptual memories are organized	78
4.1	Abstract	78
4.2	Narrative	79
4.3	Methods	88
4.4	Supplemental materials	90
5	Oscillatory patterns in temporal lobe reveal context reinstatement during memory search	95
5.1	Abstract	95
5.2	Introduction	96
5.3	Results	98
5.4	Discussion	105
5.5	Methods	107
5.6	Supplemental materials	109
6	General discussion	122
6.1	Contributions of this dissertation	122

6.2	Future directions: Chapter 2	123
6.3	Future directions: Chapter 3	124
6.4	Future directions: Chapters 4 and 5	127
6.5	Concluding remarks	133
References		134

List of tables

3.1	Summary of observed neurons	77
4.1	Bandpass filters employed by our collaborating hospitals	90
4.2	Patient and task information	94
5.1	Patient and task information	111
5.2	Bandpass filters employed by our collaborating hospitals	112

List of figures

1.1	Electrodes used for human intracranial recordings	3
2.1	Three simple render matrices with varying optical blur and spectral sensitivity	21
2.2	Optimized error (O_F^{opt}) for equal signal-to-noise case	27
2.3	Comparison of dichromatic and monochromatic vision for equal photoreceptor signal-to-noise and high color and spatial correlations	29
2.4	Comparison of dichromatic and monochromatic retinæ for varying color and spatial correlations	30
2.5	Optimized error (O_F^{opt}) for unequal dark noise case	32
2.6	Comparison of dichromatic and monochromatic vision for unequal photopigment dark noise and high color and spatial correlations . .	33
2.7	Best arrangement as a function of illumination intensity (ω)	34
2.8	Effect of spectral sensitivity	37
3.1	LFP power and neuronal firing time series	61
3.2	A representative neuron exhibiting a positive correlation between firing rate and broadband LFP power	62
3.3	Categories of observed neurons	64
3.4	Broadband-shift neurons throughout the brain	66

3.5	Percentage of positive broadband-shift neurons observed in each brain region	67
3.6	LFP components that predict firing rate	68
3.7	Average power spectra for different subsets of neurons	69
3.8	Schematic of our analysis framework	76
4.1	Illustration of behavioral and electrophysiological methods	81
4.2	Regions of interest (ROIs)	83
4.3	Neural activity predicts clustering during recall	84
4.4	Detailed region of interest results	93
5.1	Illustration of behavioral and electrophysiological methods	99
5.2	Evolution of ECoG activity as participants study lists of words	100
5.3	Predicted neural similarity as a function of lag according to three models	102
5.4	A neural signature of temporal context reinstatement	103
5.5	Detailed simulation results	116
5.6	Evidence for context reinstatement in the temporal lobe	119
5.7	Mean contributions of each frequency band to selected features	120
6.1	Putting this dissertation into context	123
6.2	Applying feature selection to principal components	129

Chapter 1

Introduction

Weighing in at under 1.5 kilograms and containing a hundred billion neurons, sixty trillion synapses, and more than 170 megameters of axon¹ (Kandel, Schwartz, & Jessell, 2000; Marner, Nyengaard, Tang, & Pakkenberg, 2003), the human brain is by far the most sophisticated computing device our species has ever produced — and perhaps ever will. Throughout history, our best and brightest philosophers, psychologists, and neuroscientists have agonized over how our brains learn about and interact with our physical world, and how we retrieve learned information and memories of our past experiences. In this dissertation, I use computational modeling and analyses of human brain recordings to study these questions. In this chapter I briefly discuss the intracranial recording methods used in Chapters 3 – 5 and provide a general overview of this dissertation.

¹For comparison, the Earth's circumference around the equator measures just over 40 megameters, and the distance from the Earth to the moon is approximately 380 megameters.

1.1 Intracranial recordings from human epilepsy patients

The physiological data I analyze in Chapters 3–5 come from neurosurgical patients with drug-resistant epilepsy. In order to identify epileptogenic tissue for potential subsequent resection, clinical teams first implant electrodes in the brains of these patients. Over the course of a patient’s hospital stay, neurologists and technicians analyze the electrical activity recorded from these electrodes as the patients have seizures, in order to localize the epileptic foci. Although seizure frequency and duration of hospital stay varies from patient to patient, the patients are often confined to their hospital beds for several weeks. During this time, many patients generously volunteer to participate in our studies.

Recordings are made from three types of electrodes (Fig. 1.1). In Chapter 3, I analyze previously collected data from microwire electrodes (Figs. 1.1A–C), which extend from the tips of implanted depth electrodes, and lie in contact with neurons’ somata. These electrodes, which measure $40\ \mu\text{m}$ in diameter, record both local field potentials, which reflect the aggregate activities of the local population of neurons within approximately $140\ \mu\text{m}$ of the tip of the electrodes (Buzsáki, 2004), and action potentials from the directly adjacent neurons (Fig. 1.1C). A single microelectrode can record action potentials simultaneously from multiple neurons; these action potentials must then be isolated and assigned to individual neurons (Quiroga, Nadasdy, & Ben-Shaul, 2004). In Chapters 4 and 5, I analyze data from depth electrodes (Figs. 1.1A,B) and surface electrodes (Figs. 1.1D,E). The exposed recording surfaces of the depth electrodes measure 1.3 mm in diameter, while the recording surfaces of the surface electrodes measure 2.3 mm in diameter. These electrodes record from much larger areas than the microwire electrodes, capturing

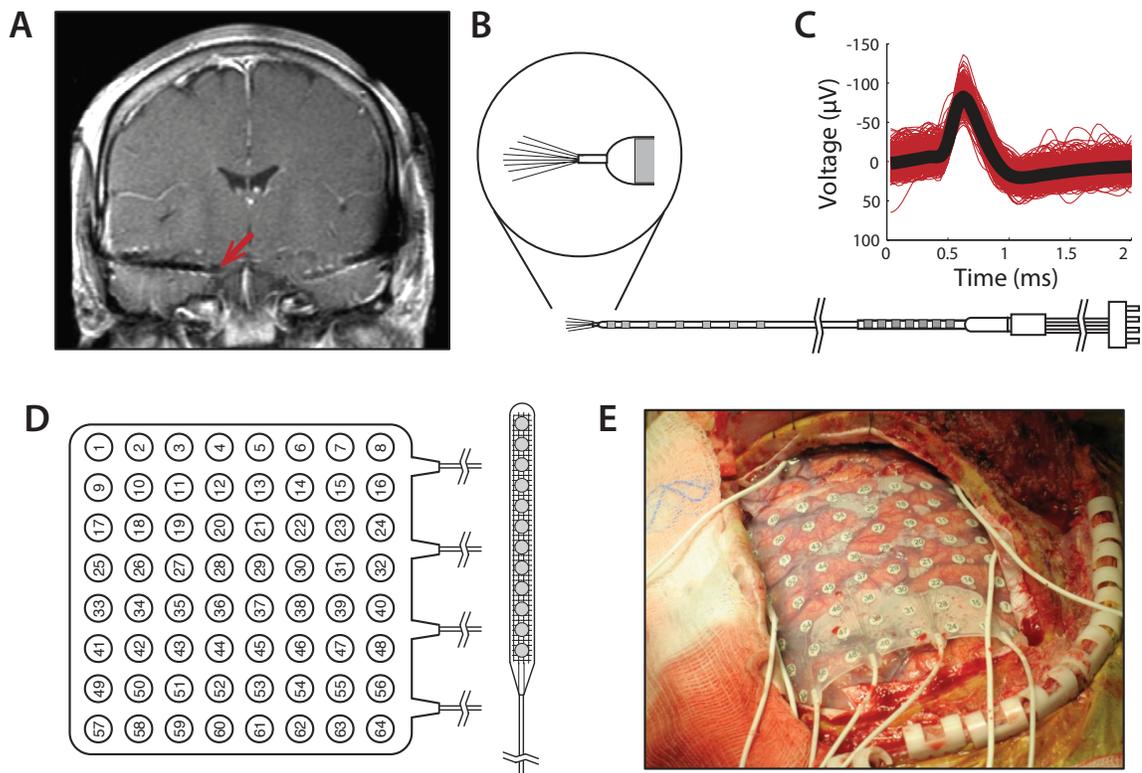


Figure 1.1. Electrodes used for human intracranial recordings. **A.** This panel shows a magnetic resonance image (MRI) of a patients' brain. The red arrow indicates the location of the tip of a depth electrode implanted in the right hippocampus. **B.** This panel shows a schematic of a depth electrode with eight microwire electrodes extending from its tip. Standard depth electrodes look similar, but do not have the microwire extensions (not shown). **C.** This panel shows action potential waveforms isolated from a microwire recording. The individual waveforms (shown in red) have been aligned to lie on top of one another. The black line indicates the average waveform. **D.** This panel shows a schematic of two surface electrodes, arranged in an 8×8 "grid" configuration (left) and a 1×12 "strip" configuration (right). Electrodes in both configurations are spaced at 1 cm intervals. **E.** Here an 8×8 grid of electrodes has been placed on the surface of a patient's brain. Panels A and C are courtesy of Joshua Jacobs. Panels B and D are courtesy of Ad-Tech, Inc. Panel E is courtesy of Brian Litt.

population activity from approximately 500,000 neurons (K. Miller, Zanos, Fetz, den Nijs, & Ojemann, 2009).

Whereas source localization poses serious challenges to scalp electroencephalography (EEG), intracranial electrodes sample from much smaller volumes (Nunez et al., 1997), are not subject to distortions produced by the skull (Pfurtscheller & Cooper, 1975), and are relatively impervious to confounds produced by eye movements, blinks, and vocalizations. These factors all contribute to the high signal-to-noise ratio characteristic of intracranial recordings. We synchronize the behavioral experiments and physiological recordings by sending a train of optically isolated pulses from the testing computer to the recording system and computing a linear regression across all synchronization events. With this technique, the synchronization error is typically smaller than the sampling resolution of the recorded data.

Despite the high quality of human intracranial recordings, there are several factors one should consider when interpreting the results of any intracranial study of human epilepsy patients, including those I report in Chapters 3 – 5. First, whereas in animal studies electrodes are placed according to researchers' needs, the placements of implanted human electrodes are determined solely by clinical teams with the goal of localizing the seizure focus in order to ensure the best possible outcome for the patient. This means that, for some patients, the brain areas most relevant to a particular research question may receive little or no electrode coverage. In order to obtain adequate coverage of all relevant brain areas, I analyzed data from many patients. A second concern is that medications or recent seizures might change the electrophysiological properties of the brain. For this reason, we refrained from collecting data while the patients were on high dosages of pain medications or anti-epileptic drugs, or during the 6 hour period following any

clinically significant seizure. A third issue is that the brain is known to re-wire itself to compensate for damage, including damage caused by epilepsy (Ribak & Dashtipour, 2002), which could lead to cognitive remapping. While I cannot control for cognitive remapping that may have occurred in individual participants, I have averaged each of my anatomical analyses over many patients; thus results due to remapping in one patient will average out in the population analyses. A fourth concern is that severe epilepsy can lead to cognitive impairment. To address this issue, I analyzed data only from patients with scores on the Wechsler Intelligence and Wechsler Memory Scales within 1.5 standard deviations of the mean for their age group.

1.2 Overview

In the next chapter, I present a theoretical model of the retina, and use the model to explore how the way in which photoreceptors are arranged can affect the ability of the visual system to reconstruct visual stimuli from the observed receptor responses. Although the model is framed as a means of examining the visual system, in principle the core features of this approach should apply to any sensory system, biological or digital. The framework suggests a means of designing sensory systems that can best make inferences about the world. This work also provides a context for interpreting why biological sensory systems may have evolved as they did.

In Chapter 3, I present an analysis of intracranial recordings taken from human neurosurgical patients as they performed a virtual navigation task. This work is intended to elucidate the relation between action potentials generated by individual neurons and the behaviors of large populations of neurons (local field potentials). I

view this work as furthering our understanding of the nature of the brain's internal language and of its electrical properties.

In Chapter 4, I present an analysis of recordings taken from a different population of human neurosurgical patients as they studied and recalled lists of words. In this work I attempt to elucidate how the brain represents the meanings of individual words as they are actively manipulated by the mind. Although a number of studies have found similarities in how different people's brains represent words (Shinkareva et al., 2008; Mitchell et al., 2008), my work shows that individual differences in these neural representations predict the order in which people will recall the words — a measure of how the studied words are organized in and retrieved from memory.

In Chapter 5, I present an analysis of neurosurgical recordings during the same memory task. In this analysis, I seek to understand how we probe our brains for memories of experienced events. This work shows that our brains contain a representation of the temporal contexts in which we experience each event, much like a clock marking time along our autobiographical timeline. The state of this gradually evolving context representation becomes associated with each event we experience. When we retrieve an event from our memory, the context representation associated with that event is partially reinstated along with the neural representation of the event itself.

Taken together, this body of work takes us one small step closer to understanding how the brain acquires, stores, and retrieves information. In the concluding chapter, I propose future directions for these studies.

Chapter 2

Optimal design of photoreceptor mosaics: why we do not see in color at night

Jeremy R. Manning & David H. Brainard. *Visual Neuroscience*, 26, 5–19.

2.1 Abstract

While color vision mediated by rod photoreceptors in dim light is possible (Kelber & Roth, 2006), most animals, including humans, do not see in color at night. This is because their retinae contain only a single class of rod photoreceptors. Many of these same animals have daylight color vision, mediated by multiple classes of cone photoreceptors. We develop a general formulation, based on Bayesian decision theory, to evaluate the efficacy of various retinal photoreceptor mosaics. The formulation evaluates each mosaic under the assumption that its output is processed to optimally estimate the image. It also explicitly takes into account

the statistics of the environmental image ensemble. Using the general formulation, we consider the tradeoff between monochromatic and dichromatic retinal designs as a function of overall illuminant intensity. We are able to demonstrate a set of assumptions under which the prevalent biological pattern represents optimal processing. These assumptions include an image ensemble characterized by high correlations between image intensities at nearby locations, as well as high correlations between intensities in different wavelength bands. They also include a constraint on receptor photopigment biophysics and/or the information carried by different wavelengths that produces an asymmetry in the signal-to-noise ratio of the output of different receptor classes. Our results thus provide an optimality explanation for the evolution of color vision for daylight conditions and monochromatic vision for nighttime conditions. An additional result from our calculations is that regular spatial interleaving of two receptor classes in a dichromatic retina yields performance superior to that of a retina where receptors of the same class are clumped together.

2.2 Introduction

The vertebrate retina contains two broadly distinguished classes of photoreceptors, rods and cones. Rods are characterized by low noise (Barlow, 1956; Baylor, Nunn, & Schnapf, 1984; Schneeweis & Schnapf, 2000) and are effective at low light levels when photons are scarce. On the other hand, they saturate more easily than cones and their contribution to vision diminishes at daylight light levels (Aguilar & Stiles, 1954; Tamura, Nakatani, & Yau, 1989; Demontis, Bisti, & Cervetto, 1993; Burns & Arshavsky, 2005; Yin, Smith, Sterling, & Brainard, 2006). Cones are noisier than rods (Schnapf, Nunn, Meister, & Baylor, 1990; Rieke & Baylor, 2000; Fu, Kefalov,

Luo, Xue, & Yau, 2008), but operate without saturation at much higher light levels (Schneeweis & Schnapf, 1999). Cones are therefore useful for daylight vision, when there is plenty of light. Across species, the relative numbers of rods and cones vary enormously, with nocturnal animals generally having a higher rod–cone ratio than diurnal animals (Walls, 1942).

Within the rod and cone systems, there is additional across-species variability. Most vertebrates have multiple classes of cones, distinguished primarily by different spectral sensitivities. The presence of multiple cone classes enables color vision, as the relative responses of cones of different classes provides information about the relative spectrum of the incident light. Across species that have multiple classes of cones, there are additional variations. These include the number of distinct classes of cones present (Walls, 1942; G. H. Jacobs, 1981), the spectral sensitivities of the individual cone classes (Bowmaker, 1991; G. H. Jacobs, 1996; G. H. Jacobs & Rowe, 2004), and the pattern of how the cones are arranged in the retinal mosaic (Scholes, 1975; Wassle & Riemann, 1978; Bowmaker & Kunz, 1987; Mollon & Bowmaker, 1992; Hofer, Carroll, Neitz, Neitz, & Williams, 2005). A small minority of mammalian species that operate in low light conditions have only one spectral class of cone (G. H. Jacobs, 1996), and are thus monochromats under cone-mediated viewing conditions.

In contrast to their cone vision, most vertebrate retinæ have only a single class of rod. Vision mediated by a single class of rod is monochromatic, since variation in image intensity is perfectly confounded with variation in image relative spectrum. The restriction to a single class of rod is not due to any fundamental biophysical constraint, as a few species (e.g. the Nocturnal Hawkmoth, *Deilephila elpenor*) do have retinæ with multiple classes of rods, and have been shown behaviorally to have color vision at rod-mediated light levels (Kelber & Roth, 2006).

Previous authors have speculated about why color vision is rare at night. Walls (1942), for example, indicates that color vision may only be useful under conditions where spatial and contrast acuity are high and that it would not be useful when photons are scarce. Rushton (1962) makes a similar statement. To our reading, however, neither of these authors offer reasoning to support their assertions. Moreover, it is now known that some plants and bacteria detect light with low spatial resolution but none-the-less have color vision (see Skorupski & Chittka, 2008), in apparent contradiction to Walls' and Rushton's hypothesis. Also note that human rod-mediated visual acuity at approximately 5 to 10 degrees eccentricity exceeds cone-mediated visual acuity at 40 to 50 degrees eccentricity (see Weymouth, 1958 who replots data from Fick, 1898), also suggesting that factors other than those that limit resolution contribute to whether a retina should be monochromatic or not.

Land and Osorio (2003) articulate what we take, based on an informal polling of our colleagues, to be the standard intuition on this topic. This intuition is based on the observation that color vision depends on the difference between the responses of two receptor classes, and the magnitude of this difference signal will on average be smaller than that of a luminance signal that arises from the sum of responses of the two classes or from a single receptor class.¹ At the same time, the net effect of photon and dark noise is the same for both difference and summed signals. Thus the idea is that color vision becomes less useful at low light levels because overall signal-to-noise drops, and that noise swamps the output of an opponent color channel before it swamps that of a luminance channel. A related idea was formalized by van Hateren (1993), who showed that as signal-to-noise drops, the optimal post-receptoral processing favors a luminance channel over a chromatic

¹This statement holds as long as the responses of the two receptor classes are positively correlated.

channel.

Note, however, that neither the standard intuition nor van Hateren's (1993) calculations lead immediately to the conclusion that monochromatic vision is superior to color vision at low signal-to-noise. To do so requires an analysis that shows not just that the value of color vision decreases with light level, but that a change to monochromatic vision leads to improved performance. That is, the fact that a luminance channel carries more information than an opponent color channel tells us neither that a monochromatic retina is superior to one with multiple receptor types, nor that it is not advantageous to add a color opponent channel to a luminance channel as part of post-receptoral processing.

Barlow (1957) argued that the spontaneous photopigment isomerization rate of photoreceptors might depend strongly on their wavelength of maximal sensitivity, with maximal sensitivity at shorter wavelengths leading to lower noise. This effect would be expected to occur for fundamental biophysical reasons: short-wavelength quanta have more energy than long-wavelength quanta, and this in turn would allow a photopigment sensitive to shorter wavelengths to have a higher energy barrier for isomerization and be more resistant to thermal fluctuations (Platt, 1956). Barlow argued that the overall shift in human visual sensitivity towards shorter wavelengths between cone- and rod-mediated vision (the Purkinje shift) maximizes absolute sensitivity by minimizing thermal noise. Barlow did not explicitly address tradeoffs between color and monochromatic vision. However, his observation that the dependence of dark noise on spectral sensitivity has important implications for differences between daylight and nighttime vision plays a key role in the analysis we present below. Other closely related antecedents to the current chapter are a treatment of the design of trichromatic mosaics (Garrigan et al., 2006, 2010) and a treatment of optimal choice of cone spectral sensitivities (Lewis & Zhaoping, 2006);

these authors used an information theoretic approach and emphasize the central role played by asymmetries in the information available at different wavelengths. We return in the discussion to review measurements of photoreceptor thermal noise.

Here we assess the efficacy of various choices of the design of the photoreceptor mosaic in the context of the vertebrate eye. We formulate the design question in terms of an explicit model that defines optimal performance for a well-defined statistical ensemble of stimuli. Our model allows us to assess how well an ideal observer could estimate the incident image from the photoreceptor responses, and we examine how this performance measure varies with different aspects of eye design. Although our formulation is very general, we focus on the performance tradeoff between dichromatic (two receptor classes) and monochromatic (one receptor class) vision, and how this tradeoff interacts with overall light level. To address this question, however, the work also touches on other aspects of mosaic design, including the choice of photopigment spectral sensitivity and the packing arrangement of the mosaic.

In the next section we introduce a general formulation. To make computational progress, however, we then make a number of simplifying assumptions. These include restricting attention to a model environment in which there are just two discrete wavelengths, and where all images are characterized by the power in each wavelength at discrete image locations on a line (one spatial dimension.) We consider image ensembles characterized by Gaussian distributions, evaluate image estimation using a mean squared error criterion, and do not take into account the energetic cost of computation. Although *a priori* it might appear that these simplifying choices reduce the complexity of the model too severely, the simplified model is surprisingly rich and studying it leads to interesting insight.

2.3 General formulation

We start with a general formulation. We model the visual environment as an ideal image, which specifies the intensity of light incident on the eye as a function of two angular spatial dimensions (x and y) and one spectral dimension (λ). For simplicity, we consider the case where each of these dimensions has been finely discretized. Thus the ideal image is specified as $I(x_i, y_j, \lambda_k)$. Note that in most three-dimensional scenes, the retinal image is not formed by viewing a planar ideal image. Rather, light reflects to the eye from objects located at various distances. The ideal image should be conceived as one that would produce the same retinal image as the actual illuminated objects in the scene.

Light from the ideal image passes through an optical apparatus consisting of a cornea, pupil, and lens. Refraction at the interfaces between the optical elements creates an image on the retina. Aberrations in the optics combine with diffraction to blur the retinal image; the blurring may be characterized by the point spread function of the optics. In general, the point spread function varies with wavelength because of chromatic aberration. It also depends on the pupil area \mathbb{A} . In the typical regime where effects of optical aberrations dominate effects of diffraction, increasing pupil size both increases the intensity of the retinal image and the degree to which it is blurred.

We denote the point spread function by $P_{\mathbb{A}}(x_i, y_j, \lambda_k)$, where \mathbb{A} is the pupil area. It specifies the spread of light in the retinal image from a point source, and we assume it to be independent of the location of that source. The subscript \mathbb{A} denotes the fact that the shape of the point spread function depends on pupil area. The retinal image may be computed as the convolution of the ideal image and the point

spread function:

$$R(x_i, y_j, \lambda_k) = \mathbb{A} \left[I(x_i, y_j, \lambda_k) * P_{\mathbb{A}}(x_i, y_j, \lambda_k) \right].$$

By convention, $P_{\mathbb{A}}(x_i, y_j, \lambda_k)$ is normalized to have unit volume; the effect of pupil area on overall light level is accounted for explicitly in the equation above.

The retinal image is sampled by an interleaved array of photoreceptors, with one photoreceptor at each location. The photopigment isomerization rates of each photoreceptor are the information about the image available to the visual system. We model the relation between the retinal image and the mean isomerization rate of a single photoreceptor of class l centered at location (x_m, y_n) :

$$\bar{\mu}_l = \sum_k ([R(x_i, y_j, \lambda_k) * Y_l(x_i, y_j)]|_{(x_m, y_n)}) S_l(\lambda_k) + d_l.$$

Here $Y_l(x_i, y_j)$ represents the acceptance aperture of the l^{th} photoreceptor class, $S_l(\lambda_k)$ represents the spectral sensitivity of the l^{th} class, and d_l represents the mean number of spontaneous isomerizations per second of the l^{th} class. The actual number of isomerizations in response to an image shown for a duration Δt is distributed as a Poisson process with mean $\bar{\mu}_l \Delta t$.

We take the task of visual processing as estimating the ideal image from the noisy array of photoreceptor responses. At this juncture, it is convenient to express the ideal image, which consists of a finite array of intensities, as a vector \vec{i} . The estimator may then be expressed as

$$\hat{i} = F(\vec{r}),$$

where \hat{i} is the estimate of \vec{i} and \vec{r} is a vector representing the entire array of noisy

responses. The appropriate estimator $F()$ depends on the exact design properties of the model visual system, on the statistical ensemble of images that will be encountered, on a loss function that describes how bad it is to approximate any given ideal image with any particular estimate, and on a decision about how loss should be aggregated over the statistical image ensemble. Once choices for each of these factors are made, however, the estimator $F()$ can be chosen to optimize aggregate performance. Moreover, we can use optimized aggregate performance as a metric to ask which eye design parameters are best.

To construct an optimal estimator, we turn to Bayesian decision theory (Blackwell & Girschick, 1954; Berger, 1985; Gelman, Carlin, Stern, & Rubin, 2004). We start by expressing the relation between the ideal image and noisy response vector as a probability distribution $P(\vec{r}|\vec{i})$. This is called the *likelihood* and represents how probable any response vector is, conditional on knowing the ideal image.

The second step is to express the statistical structure of the environment as a probability distribution, $P(\vec{i})$. This is called the *prior*. The prior captures what is known or assumed about stimuli in the model environment (here the ideal image), independent of the observed photoreceptor responses. *Bayes' rule* then yields the *posterior*

$$P(\vec{i}|\vec{r}) = \mathbb{C}P(\vec{r}|\vec{i})P(\vec{i}),$$

where \mathbb{C} is a normalizing constant. The posterior tells us how probable any particular image is, given the observed photoreceptor responses.

To obtain a particular estimate \hat{i} from the posterior, we need to specify a *loss function* $L(\hat{i}, \vec{i})$. This function provides the cost of choosing any estimate \hat{i} when the actual image is \vec{i} . Once the loss function is given, we can define the *Bayes risk*

(Berger, 1985) of an estimator $F(\vec{r})$ as

$$R_F = \int_{\vec{i}} \left[\int_{\vec{r}} L(F(\vec{r}), \vec{i}) P(\vec{r}|\vec{i}) d\vec{r} \right] P(\vec{i}) d\vec{i}.$$

The Bayes risk provides the expected loss associated with repeated applications of the estimator, when the images leading to the responses are drawn from the prior distribution $P(\vec{i})$. To see this, note that the inner integral takes the expectation over responses \vec{r} given the image \vec{i} (according to the likelihood), while the outer integral takes the expectation over images (according to the prior). In the absence of any computational constraints, a widely used optimality principle is to choose the estimator $F()$ that minimizes R_F . In the development below, however, we will use the notation O_F to denote the quantity to be minimized, with $O_F = R_F$. Although not immediately relevant, this change in notation reminds us that there are alternative possible formulations of optimality. For example, in the discussion we briefly treat the possibility of including energetic cost in addition to performance.

Given a specified likelihood, prior, and loss function, denote by $F^{opt}()$ an estimator that minimizes O_F , and by O_F^{opt} the corresponding minimized value of O_F . We refer to O_F^{opt} as the *optimized error*. The prior and loss functions describe the features of the environment and organism that define the visual system design problem, while the likelihood function describes the visual system's solution. That is, if we hold $P(\vec{i})$ and $L(\hat{i}, \vec{i})$ constant, we can regard O_F^{opt} as a function of the likelihood. The optimal design is then the one corresponding to the likelihood $P(\vec{r}|\vec{i})$ that minimizes $O_F^{opt} [P(\vec{r}|\vec{i})]$. This is the approach we implement below. At this level of generality, it closely parallels the approach used by Srinivasan, Laughlin, and Dubs (1982) to consider the receptive fields of ganglion cells, as well as that developed recently by Levin, Durand, and Freeman (2008) in the context of evaluating the design of

digital cameras.

2.4 Simplified formulation

To apply our general framework to mosaic design, and in particular to the tradeoff between color and monochromatic vision, we make concrete simplifying assumptions. When the simplified version of the general model is able to predict observed biological design features, we gain insight about what properties of the environment are likely to have driven the evolution of these features. When the simplified version fails to predict the biology, we gain insight by considering ways in which the model is over-simplified or, alternatively, considering ways in which evolution has failed to arrive at an optimal design.

2.4.1 Image representation and prior

We start with one spatial dimension with discretized spatial locations and wavelengths. Multi-wavelength-band images of this sort may be represented as column vectors:

$$\vec{i} = [m_{x_1, \lambda_1}, m_{x_2, \lambda_1}, \dots, m_{x_{\mathbb{I}}, \lambda_1}, m_{x_1, \lambda_2}, m_{x_2, \lambda_2}, \dots, m_{x_{\mathbb{I}}, \lambda_2}, \dots, m_{x_1, \lambda_{\mathbb{K}}}, m_{x_2, \lambda_{\mathbb{K}}}, \dots, m_{x_{\mathbb{I}}, \lambda_{\mathbb{K}}}]^T,$$

where m_{x_i, λ_k} is the intensity of wavelength λ_k at pixel i . We have $1 \leq i \leq \mathbb{I}$ and $\lambda_1 \leq k \leq \mathbb{K}$. For our calculations, we used 36 spatial locations and 2 wavelengths, so $\mathbb{I} = 36$ and $\mathbb{K} = 2$.

Natural images are characterized by high correlations between the intensities at neighboring spatial locations (Pratt, 1978; Field, 1987; Burton & Moorehead, 1987; Simoncelli, 2005), and between intensities in the same location but at different

wavelengths (Burton & Moorehead, 1987; Ruderman, Cronin, & Chiao, 1998).² These statistical features of natural images may be captured using a multivariate Gaussian distribution over multi-wavelength–band image vectors \vec{i} :

$$P(\vec{i}) = N(\vec{\mu}_{prior}, \mathbf{K}_{prior}),$$

where $\vec{\mu}_{prior}$ and \mathbf{K}_{prior} are the mean and covariance matrix of the distribution, respectively. We required that $P(\vec{i})$ be separable between space and wavelength, so that the spatial statistics were the same in each waveband, up to multiplicative scaling. In addition, we viewed the distribution $P(\vec{i})$ as induced when a spatially uniform illuminant reflected from a set of surfaces, which allowed us to parametrically vary the intensity of the illuminant. For illuminant intensity ω and mean surface reflectances \bar{m}_{λ_1} and \bar{m}_{λ_2} at the two wavelengths, $\vec{\mu}_{prior}$ was given by

$$\vec{\mu}_{prior} = \omega \times [\bar{m}_{\lambda_1} \times O(1, \mathbb{I}), \bar{m}_{\lambda_2} \times O(1, \mathbb{I})]^T,$$

where $O(a, b)$ is a matrix of 1’s with a rows and b columns. The covariance matrix \mathbf{K}_{prior} was given by the Kronecker product $\mathbf{K}_S \otimes \mathbf{K}_C$, where \mathbf{K}_S was the covariance matrix in the spatial domain and \mathbf{K}_C was the covariance matrix in the wavelength domain. We constructed \mathbf{K}_S to represent a first-order Markov process (Pratt, 1978), so that

$$\mathbf{K}_S(u, v) = r_S^{(u-v)},$$

²The literature generally shows not correlations across wavelength, but rather correlations across responses of different classes of cones. The reported correlations across cone classes arise both because of correlations across wavelength and because of the broadband spectral sensitivity of the cones. We have verified, using the data of Ruderman et al. (1998), that the correlations across wavelength are in fact strong. For two narrowband images separated by 50 nm, the average correlation is about 0.9. The correlation remains above 0.8 for wavelength separations of 200 nm.

where $\mathbf{K}_S(u, v)$ represents the entry in the the u^{th} row and v^{th} column of \mathbf{K}_S , and where r_S is the within-wavelength correlation between image intensities at neighboring locations. The use of a Markov model allows us to adjust the correlational structure of the image ensemble with a single parameter, and makes exploring how this structure affects optimal design more practical. We defined \mathbf{K}_C as

$$\mathbf{K}_C = \omega^2 \times \begin{bmatrix} \sigma_{\lambda_1}^2 & r_C \sigma_{\lambda_1} \sigma_{\lambda_2} \\ r_C \sigma_{\lambda_1} \sigma_{\lambda_2} & \sigma_{\lambda_2}^2 \end{bmatrix},$$

where σ_{λ_1} is the standard deviation of reflectances at the first wavelength, σ_{λ_2} is the standard deviation of reflectances at the second wavelength, and r_C is the correlation (taken over locations) of the reflectances at the two wavelengths. This formulation for \mathbf{K}_C could be easily generalized to $\mathbb{K} > 2$ wavelengths by treating it as representing a first-order Markov process over wavelength and regarding r_C as the correlation between neighboring wavelengths.

2.4.2 Image formation and likelihood

Given \mathbb{I} pixels and \mathbb{K} wavelengths, a *render matrix* \mathbf{R} determines the responses of each of the $\mathbb{L} \leq \mathbb{I}$ photoreceptors in the modeled retina. The render matrix has \mathbb{L} rows and $(\mathbb{I} \times \mathbb{K})$ columns, where the value of the element in the u^{th} row and v^{th} column is the responsiveness of photoreceptor u to light at pixel $((v - 1)\% \mathbb{I}) + 1$ of wavelength $\lceil \frac{v}{\mathbb{K}} \rceil$. Here $\%$ represents the modulus operator. To illustrate the idea, consider a retina with 4 evenly spaced photoreceptors that alternate between two classes, where each class is sensitive to only one wavelength.

For illustrative purposes we will assume that there is no optical blurring and that each photoreceptor is sensitive to light from only one spatial location and one

wavelength. Then, for a environment where there are 8 spatial locations and 2 wavelengths, we would have

$$\mathbf{R} = \left[\begin{array}{cccccccc|cccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right].$$

The first 8 columns of this matrix represent the sensitivity of each photoreceptor to the first wavelength while the second 8 represent the sensitivity of each photoreceptor to the second wavelength. Rows 1 and 3 thus describe the first class of photoreceptors, and their spatial location is indicated by the column containing the 1. Rows 2 and 4 represent the second class of photoreceptor in a similar fashion. This render matrix is shown pictorially in Figure 2.1A. Figure 2.1B illustrates the addition of optical blurring to our four photoreceptor example. Blurring was computed by centering a Gaussian at each photoreceptor’s position. Figure 2.1C illustrates a render matrix for a retina containing photoreceptors sensitive to both wavelengths, but to varying degrees. The general formulation of the render matrix allows arbitrary spatial and wavelength sensitivity to be specified.

Given a render matrix \mathbf{R} and a draw from the stimulus distribution \vec{i} , the photoreceptor responses to \vec{i} are given by the \mathbb{L} -element column vector \vec{r} :

$$\vec{r} = \mathbf{R}\vec{i} + \epsilon,$$

where ϵ represents photoreceptor noise and was a draw from the Gaussian distribution $N(\vec{\mu}_{noise}, \mathbf{K}_{noise})$. For our simulations, we used an \mathbb{L} -dimensional column vector as $\vec{\mu}_{noise}$, where the u^{th} entry is given by the mean dark noise d_l of the photoreceptor

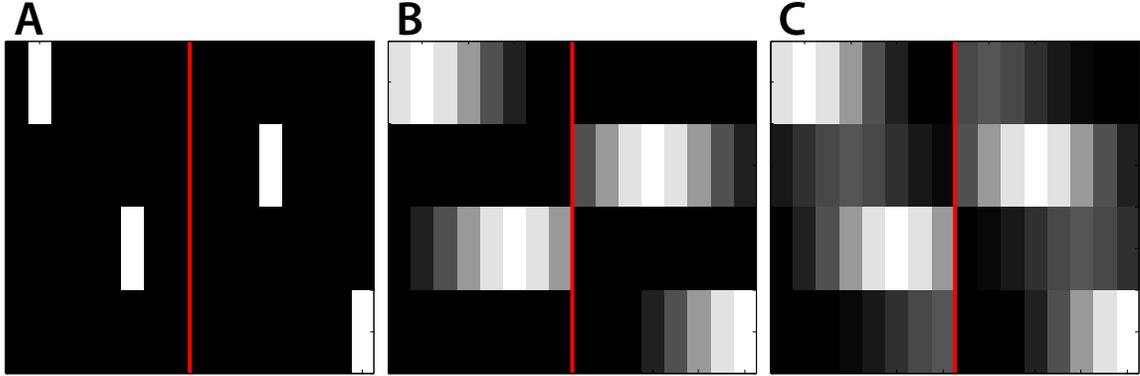


Figure 2.1. Three simple render matrices with varying optical blur and spectral sensitivity. In these examples, we show four photoreceptors, eight pixels, and two wavelengths. Each row corresponds to one photoreceptor. Columns to the left of the red line correspond to the first wavelength, while columns to the right correspond to the second wavelength. Within a wavelength, columns correspond to spatial locations. Darker areas correspond to lower sensitivity, while lighter areas correspond to higher sensitivity. **A.** This render matrix corresponds to the example given numerically in the text, with each photoreceptor sensitive to incident light at a single punctate location and at a single wavelength. **B.** Effect of optical blur. The standard deviation of the Gaussian used to compute optical blur was set to 2 pixels, and each photoreceptor is sensitive to one of the two wavelengths. **C.** Spatial blur is the same as in B, but photoreceptors are sensitive to 75% of one wavelength and 25% of the other. Each matrix shown was normalized to have the same maximum intensity for display.

represented by the u^{th} row of the render matrix. The noise covariance matrix $\mathbf{K}_{\text{noise}}$ is an $\mathbb{L} \times \mathbb{L}$ diagonal matrix where the diagonal is given by $\omega \times \bar{m}_{\lambda_k} + d_l$. With this choice of noise, the likelihood function is then given by the Gaussian

$$P(\vec{r}|\vec{i}) = N(\mathbf{R}\vec{i} + \vec{\mu}_{\text{noise}}, \mathbf{K}_{\text{noise}}).$$

We used a Gaussian approximation to the more realistic Poisson noise distribution so that we could leverage analytic results (see below) that apply for the Gaussian case.

2.4.3 Optimal estimator

Given the prior and likelihood, the optimal estimator depends on the loss function. The appropriate loss function for biological vision is almost certainly a highly complicated function of the image and the estimate, as some estimation errors will have much more serious consequences for survival and reproduction than others. Nonetheless, for practical reasons, we simplify in the remainder of the chapter by restricting attention to a simple squared error loss function

$$L(\hat{i}, \vec{i}) = \|\vec{i} - \hat{i}\|^2.$$

Although this loss function does not fully capture the biologically relevant aspects of estimation error, it has the feature that the estimator that minimizes the corresponding R_F is the mean of the posterior distribution.

Our prior and likelihood distributions are multivariate Gaussian with known mean and covariance. Therefore, the posterior is also multivariate Gaussian, and its mean and covariance may be computed in closed form (Gelman et al., 2004). The estimated image \hat{i} is given by:

$$\hat{i} = \vec{\mu}_{posterior} = \mathbf{F}\vec{r} + \vec{i}_0, \text{ where}$$

$$\mathbf{F} = \mathbf{K}_{prior} \mathbf{R}^T (\mathbf{R} \mathbf{K}_{prior} \mathbf{R}^T + \mathbf{K}_{noise})^{-1} \text{ and}$$

$$\vec{i}_0 = \vec{\mu}_{prior} - \mathbf{F} \mathbf{R} \vec{\mu}_{prior} - \mathbf{F} \vec{\mu}_{noise}.$$

Moreover, for our choice of squared error loss, $O_F^{opt}[P(\vec{r}|\vec{i})]$ is given by the trace of

the posterior covariance matrix:

$$O_F^{opt} [P(\vec{r}|\vec{i})] = \text{trace}(\mathbf{K}_{posterior}), \text{ where}$$

$$\mathbf{K}_{posterior} = (\mathbf{K}_{prior}^{-1} + \mathbf{R}^T \mathbf{K}_{noise}^{-1} \mathbf{R})^{-1}.$$

2.5 Results

2.5.1 Dichromatic versus monochromatic vision at high illumination levels

We begin by comparing the performance of dichromatic and monochromatic vision at a high illumination level. Consider two fixed classes of photoreceptor, with class 1 expressing photopigment sensitive only to wavelength λ_1 and class 2 expressing photopigment sensitive only to wavelength λ_2 . We assume that the mean and variance of the image is the same for the two wavelengths and that the two classes of photoreceptors have the same dark noise. These choices make the model environment and visual system symmetric with respect to wavelength. To mimic the statistical structure of natural images, we choose high values for the spatial and wavelength correlations: r_S was set to 0.9 and r_C was set to 0.8. We fixed the optical blur (Gaussian with standard deviation of 2 pixels), photoreceptor aperture (3 pixels), illumination level ($\omega = 2,000$), and dark noise of each photoreceptor class ($d_1 = d_2 = 1$). These choices lead to a good signal-to-noise ratio (SNR) in individual photoreceptors. To see this, note that receptor SNR is given by

$$\text{SNR} = \frac{\omega \times \sigma_{\lambda_1}}{\sqrt{\omega \times \bar{m}_{\lambda_1} + d_1}}.$$

Here, SNR = 44.7 or 16.5 dB.

To compare dichromatic and monochromatic vision, we computed O_F^{opt} for all 4,096 possible regularly spaced arrays of 12 photoreceptors on a 36 pixel line. Two of these arrangements, one with all class 1 receptors and the other with all class 2 receptors, represent monochromatic vision. The other 4,094 arrangements represent dichromatic vision with varying choices of class 1 to class 2 receptor ratio and with various arrangements of the two classes. We then asked which mosaic arrangement led to the minimum value of O_F^{opt} . We found that an alternating array of class 1 and class 2 receptors, corresponding to dichromatic vision, performs best. Within our model environment then, we have identified a set of environmental conditions under which color vision is favored.

It is instructive to examine the results in more detail. We begin by considering the effect of the relative numbers of the two photoreceptor classes. Following Hofer et al. (2005), we define a receptor class asymmetry index that groups together mosaics from the full set of 4,096 that share the same number of class 1 and class 2 receptors. The asymmetry index α is defined as:

$$\alpha = \left| \frac{\mathbb{L}_1 - \mathbb{L}_2}{\mathbb{L}} \right|,$$

where \mathbb{L}_1 is the number of photoreceptors of type 1 and \mathbb{L}_2 is the number of type 2.³ The asymmetry index takes on a value of 0 for retinae with equal numbers of each photoreceptor class, and a value of 1 for retinae that contain only one class. The index is insensitive to interchange of class 1 and class 2 labels or shuffling of receptor locations.

For each value of asymmetry index α , Figure 2.2A plots the minimum value

³In using and interpreting this and our other index expressions, we only consider the case where \mathbb{L} is even, a condition that holds for all results reported in this article.

O_F^{opt} , with the minimum taken over all arrangements with that value of α . The plot confirms our statement above, that for these conditions the best performance (minimum O_F^{opt}) is obtained for $\alpha = 0$ (dichromatic). In addition, it shows that the worst performance is obtained for monochromatic retinæ ($\alpha = 1$). As the mosaic moves from equal numbers of the two classes to all of one, performance degrades in a smooth and regular fashion.

One might reasonably argue that the comparison above is too restrictive, as we considered only receptors with spectral sensitivity confined to one wavelength. More generally, we could consider monochromatic retinæ with other spectral sensitivities. Given that our model environment has just two wavelengths, we can express spectral sensitivity through a parameter γ that describes the relative amounts of each of the two photopigments contained in a given photoreceptor. The calculations above effectively considered only $\gamma = 0$ or $\gamma = 1$, corresponding to photoreceptors containing only class 1 or class 2 photopigment. Figure 2.2B shows the effect of allowing other values of γ on the performance of a monochromatic retina. The optimal value of γ is 0.5, an equal mix of the two underlying photopigments. Note, however, that the value of O_F^{opt} for this optimal mix ($\sim 4.5 \times 10^7$) is still considerably higher than the level of performance obtained with the best dichromatic retina ($\sim 3.5 \times 10^7$). Thus even when spectral sensitivity is optimized for monochromatic vision, a dichromatic retina is better than a monochromatic retina, at least for the high illumination intensity we simulated.

Finally, note that for a given value of the asymmetry index, there may be multiple mosaics that only differ in how the two cone classes are arranged. So, for example, given that an equal number of the two receptor classes leads to best performance, we can then ask how the receptors should be arranged. We defined an alternation index, ρ , that captures the extent to which the two receptors are

interleaved in the overall mosaic. The computation of the alternation index begins with the computation of a raw alternation index:

$$\rho_{raw} = \sum_{n=1}^{L-1} D(n, n+1),$$

where $D(u, v) = 1$ if the photoreceptor at positions u and v are of different types, and 0 otherwise. The value of ρ_{raw} is a count of the number of alternations of receptor type as one moves along the one-dimensional mosaic. Because the number of possible alternations depends on the relative number of the two receptor classes, we convert the raw alternation index to the alternation index ρ by taking the percentile ranking of ρ_{raw} within the set of all possible arrangements that share the same asymmetry index α , and dividing by 100. This leads to an alternation index that is always in the range 0 to 1, that takes on a value of 0 when the two receptor classes are maximally grouped together in the mosaic, and that takes on a value of 1 when the two classes alternate as much as possible.

Figure 2.2C shows optimized error plotted against alternation index ρ . The plot makes it clear that arrangements which interleave the two photoreceptor classes as much as possible are preferred over less alternating arrangements. This is true for the optimal value of the asymmetry index ($\alpha = 0$) as well as for other values.

2.5.2 Dichromatic versus monochromatic vision across illumination levels

The section above demonstrates conditions where the performance of a dichromatic retina dominates that of a monochromatic retina. The standard intuition, as discussed in the introduction, is that this situation should reverse when light level drops. To investigate, we repeated the calculations for a range of illuminant inten-

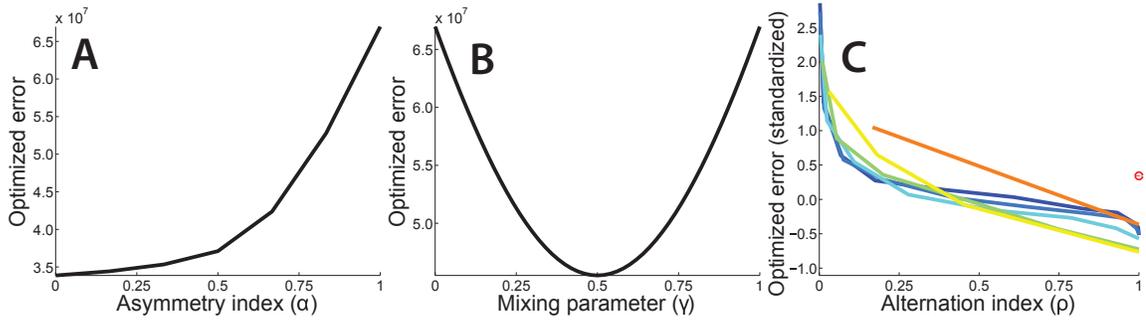


Figure 2.2. Optimized error (O_F^{opt}) for equal signal-to-noise case. This figure presents results where the mean and variance of the signals in the two wavelength bands is the same, and where the two classes of photopigment have the same dark noise. **A.** O_F^{opt} is plotted as a function of the asymmetry index, α . For each value of α , performance is plotted for the particular arrangement with that α that minimized O_F^{opt} . The two photoreceptor classes had $\gamma = 0$ and $\gamma = 1$ respectively, meaning that each contained a completely separate class of photopigment. Best performance is obtained for $\alpha = 0$, which corresponds to equal numbers of the two photoreceptor classes. **B.** O_F^{opt} is plotted as a function of the photopigment mixing parameter, γ , for a retina with only one class of photoreceptor. Best performance is obtained for $\gamma = 0.5$, corresponding to an equal mix of the two photopigment types. **C.** Effect of regularity on performance. Each line in the plot shows O_F^{opt} as a function of the regularity index, ρ . Each line in the plot corresponds to a different choice of α : 0 (dark blue), $\frac{1}{6}$ (blue), $\frac{1}{3}$ (light blue), $\frac{1}{2}$ (light green), $\frac{2}{3}$ (yellow), $\frac{5}{6}$ (orange), and 1 (red dot). For each value of α and ρ , performance is plotted for the particular arrangement that minimized O_F^{opt} . In this panel, O_F^{opt} is plotted in standardized (z-score) units computed separately for each value of α . This allows us to compare the effect of ρ for each α without undue expansion of the scale of the y-axis. For each α , the best performance is obtained for a maximally regular arrangement ($\rho = 1$). As in Panel A, these calculations were performed for two photoreceptor classes with $\gamma = 0$ and $\gamma = 1$ respectively. For all panels, illumination intensity (ω) was set to 2,000. We set the dark noise for both photopigments to 1 spontaneous isomerization per unit time. We used a duration of 1 unit time for this and all other calculations in the chapter. The mean and variance in number of reflected quanta for both color bands was set to 1. Color correlation (r_C) was set to 0.8 and spatial correlation (r_S) was set to 0.9. The standard deviation of the Gaussian blur was set to 2 pixels and the photoreceptor aperture was 3 pixels. All simulations were performed using 12 photoreceptors and 36 image pixels.

sities ω . Note, by the way, that here we are not attempting to study what factors might mediate the known shift between cone-mediated and rod-mediated vision as light level drops. To do so would require adding a model of the differences between cones and rods to our formulation and then comparing performance of each system across a range of light levels. Rather, our present goal is to compare performance of monochromatic and dichromatic mosaics within a single type of receptor. Given our focus on what happens at low light levels, this type should be conceived of as the rods.

Figure 2.3 plots the difference in optimized error O_F^{opt} between the best monochromatic retina and best dichromatic retina as a function of light level. For the monochromatic retina, we found the best value of spectral sensitivity γ for each choice of illuminant intensity ω and used this value in the comparisons. In the event, the best value was always $\gamma = 0.5$. For the dichromatic retina, we restricted attention to the case where each of the two photoreceptor classes was sensitive only to a single wavelength, and used an alternating arrangement with equal numbers of photoreceptors of each class. Given the restriction to two classes, the alternating/equal number arrangement was in fact the best for every illumination intensity. The figure shows that at every light level, the dichromatic retina outperforms the monochromatic retina.

Our calculations do not confirm the standard intuition about why rod-mediated vision is almost always monochromatic. Nonetheless, it is worth noting that our calculations are consistent with some aspects of previous thinking. The core of the standard intuition is that the advantage for color vision decreases as SNR drops, and our calculations reveal this effect: the difference in O_F^{opt} between monochromatic and dichromatic vision tends to zero with decreasing SNR. Indeed, this is not surprising in the limit, since when signal-to-noise reaches zero useful vision

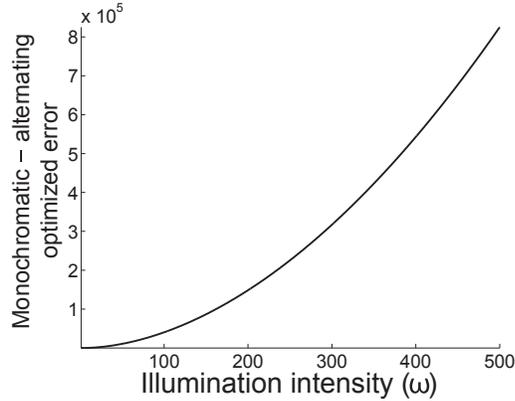


Figure 2.3. Comparison of dichromatic and monochromatic vision for equal photoreceptor signal-to-noise and high color and spatial correlations. This figure compares the optimized error for the best dichromatic and monochromatic arrangements. The dichromatic arrangement was constrained to have $\gamma = 0$ and $\gamma = 1$ for the two photoreceptor classes and had $\alpha = 0$ and $\rho = 1$. The monochromatic arrangement had $\gamma = 0.5$. The plot shows the difference between monochromatic and dichromatic O_F^{opt} as a function of illumination intensity. Positive values correspond to the case where the dichromatic arrangement has a smaller O_F^{opt} . For this case, dichromatic vision is better than monochromatic vision at all light levels. Signal and photoreceptor parameters were the same as in Figure 2.2.

disappears and the organism's estimates of the image are optimizing by guessing based on the mean of the image prior, independent of the receptor responses. What our results now add is the fact that a decreasing advantage of color vision does not necessarily translate into an advantage for monochromatic vision. To understand the selective advantage for monochromatic vision at night, some additional consideration must be added to the model system. We explore two possibilities below. One is the statistical characterization of the image ensemble, and the other is asymmetric signal-to-noise across the two receptor classes.

2.5.3 Effect of image statistics

All simulations discussed above used $r_s = 0.9, r_c = 0.8$. Although this choice reflects a reasonable characterization of natural images, it is of interest to explore the effect of varying these parameters. Figure 2.4 shows the comparison of dichro-

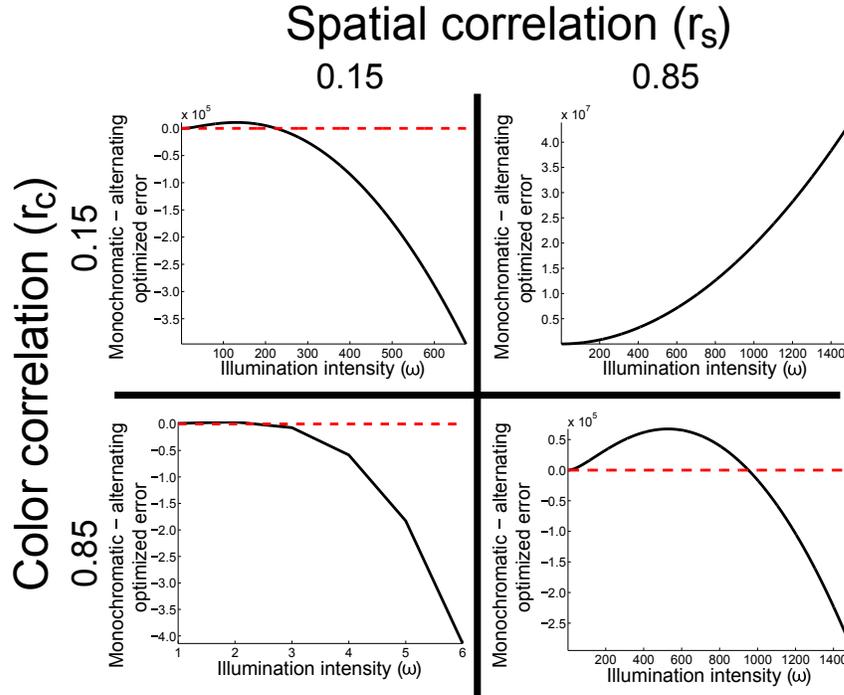


Figure 2.4. Comparison of dichromatic and monochromatic retinae for varying color and spatial correlations. All panels compare O_F^{opt} for the alternating dichromatic and best monochromatic arrangements. The dichromatic arrangement was constrained to have $\gamma = 0$ and $\gamma = 1$ for the two photoreceptor classes, with $\alpha = 0$ and $\rho = 1$. For monochromatic arrangements, the optimal value of γ was allowed to vary with illumination intensity. The plot shows the difference between monochromatic and dichromatic O_F^{opt} as a function of illumination intensity. Positive values correspond to the case where the dichromatic retina is better. Spatial correlation (r_S) was set to either 0.15 (left panels) or 0.85 (right panels). Color correlation (r_C) was set to either 0.15 (top panels) or 0.85 (bottom panels). A red dashed horizontal line representing equal error for the monochromatic and dichromatic arrangements is drawn in panels containing a crossover from monochromatic to dichromatic vision being optimal. All signal and photoreceptor parameters were the same as in Figure 2.2. Note that the scale of the x -axis is chosen differently in each panel, so as to better illustrate the transition between dichromatic and monochromatic advantage.

matic and monochromatic retinal designs for four choices of r_S and r_C , with other parameters the same as used to produce Figure 2.3.

The figure shows that dichromatic vision dominates at low light levels, regardless of spatial and color correlations. Dichromatic vision also dominates at high light levels, when $r_S > r_C$ (top right). However, when $r_S \leq r_C$ (top left, bottom panels), the high light-level pattern reverses and monochromatic vision becomes

better at high light levels.

We draw two conclusions from this analysis. First, the tradeoff between dichromatic and monochromatic vision is quite sensitive to the statistical structure of the environment. This insight is consistent with consideration of limiting cases. If the color correlation $r_C = 1$ then there is no chromatic variation in the environment and no added benefit of a second receptor class. On the other hand, if $r_S = 1$ there is no spatial variation in the environment and having multiple receptor classes has clear benefits. Our calculations show that a fairly rich set of tradeoffs is available across cases intermediate to these two environmental extrema.

The second conclusion we draw is that none of the r_S, r_C pairs we considered lead to a tradeoff pattern consistent with the biological observations, making it likely that additional factors played an important role in the evolution of vertebrate mosaic design.

2.5.4 Asymmetric photopigment dark noise

All of our calculations up to this point have assumed that the two photopigment classes are equally reliable and that signals in the two color bands are equally informative. As noted in the introduction, however, for biological photoreceptors the spontaneous isomerization rate might be expected to vary systematically with the wavelength of peak sensitivity (Platt, 1956; Barlow, 1957). For this reason, we explored the effect of specifying different levels of dark noise for the two photopigments.

Figure 2.5 shows, for a single intermediate illumination intensity, the effects of making one photopigment 100 times noisier than the other, by setting the dark noise ratio $d_1 : d_2 = 100$. This manipulation introduces an asymmetry into the results. Unlike the symmetric case, the best dichromatic photoreceptor arrangements

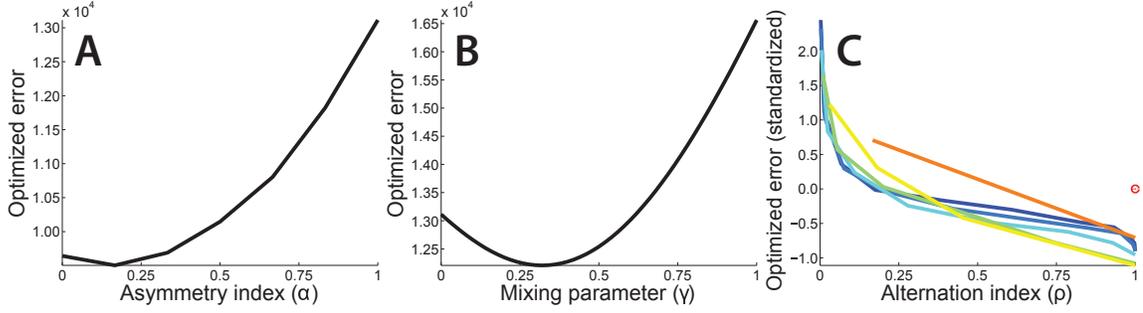


Figure 2.5. Optimized error (O_F^{opt}) for unequal dark noise case. This figure presents results where the mean and variance of the signals in the two wavelength bands are the same, but where the dark noise of class 1 photopigment is 100 times greater than the dark noise of class 2 photopigment. **A.** O_F^{opt} is plotted as a function of the asymmetry index, α . The two photoreceptors had $\gamma = 0$ and $\gamma = 1$ respectively. Best performance is obtained for $\alpha = \frac{1}{6}$, which corresponds to an arrangement containing five class 1 photoreceptors for every seven class 2 photoreceptors. **B.** O_F^{opt} is plotted as a function of the photopigment mixing parameter, γ , for a retina with only one class of photoreceptor. Best performance is obtained for $\gamma = 0.32$, which corresponds to a mix of 32% class 1 photopigment and 68% class 2 photopigment. **C.** Effect of regularity on performance. Each line in the plot corresponds to a different choice of α : 0 (dark blue), $\frac{1}{6}$ (blue), $\frac{1}{3}$ (light blue), $\frac{1}{2}$ (light green), $\frac{2}{3}$ (yellow), $\frac{5}{6}$ (orange), and 1 (red dot). For each value of α and ρ , performance is plotted in standardized (z-score) units computed separately for each value of α . For all values of α , best performance is obtained for a maximally regular arrangement ($\rho = 1$). As in Panel A, these calculations were performed for two photoreceptor classes with $\gamma = 0$ and $\gamma = 1$ respectively. For all panels, illumination intensity (ω) was set to 75. The mean and variance in number of reflected quanta for both colorbands was set to 1. Color correlation (r_C) was set to 0.8 and spatial correlation (r_S) was set to 0.9. The standard deviation of the Gaussian blur was set to 2 pixels and the photoreceptor aperture was 3 pixels. All simulations were performed using 12 photoreceptors and 36 image pixels.

now contain more of the lower-dark noise photoreceptors (Figure 2.5A). Similarly, the best monochromatic arrangements contain more of the lower-dark noise photopigment (Figure 2.5B). As for the symmetric case, however, Figure 2.5C shows that maximizing the alternation index ($\rho = 1$) optimizes performance.

The most striking new result to emerge from the asymmetric dark noise calculations is shown in Figure 2.6. Recall that dichromatic arrangements always out-performed monochromatic arrangements for the symmetric case, regardless of SNR (Figure 2.3). With the more realistic modeling of asymmetric dark noise, dichromatic arrangements still out-perform monochromatic arrangements at high illumination intensities. As illumination intensity decreases and SNR drops, how-

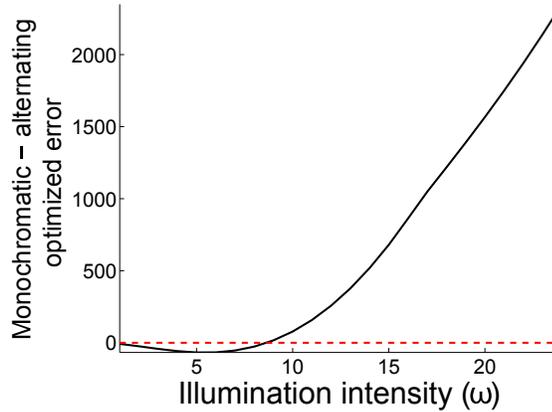


Figure 2.6. Comparison of dichromatic and monochromatic vision for unequal photopigment dark noise and high color and spatial correlations. We compare the O_F^{opt} for the alternating dichromatic and best monochromatic arrangements. The dichromatic arrangement was constrained to have $\gamma = 0$ and $\gamma = 1$ for the two photoreceptor classes and had $\alpha = 0$ and $\rho = 1$. The optimal value of γ for the monochromatic arrangement was allowed to vary with illumination intensity. The plot shows the difference between monochromatic and dichromatic O_F^{opt} as a function of illumination intensity. Positive values correspond to the case where the dichromatic arrangement has a smaller O_F^{opt} . For this case, monochromatic vision is better at low light levels, and dichromatic vision is better at high light levels. Signal and photoreceptor parameters were the same as in Figure 2.5.

ever, monochromatic arrangements lead to lower optimized error. This effect can be explained by the following intuition: at high light levels, photon noise dominates dark noise and the asymmetry between photopigment dark noise has a negligible effect. In this case, the extra information about the image transduced by a dichromatic retina is the dominant factor. At low light levels, dark noise becomes the dominant noise source, and the noise advantage of the less noisy photopigment drives the optimal design.

The crossover illumination intensity, below which monochromatic vision dominates and above which dichromatic vision dominates, depends on the relative reliability of the two photopigments. As one photoreceptor class becomes more unreliable relative to the other, the illumination intensity at which dichromatic vision becomes beneficial increases (not shown).

In Figure 2.7, we systematically explore how the optimal dichromatic and

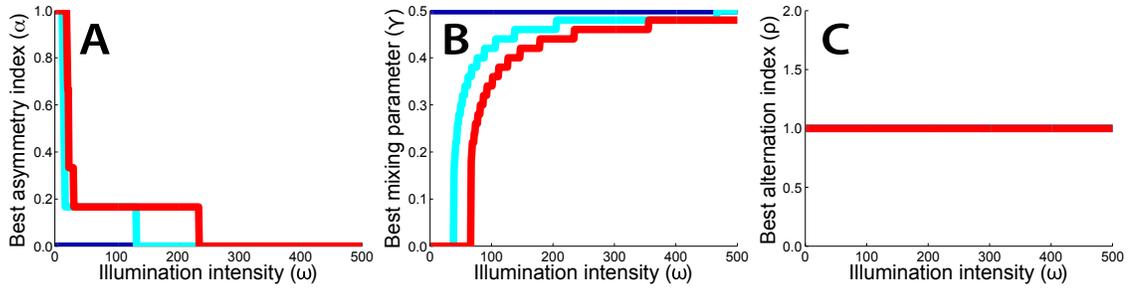


Figure 2.7. Best A. asymmetry index, B. mixing parameter, and C. alternation index, as a function of illumination intensity (ω). For all panels, class 1 photopigment is as reliable as class 2 photopigment (blue); 400x more reliable than class 2 photopigment (cyan); or 1,000x more reliable than class 2 photopigment (red). Class 1 photopigment had a dark noise of 1 spontaneous isomerization per unit time. Signal and photoreceptor parameters were the same as in Figure 2.2. Note that in Panel C, all three curves lie on top of each other.

monochromatic photoreceptor arrangements change with illumination intensity. In each panel, the three colored lines correspond to three $d_1 : d_2$ ratios. For reference, the case where $d_1 = d_2$ is shown in dark blue. For this case, independent of illumination intensity, the best dichromatic arrangements have equal numbers of the two photoreceptor classes, arranged in an alternating pattern ($\alpha = 0, \rho = 1$), while the best monochromatic arrangements have equal amounts of the two photopigments ($\gamma = 0.5$). As $d_1 : d_2$ is increased, optimal arrangements exhibit asymmetries in the relative numbers of photoreceptors (or amounts of photopigment) at low illumination intensities. The figure shows that the rate at which the arrangements return to having equal number of photoreceptors of each class (or equal amounts of the two photopigments) as a function of illumination intensity depends on the $d_1 : d_2$ ratio. The shifts in optimal spectral sensitivity with light level for monochromatic mosaics (Panel B) are an analog, within our model system, of the Purkinje shift as analyzed by Barlow (1957). For all conditions studied, it is always best to interleave photoreceptors as much as possible (Panel C, $\rho = 1$).

Introducing an asymmetry in the level of dark noise for different photopigments is not the only way to produce an asymmetry in the signal-to-noise ratio of the out-

put of different photoreceptor classes. Such asymmetries can also be produced if the mean light level is different across wavelengths, if there is a difference in the degree of retinal absorption across wavelengths (as might be produced by protective UV-absorbing macular pigment), a difference in the quantal efficiency of different photopigments, or if the variance in the signals at different wavelengths differs. We would expect that effects similar to those shown above for the asymmetric dark noise case would also be produced by signal asymmetries. Indeed, we explored this explicitly for the case of unequal signal variance and found a pattern of results very similar to that shown in Figures 2.5 and 2.6. In the interest of brevity, we do not present those results here.

2.5.5 Optimizing dichromatic spectral sensitivity

In the calculations above for dichromatic mosaics, we considered only photoreceptor classes with $\gamma = 0$ and $\gamma = 1$ respectively. Across species, there is variation in the degree of overlap between photoreceptor spectral sensitivities between photoreceptors of different classes. For example, in goldfish the peak sensitivities of the M and L cones are noticeably more separated, at 530 nm and 620 nm, than in primates, where the peak sensitivities are approximately 530 nm and 560 nm (Bowmaker, 1991). We wondered what the optimal choices of spectral sensitivity were in our model system.

To study this while avoiding parameter explosion, we considered how O_F^{opt} varied with a parameter η , where the photoreceptor sensitivities of the two classes depend on η through $\gamma_1 = \eta$ and $\gamma_2 = 1 - \eta$. We studied mosaics with equal numbers of receptors from each class in an alternating arrangement. Note that when $\eta = 0.5$, the dichromatic arrangement is equivalent to the monochromatic arrangement with $\gamma = 0.5$, while $\eta = 0$ and $\eta = 1$ reduce to the dichromatic arrangement studied

above.

Figure 2.8 plots O_F^{opt} as a function of η , for the alternating dichromatic arrangement, at a single illumination intensity. The plot has two minima, at the symmetric locations $\eta = 0.27$ and $\eta = 0.73$.⁴ Adding some degree of spectral overlap between the two photoreceptor classes improves performance, but too much overlap has a significant deleterious effect. Indeed, in the limit of $\eta = 0.5$ this is simply a re-expression of the fact that monochromatic retina does worse than a dichromatic retina.

Figure 2.8 shows results for the case of symmetric receptor dark noise. In the asymmetric case, the same value of η characterizes optimal design for high light levels (since at high light levels dark noise is dominated by photon noise.) Because our results in Figure 2.6 were obtained for $\eta = 0$ (or equivalently $\eta = 1$), it is possible that the advantage for monochromatic vision at low light levels shown in the figure would reverse if the calculations were performed with optimized dichromatic spectral sensitivity. We ruled out this possibility by verifying that monochromatic vision continues to dominate at low light levels in the asymmetric dark noise case, even for $\eta = 0.27$ (or equivalently $\eta = 0.73$).

More generally, one could consider whether any dichromatic retina outperforms a monochromatic retina at low light levels, for the asymmetric dark noise case. Exploring all possible choices of spectral sensitivity for the two receptor classes for all possible arrangements is computationally prohibitive. But since a monochromatic retina represents a limiting case of a dichromatic retina where the spectral sensitivities of the two classes become progressively more similar, and where the asymmetry index approaches unity, we can ask whether making a monochromatic

⁴The symmetry arises because for the case studied there is no effect of reversing the labels given to the two photoreceptor classes. For this reason, only half of the plot is shown.

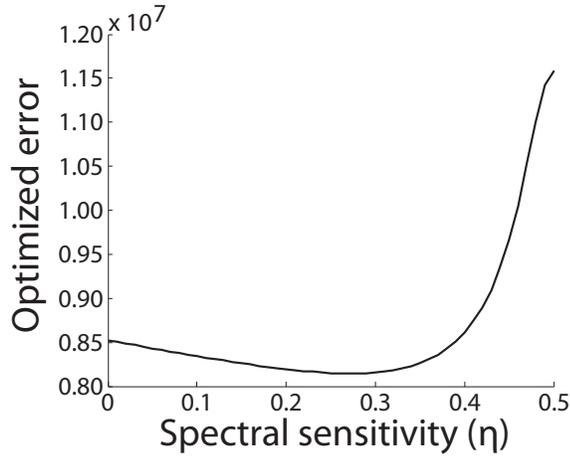


Figure 2.8. Effect of spectral sensitivity. This figure plots O_F^{opt} as a function of the spectral sensitivity parameter η , for the alternating arrangement ($\alpha = 0, \rho = 1$), at a single illumination intensity ($\omega = 1,000$). The optimal values of η are 0.27 and 0.73, corresponding to mixes of 27% of one photopigment and 73% of the other. Aside from illumination intensity and photopigment mixing parameters, the signal and photoreceptor parameters were the same as in Figure 2.2. Since this figure is symmetric about $\eta = 0.5$, we restrict the domain displayed to $0 \leq \eta \leq 0.5$.

retina just a little bit dichromatic improves performance or worsens it. To do so, we compared the performance of a monochromatic retina with all receptors containing the low-dark-noise photopigment to that of a retina where a small amount (10%) of the high-dark-noise photopigment had been added to a single photoreceptor in the array. We found that for the asymmetric dark noise case, the purely monochromatic retina led to superior performance at low illumination levels. This in turn suggests that indeed the purely monochromatic retina represents a global optimum in design.

2.6 Discussion

Animals have evolved sense organs that are near optimal for detecting and measuring changes in the environment. For example, rod photoreceptors can reliably transduce the energy of a single quantum of light (Baylor, Lamb, & Yau, 1979);

hair cells in the ear are sensitive to deflections as small as the radius of a single hydrogen atom (Sellick, Patuzzi, & Johnstone, 1982); and olfactory sensory neurons can respond to the arrival of a single molecule of odorant (Schneider, 1969).

More generally, many properties of sensory systems have been successfully understood as near optimal solutions to information uptake and processing problems. For example, the design of a fly's compound eye has been shown to represent an optimal tradeoff between spatial resolution and sensitivity to contrast (Snyder, Stavenga, & Laughlin, 1977); the spectral properties of primate cone photoreceptors optimize discrimination of fruit from foliage (Regan et al., 2001); properties of retinal ganglion cells optimize information transmission down the optic nerve (Buchsbbaum & Gottschalk, 1983; Atick, Li, & Redlich, 1992; Atick, 1992; van Hateren, 1992, 1993; Twer & MacLeod, 2001; Koch et al., 2004); and adaptation maximizes the use of limited neural dynamic range across changes in environmental conditions (Laughlin & Hardie, 1978; J. Walraven, Enroth-Cugell, Hood, MacLeod, & Schnapf, 1990).

Optimality calculations provide a principled null model against which to benchmark the performance of sensory and other information processing systems (Geisler, 1987; Watson, 1987; Brainard, 1993). In addition, showing how a feature of a biological system is closely matched to predictions derived from considerations of optimality provides a satisfying, if speculative, answer to the question of why that particular feature evolved.

In this chapter, we develop a theoretical methodology to evaluate the quality of different choices of retinal photoreceptor mosaic. Our method is based on Bayesian decision theory (Gelman et al., 2004; Berger, 1985; Blackwell & Girschick, 1954). Similar underlying logic was employed by Srinivasan et al. (1982) in the context of understanding the receptive fields of retinal ganglion cells, and our general

formulation closely parallels that developed independently by Levin et al. (2008) to evaluate the design of digital cameras.

We applied our method in the context of a simplified model system to investigate three fundamental questions about retinal design. First, we showed that for dichromatic retinae, regular arrangements that maximally alternate between the two types of photoreceptors lead to better performance than less regular arrangements. Second, introducing overlap between the spectral sensitivities of the photoreceptors of dichromatic retinae can improve performance. Third, when we compare dichromatic and monochromatic retinae, we find that which is preferred is quite sensitive to the specific parameters chosen within our model system. In the symmetric case where the different photoreceptor types have the same level of intrinsic noise and the different wavelengths carry the same amount of signal, a dichromatic retina yields best performance at all signal-to-noise ratios, when the correlations between neighboring image locations are higher than those between neighboring wavelength bands. To obtain results that mirror what is typically found in biological systems (monochromatic vision for low signal-to-noise conditions and color vision for high signal-to-noise conditions), we had to introduce an asymmetry between the intrinsic noise level of the two photoreceptor classes.

2.6.1 Measurements of photoreceptor dark noise

A number of laboratories have measured photoreceptor dark noise for receptors with different pigment spectral sensitivities. The most relevant measurements to our analysis are those for rods, which mediate vertebrate vision at low light levels. Baylor et al. (1984) showed that thermal like events dominate the dark noise of primate rods, suggesting strongly that at the lowest light-levels it is thermal noise rather than later noise that sets the limit on performance. Ala-Laurila,

Pahlberg, Koskelainen, and Donner (2004) measured the activation energy for 12 visual pigments varying in λ_{max} and found that activation energy decreased with λ_{max} , although the decrease was shallower than the $1/\lambda_{max}$ dependence that would be predicted if the only factor involved were the energy of photons at wavelength of peak sensitivity. This result does not directly show that receptor dark noise depends on λ_{max} , but does confirm a key feature of the theory that predicts such dependence. More recently, Ala-Laurila, Donner, Crouch, and Cornwall (2007) estimated thermal noise rates for two forms of rhodopsin in isolated salamander rods and concluded that this was larger for the form with longer λ_{max} . Taken together, these results make it highly plausible that there are fundamental dark noise asymmetries across rod photoreceptors with different wavelengths of peak sensitivity, as Barlow (1957) hypothesized.

Rieke and Baylor (2000) compared dark noise in salamander S and L cones, and found that the noise in L cones was much greater. For L cones, they concluded that dark noise was dominated by thermal isomerizations, while for S cones, other noise sources dominated. Their results are consistent with a dependence of dark noise on $1/\lambda_{max}$. On the other hand, Schneeweis and Schnapf (1999) found little difference in the dark noise levels of primate M and L cones, perhaps because the dark noise in these receptors is dominated by sources other than thermal isomerizations. It is probably wise to keep in mind that across receptor types and species, the relationship between λ_{max} and dark noise may not be simple. For our present purposes, however, the crucial point is that at the lowest light levels, asymmetries in rod SNR across receptor types seem likely to play an important role in retinal design.

2.6.2 Relation to other work: color vision at night

As noted in the introduction, a few other authors have considered why monochromatic vision prevails at low light levels. Although our calculations do not contradict the core observation of what we have called the standard intuition on this matter (van Hateren, 1993; Land & Osorio, 2003), they show that the standard intuition alone is insufficient and requires more careful elaboration. Indeed, in the absence of any asymmetry in the noise properties of photoreceptor types or in the signal carried by different image wavelengths, a dichromatic retina can outperform a monochromatic retina at all signal-to-noise ratios. This conclusion holds for the case where the spatial correlation between neighboring image locations is higher than the color correlation across neighboring image wavelengths.

We are, however, able to show that monochromatic vision dominates dichromatic vision at low light levels if we introduce an asymmetry across photoreceptor types or between the statistics at different image wavelengths. Such asymmetries interact with the signal-to-noise ratio to drive a shift in the optimal mosaic arrangement. The critical role of the noise asymmetry parallels Barlow's (1957) explanation of the Purkinje shift. As discussed in the previous section, there is good reason to believe that there are in fact asymmetries in photoreceptor noisiness (Barlow, 1957; Rieke & Baylor, 2000). Less is known about possible asymmetries in the signals carried at different image wavelengths, but these may also exist. Thus, our analysis provides a plausible explanation for the rarity of color vision at night, but only with the assumption of asymmetries across photoreceptors and/or wavelengths.

Note that our work does not contradict van Hateren's (1993) results, as there are several key differences between his analysis and that presented here. First, we explicitly compare the performance of interleaved dichromatic retinæ with that of monochromatic retinæ. This is quite different than asking, for any particular

retina, whether the optimal processing of the output of that retina tends to towards luminance processing at low light levels. Second, considering the properties of the optimal single channel, as van Hateren did, is different from asking whether adding a second channel might improve performance.

Finally, we note that we can replicate van Hateren's (1993) basic conclusion within the context of our current model. We asked, for the case where the two receptor classes have equal SNR, how the reconstruction of luminance (both receptor types stimulated together) and chromatic (the two receptor types stimulated in opposition) signals compared as a function of SNR (results not shown). We found, consistent with van Hateren, that the magnitude of the reconstructed chromatic signal decreased relative to that of the reconstructed luminance signal as overall SNR dropped. This calculation is consistent with the standard intuition. Interestingly, it is also consistent with the observed *achromatic interval* found for cone-mediated foveal viewing in humans: as light level drops, there is a regime where dim test lights can be detected but not be judged in color (P. L. Walraven, 1962; Graham & Hsia, 1969). Massof (1977) presents a model where detection is limited by quantum fluctuations that accounts for this interval. The central idea driving Massof's model is the same as that driving the standard intuition.

2.6.3 Relation to other work: mosaic regularity

In all of our calculations, we found that the best arrangement of the multiple (in this case two) receptor classes is regular, in the sense that it maximizes the number of alternations between classes across the mosaic. This result is consistent with the regular mosaic layout found in fish (Scholes, 1975; Bowmaker & Kunz, 1987) and with the regular arrangement of the S cone submosaic in humans (Curcio et al., 1991), but not with the quasi-random arrangement of L and M cone submosaics in

primates (Mollon & Bowmaker, 1992; Hofer et al., 2005).

A few investigators have considered the effect of photoreceptor irregularity on performance. This work is in the context of the spatial arrangement of monochromatic mosaics, rather than the arrangement of interleaved submosaics. The principles considered in the earlier work, however, are of a general nature and might reasonably be taken to apply to the current case.

Bossomaier, Snyder, and Hughes (1985) concluded that mosaic irregularity had a deleterious effect on performance, consistent with our general conclusion. Their analysis, however, was based on the assumption that the visual system processed the output of the irregular mosaic as if the receptors were in fact regularly spaced. In this case, the effect of irregular spacing may be mimicked by the addition of noise to the output of a regular mosaic, and the addition of such noise would degrade performance. Recent modeling of the appearance of very small flashed spots (Brainard, Williams, & Hofer, 2008), however, indicates that the human visual system does take the fine spatial structure of its mosaic into account. In any case, the key difference between the development of Bossomaier et al. (1985) and the one presented here is that their conclusion hinges on the assumption that the visual system's post-receptoral processing is sub-optimal, while ours evaluates each mosaic arrangement on the assumption that post-receptoral processing optimally accounts for the location of each photoreceptor.

Yellott (1982, 1983) concluded, on the other hand, that mosaic irregularity can improve visual performance. He analyzed the nature of the set of spatial aliases for different mosaic arrangements. (Two different images are aliases for a mosaic if they produce the same response in every photoreceptor in the mosaic.) For regularly arranged mosaics, regular spatial patterns at high spatial frequencies (e.g. high spatial frequency sinusoids) have as aliases other regular spatial patterns at

lower spatial frequencies (e.g. low spatial frequency sinusoids). For irregularly arranged mosaics, regular high spatial frequency patterns alias into irregular low spatial frequency patterns that appear as spatial noise. Yellot pointed out that a visual system that incorporated knowledge of the spatial structure in natural scenes would be better able to filter the low spatial frequency aliases produced by an irregular mosaic than the those produced by a regular mosaic, suggesting a possible advantage for irregular mosaics.

Yellot's argument could be formalized and evaluated within the framework presented here. The key extension to the current analysis required for this would be to choose an image prior that incorporates a preference for structured images; the Gaussian priors we used capture only the second order structure of natural images and do not allow expression of higher order structure that favors, for example, the presence of edges. Whether such an analysis would lead to a rationale for irregular mosaics within our simple model system remains an open question of considerable interest.

Finally, Hsu, Smith, Buchsbaum, and Sterling (2000) argued that in the presence of electrical coupling between receptors of different classes, irregularity in the interleaving of these classes would reduce the deleterious effects of the coupling on color vision. DeVries, Qi, Smith, Makous, and Sterling (2002) complemented this line of thinking by observing that the same coupling could reduce noise at the receptor outputs by integrating over multiple receptors, and that the cost of such coupling for spatial vision was not significant in the presence of optical blur. We have not explored these results in the context of our model system. Doing so would require imposing specific processing steps after the receptors (e.g. modeling electrical coupling), and then asking both how the added processing constraint affected retinal design as well as overall performance.

2.6.4 Relation to other work: spectral sensitivity

A number of authors have considered the interaction of spectral properties of the photic environment and cone spectral sensitivity. The theoretical approach taken by Lewis and Zhaoping (2006) is quite similar to ours, although their performance criterion is one of maximizing information rather than minimizing Bayes risk (see below). These authors conclude that the human L cone is not optimally located, as more information would be transmitted if its peak sensitivity were at a higher wavelength. As with our work, a key factor that drives their calculations is asymmetry between the information available for photoreceptors with different wavelengths of peak sensitivity. In particular, they note that as the overall signal-to-noise ratio decreases, the optimal placement of both L and M cones tends toward the wavelength that provides maximum signal-to-noise. This is analogous to our conclusion that the optimal mosaic tends towards monochromatic as signal-to-noise decreases. One important difference between our work and theirs is that we explicitly evaluate spatial and spectral performance jointly, whereas they do not treat spatial variation.

Other papers in this general tradition (Lythgoe & Partridge, 1989; Chittka & Menzel, 1992; Osorio & Vorobyev, 1996; Regan et al., 2001; Cummings, 2004) analyze performance on a particular color discrimination task thought to be important for the organism (e.g. discriminating fruits from foliage for primates; Regan et al., 2001) as a function of choice of spectral sensitivity. This work shares with our current analysis emphasis on both the statistical properties of stimuli that will be encountered by the visual system and on optimizing performance for a particular task. The specifics of the analyses differ, with our work considering a highly simplified visual environment. This allows us to explore a wider range of parametric interactions, at the cost of departing further from ecologically valid measurements.

2.6.5 Relation to other work: theory

Our approach to optimal design employs formalism taken from Bayesian decision theory. An alternative and closely related approach is to consider maximization of transmitted information, as measured in bits, rather than minimization of expected loss. Analyses based on information theory have been successfully employed to explain features of post-receptoral retinal design (e.g., Atick, 1992; Brenner, Bialek, & de Ruyter van Steveninck, 2000; Balasubramanian & Berry, 2002; Koch et al., 2004). In the current context, the difference between the two approaches is straightforward. Given a fixed prior distribution on images, the design that maximizes information transmission is the one that leads to a posterior distribution with minimum entropy (Cover & Thomas, 1991). Entropy is a measure of the uncertainty in the outcome of draws from a distribution. A posterior distribution with low uncertainty, generally speaking, is also one that will lead to low estimation error with respect to a specified loss function. However, the relation between entropy and expected loss need not be monotonic (Thomson & Kristan, 2005).

Given that minimizing expected loss and maximizing information transmitted do not always lead to the same conclusion, one could reasonably ask which approach is more appropriate. The attraction of minimizing expected loss is that it allows explicit inclusion of what matters to the organism, and it is for this reason that we have used the Bayesian approach here. At the same time, we concede that the specific choice of squared error loss is at best a crude approximation to what matters to most organisms, and in the current work serves as a placeholder in the formalism where more realistic loss can be inserted when such are available. This general point, that information theory tends to be insensitive to the metric structure of the stimulus space, is discussed by Luce (2003). We did verify that the main features of the principal results reported in this chapter (i.e. the

monochromatic/dichromatic comparisons shown in Figs. 2.3 and 2.6) continue to hold when the calculations are repeated with maximizing transmitted information as a criterion. The one exception we found is that increasing the overlap in spectral sensitivity, which improves performance for a dichromatic retina as assessed by minimizing expected squared error loss (see Fig. 2.8), does not increase information transmitted.

An advantage of information theory that drives its use in many studies of neural systems is that there are techniques for measuring the information carried by a neuron that do not require knowledge of the analytic form of the stimulus ensemble or of the representational structure of the neural response. Thus, for example, the information about natural images conveyed by a particular ganglion cell may be estimated by measuring the cell's response to a sequence of presented natural images, without need to develop an explicit algorithm for estimating images from the response (e.g., Brenner et al., 2000; Balasubramanian & Berry, 2002; Koch et al., 2004). In this sense, information theory is an approach that complements the one we have taken here. Indeed, in related work, we considered features of mosaic design using information theoretic methods (Garrigan et al., 2010).

2.6.6 Future directions

The model system explored in this chapter was very simple. Nonetheless, it allowed us to express many of the pieces required for a full theory of optimal mosaic design, including a specification of the statistical properties of the visual environment, specification of the relation between an image and its sensory representation, and specification of the goal of the visual computation as a loss function. Given these, we could explore the performance of different design choices. Despite the fact that that the model system we used is simple, we were able to demonstrate a

number of conclusions. The first is that classic arguments for why monochromatic vision is favored over color vision at low signal-to-noise levels are incomplete. That is, we were able to demonstrate conditions where color vision provides better performance as signal-to-noise drops to zero. Moreover, our analysis suggested additional factors that, when incorporated into the analysis, do yield a shift from color to monochrome vision as SNR drops. In addition, within the context of our model system, regular mosaics dominate irregular mosaics. We have not yet found factors that would predict a robust advantage for the quasi-random tiling of L and M cones found in primate retina. Here our analysis clarifies not what we understand, but what we have yet to make sense of.

Because our model system is very simple, an important goal for future work is to generalize the analysis and bring it closer to biological realism. Obvious generalizations include adding a second spatial dimension, additional wavelengths (i.e. full spectra), time and motion, and additional cone classes. There are no conceptual obstacles to these generalizations, but given present computing resources it is not currently possible to explore these added dimensions exhaustively, as we could for the simple case. The challenge now is to pick specific manipulations of interest. Moving in this direction would also allow incorporation of additional factors, such as whether the statistics of spectra shift systematically between day and night — something we have not yet considered. These generalizations would also enable us to draw the numerical values used for spectral sensitivity and dark noise more directly from biological measurements; the present calculations are sufficiently abstracted from real retinæ that we thought such an effort premature.

Computations cost energy, and energy considerations are non-trivial for the design of biological systems (Laughlin, 2001; Balasubramanian, Kimber, & Berry, 2001; Balasubramanian & Berry, 2002; Laughlin & Sejnowski, 2003). The general

framework we have presented could be extended to include considerations of energetic cost. If we think of the loss function as expressing the cost of misestimation in terms of its ultimate effect on reproductive success, we can also conceive of a function that expresses the cost of computing an estimate from the receptor responses. Denote by the *energetic cost function* $C_F(\vec{r})$ the energetic cost incurred when the estimator $F()$ acts on input \vec{r} . In analogy to the Bayes risk, we can then associate an expected computational cost with the estimator

$$C_F = \int_{\vec{i}} \left[\int_{\vec{r}} C_F(\vec{r}) P(\vec{r}|\vec{i}) d\vec{r} \right] P(\vec{i}) d\vec{i}.$$

Once computational cost is taken into account, a more general optimality principle is to choose an estimator $F()$ that minimizes $O_F = R_F + C_F$.⁵ Note that this formulation for computational cost would also allow for consequences of the time taken to do the computation (Koch et al., 2006) and for variations in cost that arise from variation in the size and mass of the computational apparatus (Laughlin & Sejnowski, 2003). It might be reasonable to make initial steps toward including energetic cost by examining the number of multiplications and additions required to closely approximate the optimal estimator for each mosaic.

As noted in the results section, we have not explicitly modeled differences between rods and cones, such as differences in the source and level of receptor dark noise (see above), nor between properties of receptors across species. Because our formulation is very general, such differences could be included. Modeling the properties of rods and cones would, for example, allow us to address questions such as at what light level a visual system that has both rods and cones should

⁵The use of a simple sum is appropriate on the assumption that the measures of loss and energetic cost can be expressed in the same units. This seems reasonable in principle but may be difficult to achieve in practice.

switch from one to the other. Similarly, comparing the details of rod properties and other retinal design factors across species might provide insight as to why some species do exhibit rod-mediated color vision (Land & Osorio, 2003; Kelber & Roth, 2006).

Even within our simple one-dimensional, two-wavelength system, we did not explore all possible parametric manipulations. For example, one could explore tradeoffs between optical blur and pupil aperture, or effects of overall receptor density. One could also try to refine the statistical model of image structure by incorporating a characterization of the edge-like features that are pervasive in natural images (Simoncelli, 2005) or the model of the likelihood function (e.g. by incorporating a Poisson rather than Gaussian noise model; L. Paninski, personal communication). Finally, it would be of interest to explore effects of varying the loss function, for example to emphasize either luminance or chromatic estimation error. This would be a first step to modeling species differences in what stimuli are most ecologically relevant. More generally, examining properties of the estimator itself could yield insights about how post-receptoral processing should be configured to optimally extract information from interleaved receptor mosaics.

Chapter 3

Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans

Jeremy R. Manning, Joshua Jacobs, Itzhak Fried, & Michael J. Kahana.
The Journal of Neuroscience, 29(43), 13613–13620.

3.1 Abstract

A fundamental question in neuroscience concerns the relation between the spiking of individual neurons and the aggregate electrical activity of neuronal ensembles as seen in local field potentials (LFPs). Because LFPs reflect both spiking activity and subthreshold events, this question is not simply one of data aggregation. Recording from 20 neurosurgical patients, we directly examined the relation between LFPs

and neuronal spiking. Examining 2,030 neurons in widespread brain regions, we found that firing rates were positively correlated with broadband (2 – 150 Hz) shifts in the LFP power spectrum. In contrast, narrowband oscillations correlated both positively and negatively with firing rates at different recording sites. Therefore, broadband power shifts were a more reliable predictor of neuronal spiking than narrowband power shifts. These findings suggest that broadband LFP power provides valuable information concerning neuronal activity beyond that contained in narrowband oscillations.

3.2 Introduction

A large and growing literature has documented the existence of oscillatory activity from large neuronal ensembles in both humans and animals, and has shown that these brain oscillations are related to a wide variety of behavioral and cognitive states (Kahana, Seelig, & Madsen, 2001; Buzsáki, 2004; Buzsáki, 2006; Crone, Sinai, & Korzeniewska, 2006; J. Jacobs & Kahana, 2009; Jerbi et al., 2009). Elucidating the relation between large-scale neural patterns and the activities of individual neurons is critical for gaining a deep understanding of the brain and how it supports behavior (Kreiman, 2007). To this end, several studies of humans and animals have examined the relation between large-scale recordings, such as local field potentials (LFPs), and the activities of individual neurons.

Recording from monkey visual cortex, Rasch, Gretton, Murayama, Maass, and Logothetis (2008) found that increases in gamma activity (40–90 Hz) and decreases in delta and theta activity (1–10 Hz) in the LFP signal were correlated with increased neuronal firing rates. Examining primary auditory cortex as human neurosurgical patients viewed brief movies, Mukamel et al. (2005) also observed that increased

gamma activity and decreased theta activity correlated with neuronal firing. The positive correlation between gamma activity and neuronal spiking is a robust finding reported in a variety of studies (Fries, Reynolds, Rorie, & Desimone, 2001; Pesaran, Pezaris, Sahani, Mitra, & Andersen, 2002).

Reports of strong correlations between neuronal firing and narrowband activity (i.e., oscillations) have supported the view that oscillations reflect synchronized spike timing in large neuronal ensembles (Singer & Gray, 1995; Logothetis, 2003; Fries, Nikolić, & Singer, 2007). This follows in part from the temporal-binding hypothesis (von der Malsburg, 1981), which proposes that synchronized neural activity can solve the “binding problem” by linking multiple neuronal signals (Koffka, 1935; Köhler, 1947; Kanisza, 1979; Pal & Pal, 1993). However, several studies suggest that apparent correlations between spikes and gamma-band LFP activity are actually due to broadband LFP patterns, rather than band-specific oscillations (Mitzdorf, 1985; Juergens, Guettler, & Eckhorn, 1999; Cruikshank, Rose, & Metherate, 2001; Kaur, Lazar, & Metherate, 2004; Kreiman et al., 2006; Chen et al., 2007; K. Miller, Leuthardt, et al., 2007). Furthermore, task-related modulations in human LFP activity have been observed throughout a broad frequency range extending to ~150–200 Hz, showing that the human brain exhibits broadband phenomena in addition to narrowband oscillations (Tanji, Suzuki, Delorme, Shamoto, & Nakasato, 2005; Lachaux et al., 2005; Canolty et al., 2006). Other work suggests that broadband changes in LFP activity are related to neuronal spiking (K. Miller et al., 2009; Milstein, Mormann, Fried, & Koch, 2009). However, no previous work has directly compared neuronal spiking in humans with simultaneous broadband LFP activity.

Here we examined the relation between LFPs and single-neuron activity in 20 neurosurgical patients during a virtual navigation task. We recorded action poten-

tials from 2,030 neurons from diverse brain regions including the neocortex and medial temporal lobe. Consistent with previous studies, we found a population of *narrowband-shift neurons*, which varied their firing in proportion to LFP power at specific frequency bands. Narrowband-shift neurons were present throughout the brain, but were especially prevalent in the frontal cortex and amygdala. In addition, we observed a larger population of *broadband-shift neurons*, which varied their firing with the overall height of the LFP power spectrum at all frequencies. Broadband-shift neurons appeared in all examined brain regions, but were especially prevalent in the medial temporal lobe. Broadband increases in LFP power were almost exclusively positively correlated with single-neuron firing, providing a robust estimate of neuronal firing. We propose that, when estimating local neuronal firing using LFPs, researchers should examine broadband power in addition to power contained in narrow frequency bands.

3.3 Methods

3.3.1 Electrophysiological recordings

We examined intracranial brain recordings from 20 neurosurgical patients undergoing treatment for drug-resistant epilepsy. During each 25–60-minute recording session, patients played a virtual navigation game, *Yellow Cab*, in which they assume the role of a taxi driver and chauffeur (virtual) passengers to their desired destinations. While playing this game, patients learn the virtual environment's layout (Newman et al., 2007) and display brain oscillations at various frequencies related to learning and sensorimotor integration (Caplan et al., 2003; Ekstrom et al., 2005).

Patients are implanted with 6–12 neurosurgical depth electrodes by clinical teams. Local field potentials (LFPs) recorded from these electrodes are used to map the seizure focus and identify functional brain regions for potential subsequent surgical resection. In addition, a set of nine small microwires extend from the tip of each depth electrode. (The ninth wire is the recording reference for the other eight.) Each microwire is 40 μm in diameter and records from a small, local region of cortex. By recording from the microwires at 32 kHz, we observe both high frequency single-neuron spiking (Fried et al., 1999) and lower frequency LFPs (Mukamel et al., 2005; J. Jacobs, Kahana, Ekstrom, & Fried, 2007). Using the WaveClus software package (Quiroga et al., 2004), we identified the action potentials of 0–3 neurons per microwire, for a total of 2,030 neurons across the 20 patients. Recordings were obtained in widespread brain regions including the frontal cortex, posterior cortex (occipital and parietal cortices), amygdala, hippocampus, and parahippocampal region (Witter, 2002). In order to make the LFP data more computationally tractable, we downsampled our recordings to 2 kHz. We then applied a second-order Butterworth notch filter at 60 Hz to remove line noise. We computed the smoothed firing rate for each time point by convolving each neuron’s spike train with a Gaussian kernel (half-width = 500 ms). To prevent the low frequency components of the action-potential waveform from contaminating the LFP signal, we replaced the data samples in the –2–8-ms window around each spike with a linear interpolation of the underlying LFP signal (J. Jacobs et al., 2007). These recordings were obtained for a previous study (J. Jacobs et al., 2007), but the analyses reported here are completely novel.

3.3.2 LFP feature extraction

We measured oscillatory power in the LFP signal using Morlet wavelets (wave number = 4) at 50 log-spaced frequencies between 2 and 150 Hz ($2 \times 10^{0.0383x}$ for $x \in \{0..49\}$). Because oscillatory power at a given frequency is χ^2 distributed (Percival & Walden, 1993; Henrie & Shapley, 2005), we log-transformed the wavelet-calculated powers to make the distributions more normal. To account for inter-electrode impedance differences, we normalized the powers recorded at each electrode such that the mean power spectrum was centered at 0 with a standard deviation of 1.

To analyze the relation between LFP spectral power and spiking activity, we next divided each recording session into 500 ms epochs. This epoch length was chosen to provide a reasonable balance between temporal resolution (which we sought to maximize) and correlations across successive measurements (which we sought to minimize). To eliminate the effect of non-biological noise on our analysis, we removed epochs with firing rates above the 99th percentile. For computing summary statistics, we calculated the mean power contained in the following narrow frequency bands: delta (2–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–150 Hz). In addition to measuring LFP power at narrow frequency bands, we computed *broadband power*. Broadband power refers to LFP voltage fluctuations that occur at a broad range of frequencies, consistent with the voltage following a random-walk process (Annibaldi & Hopcraft, 2002), unlike true oscillations limited to a narrow frequency range. To measure the broadband power for each individual epoch, we fit a line to that epoch's wavelet-calculated power spectrum via a robust regression (Holland & Welsch, 1977). Robust regression fits the majority of data points closely and is relatively unaffected by outliers; by contrast, a standard least-squares regression would be more affected by data points that did not fall on the line. This is critical for distinguishing broadband and

narrowband effects because it minimizes the impact of narrowband oscillations, which appear as local peaks in the LFP power spectrum, towards the computed broadband power. We use the term *broadband power* to refer to the mean height of the robust-regression-fitted line in each epoch.

3.3.3 Regression framework

Our primary objective was to examine how the firing rates of individual neurons related to narrowband changes (i.e., oscillations) and broadband changes in the LFP. To quantify the extent to which each of these patterns predicted neuronal firing, we used a series of least-squares regressions. For each neuron, we set up five bivariate simultaneous regressions of the form

$$R = \beta_0 + \beta_B B + \beta_F F,$$

where R was a vector containing the estimated firing rate for each epoch, F was a vector containing the average power in each epoch for one narrowband frequency (delta, theta, alpha, beta, or gamma), and B was a vector containing broadband power. Prior to computing the mean power in each narrow frequency band for each electrode, we z-transformed the power distribution at each individual frequency to have a mean of 0 and a standard deviation of 1. This step ensured that the individual frequencies in each band contributed equally, despite the overall $\frac{1}{f^\alpha}$ shape of the power spectrum. The regression was performed five times (once for each frequency band), and the regression coefficients, β_0 , β_B , and β_F were fit separately each time. When the β_B coefficients from all five regressions were significantly different from zero (see bootstrap procedure, below), and all had the same sign, we designated the neuron as a *broadband-shift neuron*. This technique assured that, for broadband-

shift neurons, broadband power explained a significant portion of the firing-rate variance beyond any one narrow frequency band alone.

We designated a neuron as a *narrowband-shift neuron* if exactly one β_F was significantly different from zero (see bootstrap procedure, below). In addition, a neuron was also labeled as a narrowband-shift neuron if it had exactly two significant β_F 's with the same sign in neighboring frequency bands (e.g., $\beta_\delta < 0$ and $\beta_\theta < 0$). We added this second condition so that our analysis was not biased against neurons whose firing rates were related to oscillations that spanned the frequency band boundaries. In this way, the firing rates of narrowband-shift neurons were correlated with oscillations in one narrow frequency band. Note that a single neuron could be tagged as both a broadband-shift neuron *and* a narrowband-shift neuron. As a potential alternative to the bivariate regression framework presented here, we also considered the use of a single simultaneous regression for each neuron (using broadband power and each of the five frequency bands as regressors). However, because the different frequency bands were highly co-linear, results using this alternative approach were difficult to interpret.

3.3.4 Multiple comparisons

Our framework for classifying neurons as broadband- and narrowband-shift neurons relied on the outcomes of several potentially correlated statistical tests. To address this issue we used a bootstrap procedure to calculate a p threshold to use for each statistic. We computed this threshold in a way that allowed us to set a single false-positive rate for the entire procedure (i.e., across all five regressions). We used a time-shifting bootstrap to estimate the probability of falsely labeling each neuron as a narrowband-shift neuron and as a broadband-shift neuron. For each neuron, we generated 1,000 simulated firing rate vectors by circularly shifting the

values of the original vector by a random number of elements. For each time-shifted vector, we re-ran the original regression analyses providing, for each neuron, 1,000 sets of bootstrap regression coefficients and p values. We used these bootstraps to determine, for each neuron, two p thresholds—one for identifying narrowband-shift neurons and one for identifying broadband-shift neurons—that gave a 5% false positive rate for each designation. Although individual LFP recordings may exhibit differing correlations between the powers at neighboring frequencies, our procedure ensured that each neuron’s false-positive rate is fixed at 5% for each effect.

3.4 Results

Using recordings of 2,030 neurons from 20 neurosurgical patients (Tab. 3.1) we analyzed the relation between the firing rates of these neurons and simultaneous variations in the local field potential (LFP). In particular, we determined how moment-to-moment variations in two distinct aspects of LFP power related to simultaneous changes in the firing rates of nearby neurons (Fig. 3.8). In addition to examining LFP oscillations in narrow frequency ranges, as traditionally done in this type of work (Kahana, 2006; Fries et al., 2007), we also analyzed a novel measure of *broadband* LFP activity. Broadband activity refers to changes in the LFP power spectrum that simultaneously appear at all frequencies, rather than being limited to a narrow frequency band. We measured broadband power using a technique that measures the overall level of the LFP power spectrum while ignoring narrowband oscillatory peaks (see *Methods*). This procedure ensures that whenever we observed a neuron whose spiking was correlated with broadband power, this activity was truly related to the overall level of the LFP power spectrum, rather

than to oscillations at specific bands. We ruled out that broadband LFP power shifts are a consequence of low frequency components of the spike waveform appearing in the LFP (see *Supplemental results*). Before presenting aggregate statistics for the full dataset, we provide some sample data to illustrate the major phenomena of interest.

Figure 3.1 illustrates the relation between broadband power shifts and spiking activity recorded from an electrode in the amygdala of Patient 3. Panel A shows the normalized LFP power spectrum (black lines) and the mean broadband LFP power (red lines) for each of thirty consecutive 500 ms epochs. Panel B shows the neuron's spiking (black tick marks) and mean firing rate (blue lines) for these same epochs. Across these epochs, variations in broadband power were strongly correlated with simultaneous variations in the neuron's firing rate (Pearson's $r = 0.92, p < 10^{-12}$). Note that both broadband power and neuronal firing rate exhibited local maxima at 0.5 s, 4 s, 6.5 s, and 10 s, and both had local minima at 1.5 s and 5 s. Critically, variations in LFP power were not limited to particular narrow frequency bands, but rather appeared as overall broadband shifts in the entire power spectrum (brown lines in Panel A).

To determine whether this pattern was robust across the entire recording session, we examined the mean broadband LFP power and the mean firing rate for each of the 3,477 half-second epochs we recorded for this neuron. Data from each epoch appear as a point in Figure 3.2A, where the horizontal coordinate indicates the firing rate and the vertical coordinate indicates the normalized broadband power. Across the entire recording session, these points were clustered along the diagonal, indicating that neuronal firing rate was positively correlated with LFP broadband power (Pearson's $r = 0.6, p < 10^{-10}$). Figure 3.2B depicts this relation in a different manner, showing the mean LFP power spectra for each of five groups of epochs

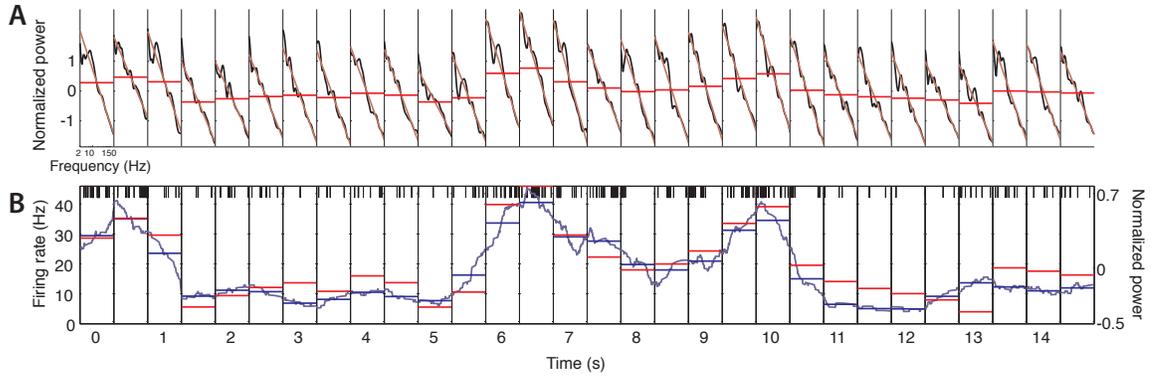


Figure 3.1. LFP power and neuronal firing time series. Each box details the activity in one 500 ms epoch. **A.** This panel illustrates how various features of the LFP change over time. In each epoch, the black lines indicate the overall LFP power spectrum, brown lines indicate robust-fit lines, and the horizontal red lines indicate mean broadband powers. **B.** This panel illustrates changes in neuronal firing rate concurrent with changes in the LFP power spectrum. Black vertical ticks represent the times when individual spikes occurred, dark blue lines indicates the smoothed firing rate (see *Methods*), and horizontal blue lines indicate mean firing rates in each epoch. Mean broadband power is shown in Panel B (horizontal red lines) on a different scale (indicated at right).

where this neuron had different firing rates (different colors in Panel A). As this neuron's firing rate increased, the LFP power spectrum exhibited a proportional upward shift at *all* frequencies.

We next sought to identify all neurons in our dataset whose firing rates varied with broadband power (as in the example above) or with narrowband power (as documented in the previous literature). Because broadband power is influenced by each narrow frequency band, disambiguating broadband and narrowband effects is critical for understanding the relation between neuronal spiking and LFP activity. To identify neurons exhibiting each of these patterns, we fit a bivariate linear regression model to the relation between firing rate and measures of both broadband and narrowband LFP power. For each neuron, we computed the firing rate for each 500 ms epoch and we also computed LFP power measured at the same electrode at five narrow frequency bands: delta (2–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–150 Hz), in addition to computing broadband power

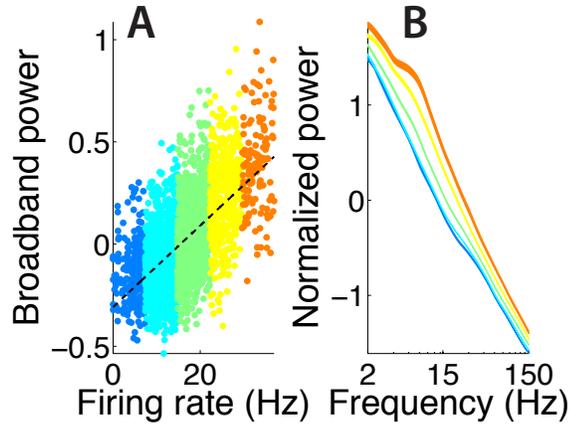


Figure 3.2. A representative neuron exhibiting a positive correlation between firing rate and broadband LFP power. **A.** Broadband power and firing rate for the neuron analyzed in Figure 3.1. Each 500 ms epoch of the recording session is represented by one colored dot. The color of each dot represents its relative firing rate. Warm colors depict epochs with high firing rates, and cool colors indicate epochs with low firing rates. The dashed black line shows an ordinary least-squares regression to these data. **B.** Average LFP power spectra for epochs with different firing rates. The same color scheme is used in both panels. As firing rate increases, the power spectrum exhibits a positive shift at all observed frequencies. The thickness of each line represents ± 1 SEM.

(see *Methods*). For each neuron we then performed a set of bivariate regressions where broadband power and the mean power in one narrow frequency band were simultaneously used to predict the neuron’s instantaneous firing rate. The two β coefficients estimated in each regression indicate the contributions of broadband activity and this particular narrowband frequency band towards each neuron’s firing rate (see *Methods*).

Combining the results of all five regressions for each neuron, we designated a neuron as a *broadband-shift neuron* when all five β coefficients for the broadband predictor were significantly different from zero in the same direction. We designated a neuron as a *narrowband-shift neuron* if, across the five regressions, either (a) one and only one narrowband β coefficient was significantly different from zero or (b) exactly two narrowband β coefficients at adjacent frequency bands (e.g. beta and gamma) were significantly different from zero in the same direction. (An

individual neuron could receive both the broadband-shift and narrowband-shift designations.) We adjusted our significance thresholds for each neuron to yield a 5% false-positive rate for each designation based on a reshuffling procedure (see *Methods*).

Figure 3.3 illustrates power spectra for neurons whose spiking (positively or negatively) correlated with narrowband LFP oscillations, neurons whose spiking correlated with broadband LFP activity, and neurons whose spiking was simultaneously influenced by both broadband and narrowband activity. Figure 3.3 also shows the proportion of neurons that fell into each category. For neurons whose firing rates were correlated with broadband or narrowband LFP patterns, the most prominent pattern in the dataset was that 34% of neurons exhibited spiking that was positively correlated with broadband LFP activity (Fig. 3.3, right column). The prevalence of this phenomenon far exceeded the level expected by chance ($p < 10^{-6}$, binomial test). In addition to this broadband effect, 29.7% of all neurons varied their spiking in relation to narrowband LFP oscillations (top and bottom rows).

The broadband and narrowband patterns were qualitatively different: the broadband shift effect was remarkably unidirectional, with 92% of all broadband-shift neurons exhibiting this effect in a positive direction. By contrast, among narrowband-shift neurons, only 66% exhibited positive correlations. Both broadband and narrowband effects often appeared simultaneously at the same recording sites; we found that 11.7% of neurons were classified as both broadband- and narrowband-shift neurons (i.e., corner entries in Fig. 3.3). This proportion is greater than would be expected if broadband and narrowband phenomena were independent ($\chi^2(1) = 5.3, p < 0.05$).

We next sought to determine whether the phenomenon of positive broadband-shift neurons was linked to any particular brain regions. We found positive

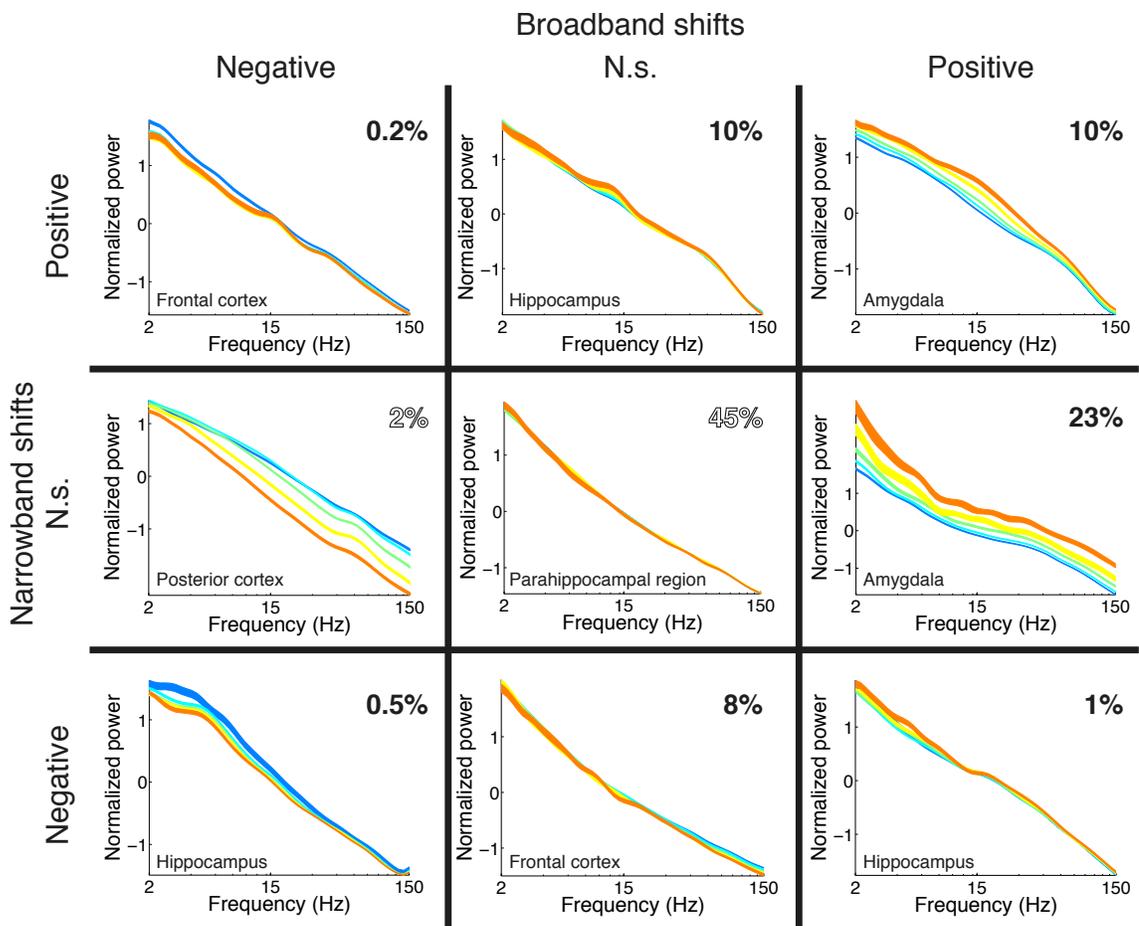


Figure 3.3. Categories of observed neurons. Each box corresponds to one of the nine possible categories into which our regression framework assigned each neuron in our dataset. The right and left columns correspond to significant positive and negative broadband shifts, respectively. The top and bottom rows correspond to significant positive and negative narrowband shifts, respectively. The middle column and middle row correspond to neurons that did not show significant broadband and narrowband shifts, respectively. The proportion of neurons placed into each category is indicated in the upper-right corner of each box. Proportions that were significantly greater than the estimated false positive rate for each category (see *Methods*) are filled in black. Each box contains power spectra (in the same format as Fig. 3.2B) of an example neuron from the corresponding category. The brain region of each example neuron is indicated in the lower-left corner of each box.

broadband-shift neurons in all observed brain regions (Fig. 3.4). To measure the prevalence of this pattern, we calculated the proportions of positive broadband-shift neurons in each of the five brain regions we examined (frontal and posterior cortices, parahippocampal region, amygdala, and hippocampus). In each of these regions, we observed significantly more positive broadband-shift neurons than expected by chance (binomial tests, $p < 10^{-6}$ for each region). However, the prevalence of this effect significantly varied between regions ($\chi^2(4) = 34.01$, $p < 10^{-6}$). In particular, the proportion of neurons exhibiting significant positive broadband shift patterns was significantly greater in medial temporal regions than in other regions (Fig. 3.5; post-hoc $\chi^2(1) = 241.56$, $p < 10^{-6}$).

The foregoing analyses focused on characterizing the broadband shift phenomenon and distinguishing it from narrowband effects. We next sought to determine whether combining our measure of broadband power with narrowband power can improve the prediction of firing rate. We performed a series of bivariate regressions with broadband power in conjunction with narrowband power, using a different narrow band (delta, theta, alpha, beta, or gamma) in each regression. Figure 3.6 shows that broadband and gamma-band power were the two dominant LFP measures that positively correlated with firing rate. The proportions of neurons exhibiting these two effects were comparable to one another and were both significantly greater than the proportions of significant positive or negative correlations observed at other frequency bands ($\chi^2(1) > 65$ and $p < 10^{-10}$ for all comparisons). In addition, we observed that in the delta, theta, and alpha bands, a greater number of neurons showed negative correlations with firing rate than positive correlations, which is consistent with recent literature (Mukamel et al., 2005). A notable feature of the broadband effects is that they are predominantly unidirectional, with nearly all (92%) significant correlations being in the positive direction. By contrast,

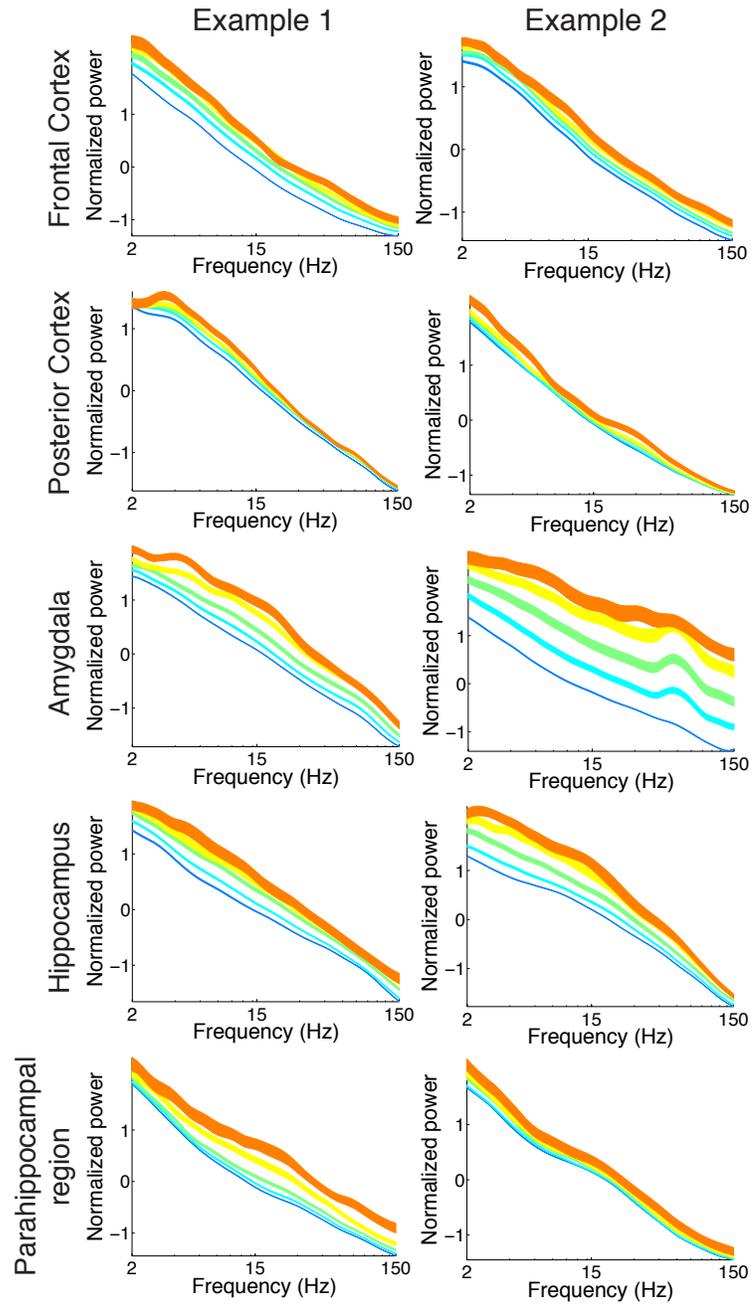


Figure 3.4. Broadband-shift neurons throughout the brain. Each row shows two example neurons from a single brain region. Within each row, each plot illustrates the power spectra from an example neuron, in the same format as Figure 3.2B. (Each of these neurons were categorized as positive broadband-shift neurons, corresponding to the middle-right box of Fig. 3.3.)

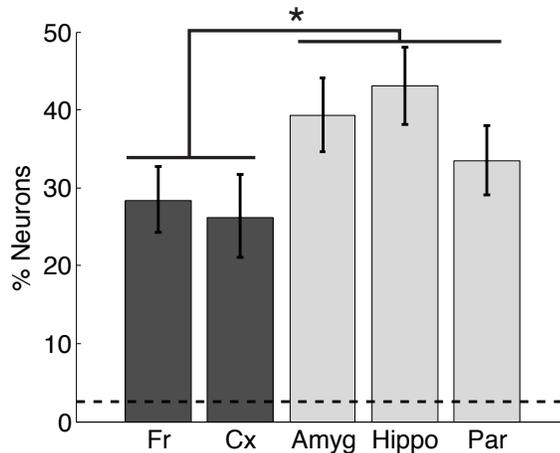


Figure 3.5. Percentage of positive broadband-shift neurons observed in each brain region. Dark gray bars correspond to neocortical regions: frontal cortex (Fr) and posterior cortex (Cx). Light gray bars correspond to medial temporal lobe structures: amygdala (Amyg), hippocampus (Hippo), and parahippocampal region (Par). Positive broadband-shift neurons were more concentrated in the medial temporal lobe than in the neocortex. The dotted horizontal black line marks the false-detection rate for positive broadband-shift neurons. Error bars indicate 95% confidence intervals based on the binomial distribution.

for neurons showing significant gamma–firing-rate correlations, only 63% showed positive correlations. Because the broadband effects were significantly more unidirectional than narrowband gamma effects ($\chi^2(1) = 171.4, p < 10^{-10}$), our results indicate that broadband power is a more specific predictor of neuronal spiking than narrowband oscillations.

To further illustrate the robust nature of the broadband phenomenon, we computed the average power spectra across all 2,030 neurons in our dataset (Fig. 3.7A). Although only 37% of the neurons included in this figure demonstrated significant firing rate-related broadband shifts in their LFP power spectra (Fig. 3.7B), there is nevertheless a clear overall effect of increasing power at all frequency bands with increased neuronal firing. However, repeating this analysis after excluding all broadband-shift neurons (left and right columns of Fig. 3.3) reveals an effect of increased gamma power with increased firing rate (Fig. 3.7C). Thus, the population

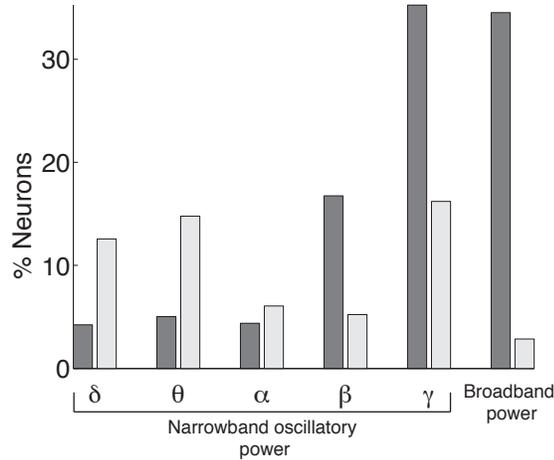


Figure 3.6. LFP components that predict firing rate. Dark gray bars indicate the percentage of neurons in each region that exhibited positive correlations between firing rate and a particular LFP feature; light gray bars show the percentage of neurons in each region that exhibited negative correlations. The bars on the left indicate the proportions of neurons whose firing rates were correlated with power in each narrow frequency band: delta (2–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–100 Hz). Each neuron may be counted in at most one direction (i.e., either positive or negative) per narrow frequency band. The bars on the right indicate the proportions of neurons whose firing rates were correlated with broadband power (i.e., broadband-shift neurons).

of non-broadband-shift neurons recovers the gamma–firing-rate relation described in the prior literature (Mukamel et al., 2005).

3.5 Discussion

We examined the relation between the firing of individual neurons and simultaneous LFP activity. In previous work, LFP recordings have typically been analyzed in terms of narrowband oscillations, which indicate that nearby neurons are synchronously spiking at a particular frequency (Buzsáki, 2006; Fries et al., 2007; Jensen, Kaiser, & Lachaux, 2007). For many of the neurons that we recorded, narrowband oscillations correlated significantly with firing rate, especially in the gamma frequency band. We also analyzed LFP recordings using a measure of

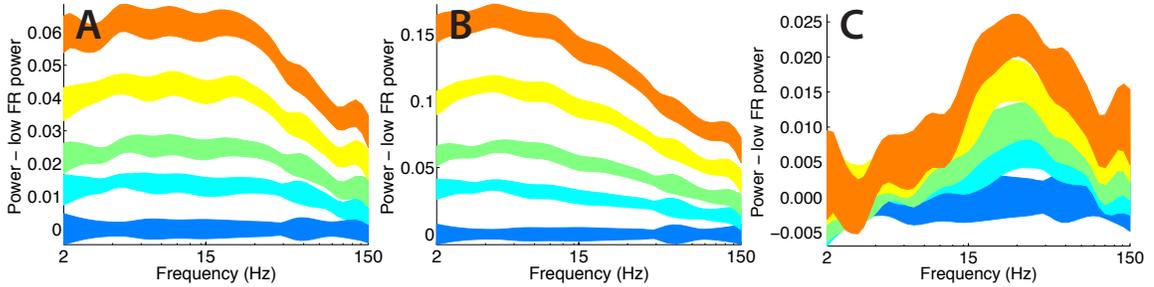


Figure 3.7. Average power spectra for different subsets of neurons. **A.** Average power spectra for all 2,030 neurons in our dataset. The power spectra are normalized by subtracting the mean of the lowest firing rate power spectrum (dark blue) from all five curves in each panel. The thickness of each line represents ± 1 SEM. As in Figure 3.2, cooler colors correspond to lower firing rates, while warmer colors correspond to higher relative firing rates. Each neuron contributes exactly once to each curve, at each frequency. **B.** Average power spectra for the 759 broadband-shift neurons. **C.** Average power spectra for the 1,271 neurons that did not exhibit the broadband-shift effect.

broadband activity, which identifies fluctuations that simultaneously appear at all frequencies. This analysis revealed that increases in the power of broadband LFP activity positively correlated with the spiking of nearby neurons. (We verified that these broadband increases in LFP power were not the result of the appearance of spike waveforms in our recordings; see *Supplemental results*). We observed this broadband effect in multiple patients and brain regions, and found that it was most prevalent in medial temporal lobe structures. Our findings indicate that both narrowband and broadband components of the LFP can be used to estimate the firing rates of nearby neurons, even though LFP recordings do not actually display the waveforms of individual action potentials. Below we discuss these findings in relation to the recent literature on narrowband and broadband LFP phenomena.

3.5.1 Narrowband effects

A wide range of studies have examined the relation between narrowband LFP oscillations and the firing of nearby neurons, identifying an array of patterns that often differ across species, frequencies, brain regions, and behaviors. However,

two common trends have emerged. Generally in neocortex, the power of low frequency ($< \sim 10$ Hz) oscillations are negatively correlated with neuronal firing rates, while the power of high frequency ($> \sim 30$ Hz) oscillations are typically positively correlated with neuronal firing rates (Mukamel et al., 2005; Rasch et al., 2008). These results comport well with the observation of task-related increases in the amplitude of high frequency activity and decreases in the amplitude of lower frequency activity (Crone, Miglioetti, et al., 1998; Crone, Miglioretti, Gordon, & Lesser, 1998; K. Miller, Leuthardt, et al., 2007). In the present study, we recorded from distributed regions throughout the brains of neurosurgical patients and found narrowband effects that were consistent with this literature. In particular, we found that LFP oscillations at low frequencies (delta, theta, and alpha bands) were negatively correlated with neuronal spiking, while high frequency LFP oscillations (beta and gamma bands) were most often positively correlated with neuronal spiking. However, we also found a significant number of neurons whose firing rates were negatively correlated with gamma oscillations (Fig. 3.6). Thus, there is considerable variability in the relation between narrowband LFP oscillations and neuronal spiking across neurons.

3.5.2 Broadband effects

Our primary finding is that neuronal spiking is positively correlated with broadband LFP power. Broadband changes in LFP power are qualitatively different from narrowband changes: whereas narrowband power changes reflect modulations in oscillatory activity, broadband power changes generally reflect modulations in the variance of the LFP time series (Annibaldi & Hopcraft, 2002). Previous research has identified broadband power changes in brain regions that are thought to be involved with particular behaviors. One example is a study by K. Miller, Leuthardt,

et al. (2007), which showed that finger and tongue movements are accompanied by highly localized broadband LFP power increases in human motor cortex (also see K. J. Miller et al., 2008; K. Miller et al., 2009). These studies show that broadband LFP shifts can be difficult to identify in practice because they often occur simultaneously with power decreases at alpha and theta frequencies (e.g. see Figs. 1 and 2 of K. Miller, Leuthardt, et al., 2007). Furthermore, oscillatory correlates of spatial and verbal memory processes are not always specific to a given frequency band, but rather appear at a very broad range of frequencies, consistent with a broadband effect (Ekstrom et al., 2007; Sederberg et al., 2007). Although few other studies have discussed their findings in terms of broadband patterns, a close inspection of the data from several studies reveals examples of behavior-related broadband LFP patterns. For example Edwards, Soltani, Deouell, Berger, and Knight (2005) report that unexpected auditory stimuli were followed by LFP changes in left-frontal and temporal cortex that appear to be broadband in nature (Fig. 3 of Edwards et al., 2005). Furthermore, the data reported by Lachaux et al. (2005) are consistent with the presence of positive broadband power shifts in the fusiform gyrus during face viewing (top panel of Fig. 2; also Fig. 5, Patient 3 in Lachaux et al., 2005). A third example can be found in Belitski et al. (2008), which examined LFP power spectra from macaque visual cortex during movie watching. Figure 2A in their manuscript shows a positive broadband shift in the LFP power spectrum during movies, relative to a baseline period. Finally, Ball et al. (2008) show data suggesting that broadband patterns can be observed non-invasively using scalp electroencephalography (EEG). For example, they show recordings from an electrode above motor cortex that exhibits a broadband power increase (coupled with a low frequency narrowband decrease) after movement (Fig. 1A in Ball et al., 2008). While each of these studies imply that broadband power indicates when a partic-

ular brain region is active, ours is the first to show that increases in broadband activity directly relate to simultaneous increases in neuronal spiking.

Given our finding that neuronal spiking is correlated with broadband power, one may wonder why many studies report task-related high frequency (gamma) modulation rather than reporting broadband effects (e.g. Lachaux et al., 2005; Belitski et al., 2008; Ball et al., 2008). As described by K. Miller, Leuthardt, et al. (2007), task-related spectral changes often appear as simultaneous low frequency decreases coupled with broadband increases in power. Due to these low frequency decreases, in practice, broadband power increases are often most visible in the gamma band. Thus many of the gamma responses reported in the literature may actually be the result of broadband power increases coupled with low frequency narrowband power decreases.

An important feature of our methodology is the robust, non-linear regression (Holland & Welsch, 1977) technique we used for distinguishing simultaneous changes in broadband and narrowband LFP activity. Because this regression is resistant to outliers, its estimate of broadband power is only minimally affected by the presence or absence of narrowband oscillations, which appear as narrowband peaks in the power spectrum. By contrast, linear methods might interpret the appearance of large narrowband oscillations as increased broadband power. This issue can be important, because broadband and narrowband changes often occur simultaneously and in opposite directions (K. Miller, Leuthardt, et al., 2007; K. Miller et al., 2009; Ball et al., 2008), which can make these effects difficult to tease apart.

3.5.3 Broader implications

A particularly exciting implication of this study is that increases (K. Miller, Leuthardt, et al., 2007) and decreases (Lachaux et al., 2008) in broadband power, which can be observed from both invasive (Edwards et al., 2005; Lachaux et al., 2005; Ekstrom et al., 2007; Sederberg et al., 2007; K. Miller, Leuthardt, et al., 2007; K. Miller et al., 2009) and non-invasive (Ball et al., 2008) recordings, may be used as a proxy for neuronal spiking. This is especially important for researchers using macroelectrode recordings to study the neural basis of human behavior, because such recordings cannot resolve single-neuron spiking, but can record these broadband shifts. To the extent that the broadband effects observed here generalize to activity recorded at broader spatial scales, one could use broadband activity to measure correlates of neuronal spiking in scalp recorded EEG and MEG signals, or even to modulate neuronal spiking using realtime feedback methods (Lachaux et al., 2007; K. Miller, den Nijs, et al., 2007). In this way, the broadband power measure proposed here may lead to new discoveries concerning how neuronal activity underlies complex human behavior and cognition.

3.6 Supplemental materials

3.6.1 Overview

This section is comprised of three main parts. First we include a supplementary figure illustrating each major component of the analysis framework used in the main text of this chapter. Next, we comment on the effect of the appearance of spike waveforms on the shape of the LFP power spectrum. Finally, we includes a table summarizing the number of broadband- and narrowband-shift neurons

identified for each patient.

3.6.2 Supplemental methods

Figure 3.8 illustrates the major components of the analysis we used to tag neurons as broadband- and narrowband-shift neurons. Synthetic data are shown in the figure in order to clearly illustrate the methods. As shown in Panel A, we first record a local field potential (*Voltage*, top sub-panel) and simultaneous spiking activity for nearby individual neurons (tick marks, bottom sub-panel). We then convolve the spike train with a Gaussian kernel to obtain a smoothed estimate of firing rate at each point in time during the recording session (*Firing rate*, bottom sub-panel). We next divide the recording session into 500 ms epochs. We thus obtain a distribution containing mean firing rates recorded during each epoch of the recording session. As shown in Panel B, we divide the distribution of firing rates into five equally spaced bins, such that each epoch is labeled according to its associated firing rate bin (indicated by different colors). We also compute mean LFP power spectra for each epoch. In Panel C, as in Figures 3.2, 3.3, 3.4, and 3.7 of the main text, the power spectra of epochs in each firing rate bin are averaged together and displayed using the same color scheme as in Panel B. We use robust regression (dotted lines) to capture the underlying $\frac{1}{f^\alpha}$ shape of the power spectra while ignoring narrowband peaks. As shown, broadband power is equal to the mean height of the robust regression line. We also measure narrowband power by averaging the normalized power in each frequency band (see *Methods*). The panel shows that when the neuron's firing rate increases, the LFP power spectrum exhibits an increase in broadband power coupled with a simultaneous decrease in power at a single narrow frequency band. In Panels D and E, each epoch is represented by a single dot. Note the overall positive relation between broadband

power and firing rate illustrated in Panel D, and the overall negative relation between narrowband power and firing rate illustrated in Panel E. As illustrated in Panel F, we assess the degree to which broadband and narrowband power inform us about the neuron's firing rate using a series of five least-squares regressions with broadband and narrowband power as regressors; each regression uses a different narrow frequency band. A neuron is tagged as a broadband-shift neuron if the β_B coefficients have the same sign and are significant in all five regressions. The neuron is tagged as a narrowband-shift neuron if only one β_f coefficient is significant, or if two β_f coefficients (corresponding to neighboring frequency bands) are significant and have the same sign. See *Methods* for details on the bootstrap procedure we used to determine the significance of each β coefficient in the regressions.

3.6.3 Supplemental results

Effect of spike waveform shape on the LFP power spectrum

In extracellular recordings, the shape of action potentials resembles an impulse function, which has spectral power at a broad range of frequencies. We thus considered the possibility that our main finding of a positive correlation between broadband power and neuronal spiking is a consequence of measuring action-potential waveforms. As described in the *Methods* section of this chapter, we used a linear interpolation to remove each spike waveform from the LFP recording. However, it is possible that our cluster-cutting algorithm missed a number of spikes, and that these undetected spikes are responsible for the broadband shifts we observed in many neurons. To test this possibility, we re-ran our entire analysis without removing spike waveforms from the LFP. If our main effect had been driven by the appearance of spike waveforms in the LFP, then keeping spike waveforms

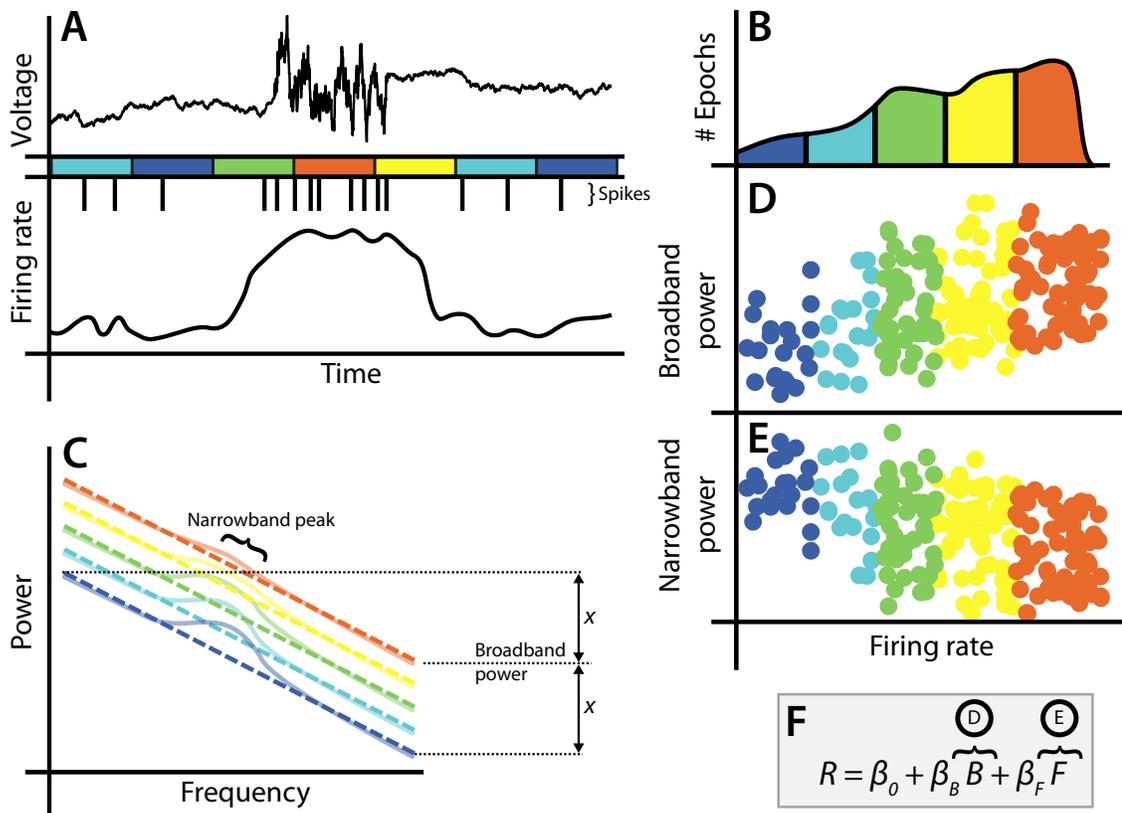


Figure 3.8. Schematic of our analysis framework. All panels contain synthetic data for illustrative purposes. **A.** This panel depicts the local field potential (*Voltage*, top sub-panel) and simultaneous spiking activity for nearby individual neurons (tick marks, bottom sub-panel). We convolve the spike train with a Gaussian kernel to obtain a smoothed estimate of firing rate at each point in time (*Firing rate*, bottom sub-panel). Rectangles are colored according to the firing rate of each epoch. **B.** This panel illustrates the distribution of mean firing rates recorded across all 500 ms epochs in the recording session. We divide the distribution of firing rates into 5 equally spaced bins (indicated by different colors). **C.** Here we show mean power spectra for epochs in each firing rate bin (solid lines). We use robust regression (dotted lines) to capture the underlying $\frac{1}{f^\alpha}$ shape of the power spectra while ignoring narrowband peaks. We also measure narrowband power by averaging the normalized power in each frequency band (see *Methods*). **D, E.** These panels illustrate the relation between firing rate and broadband power (Panel D) or narrowband power in a single frequency band (Panel E). Each epoch is represented by a single dot in each panel. **F.** We use a series of least-squares regressions to determine whether the neuron is a broadband- and/or narrowband-shift neuron.

in the LFP would be expected to increase the magnitude of the broadband-shift effect. Instead, here we found that the number of positive broadband-shift neurons actually decreased (from 697 to 602), rather than increased, without spike removal. This suggests that the appearance of the spike waveform in the LFP does not account for the positive broadband shifts we observed.

3.6.4 Patient table

Patient	Sessions	Observed neurons	BB+	BB-	$\delta+$	$\delta-$	$\theta+$	$\theta-$	$\alpha+$	$\alpha-$	$\beta+$	$\beta-$	$\gamma+$	$\gamma-$
1	1	73	16	9	1	2	0	2	0	0	7	0	3	3
2	2	103	41	5	1	1	0	1	0	0	4	1	13	9
3	1	31	15	0	0	0	1	0	2	0	4	0	5	1
4	3	183	46	8	0	2	1	1	6	0	18	0	8	24
5	2	56	22	0	3	1	1	0	1	0	6	1	9	0
6	3	161	52	0	0	2	3	5	0	2	12	1	18	8
7	2	92	37	3	1	1	0	0	1	0	6	0	14	7
8	1	81	25	3	0	0	0	0	1	0	4	0	12	12
9	3	189	40	5	0	0	3	1	0	1	4	0	24	18
10	3	164	48	3	1	2	0	2	1	1	4	0	14	20
11	3	180	42	6	1	2	0	1	5	1	13	0	22	13
12	4	195	58	5	0	6	0	0	1	0	4	4	39	13
13	2	113	46	1	2	0	1	2	4	1	6	0	26	7
14	3	80	39	2	3	1	1	0	1	0	1	2	11	3
15	4	80	45	1	0	1	0	0	1	0	5	0	17	1
16	1	22	16	0	0	0	0	0	2	0	0	0	3	0
17	1	41	17	0	0	0	0	0	0	1	1	1	2	2
18	2	63	20	6	5	0	0	0	0	0	3	0	8	8
19	3	85	53	5	1	1	0	0	1	0	6	0	13	5
20	2	38	19	0	0	0	0	0	0	0	4	0	6	3
Total	46	2,030	697	62	19	22	11	15	27	7	112	10	267	157

Table 3.1. Summary of observed neurons. Columns 4–15 indicate, for each patient, the number of identified positive (+) and negative (–) broadband-shift neurons (BB), and for each frequency band (δ , θ , α , β , and γ) the number of narrowband-shift neurons. Each neuron appears in at most one broadband column (BB+ or BB–) and at most two narrowband columns of the same sign, for neighboring frequencies (e.g. $\beta+$ and $\gamma+$).

Chapter 4

Temporal and frontal networks reveal how conceptual memories are organized

Jeremy R. Manning, Michael R. Sperling, Ashwini Sharan, Emily A. Rosenberg, & Michael J. Kahana

4.1 Abstract

How the brain represents, organizes, and retrieves semantic information is one of the great enduring questions in science and philosophy. We approached this question by analyzing electrocorticographic (ECoG) recordings taken as 46 neurosurgical patients studied and freely recalled lists of words. We first identified semantic components of brain activity that varied systematically with the meanings of each presented word, as defined by latent semantic analysis (LSA; Landauer & Dumais, 1997). We then examined these same semantic components of brain activ-

ity as the patients recalled the words. The degree to which brain activity correlated with semantic similarity during recall predicted participants' tendency to successively recall similar words. Whereas previous studies have shown that semantic representations are conserved across people (Shinkareva et al., 2008; Mitchell et al., 2008), our work shows that differences in these neural representations reveal how individuals organize and retrieve their memories.

4.2 Narrative

Psychologists have long been interested in the role of semantic or conceptual representations on the retrieval of remembered words (Tulving & Pearlstone, 1966; Glanzer, Koppenaal, & Nelson, 1972). The distributed memory hypothesis predicts that concepts are represented by overlapping sets of semantic features, each corresponding to a particular attribute (Semon, 1923; Guthrie, 1935; Hooke, 1969; Plaut & McClelland, 2010). Activating the representation of a concept entails activating its constituent semantic features. Because semantically related concepts share semantic features, activating a given concept will, in turn, partially activate representations of other similar concepts. Consistent with this hypothesis, when instructed to recall words in any order, participants often cluster their recalls by natural categories, even when the presentation order is randomized (D'Agostino, 1969). In addition, the delay between successive recalls of items from the same natural category tends to be shorter than between items from different categories (Pollio, Kasschau, & DeNise, 1968). Further, presenting a set of words that are all strong semantic associates of an unrepresented critical word often leads to "false recall" of the critical word (Deese, 1959; Roediger & McDermott, 1995).

Whereas the above behavioral studies provide indirect evidence that the brain

organizes conceptual representations by their semantic features, functional magnetic resonance imaging (fMRI) studies over the past decade have begun to directly examine the neural representations of concepts. Early studies showed that viewing (Haxby et al., 2001; Cox & Savoy, 2003; Shinkareva et al., 2008) or thinking about (Polyn, Natu, Cohen, & Norman, 2005) words from similar semantic categories evokes similar blood-oxygen-level dependence (BOLD) signatures. Over the past several years, researchers have gone beyond the neural representations of broad categories, and have now begun to study the neural representations of individual concepts. This work has shown that a word's semantic properties can be used to make accurate predictions about its BOLD signature (Mitchell et al., 2008; Just, Cherkassky, Aryal, & Mitchell, 2011), and that the neural response to a viewed object is attenuated when a similar object is viewed first (Yee, Drucker, & Thompson-Schill, 2010). An additional finding of this line of research is that the neural representations of concepts appear to be reasonably well-conserved across individuals (Mitchell et al., 2008; Shinkareva et al., 2008; Just et al., 2011). In the present study, we first use ECoG recordings to study the representations of individual concepts as they are studied and recalled during a memory task. We then ask whether (and how) variations in conceptual representations across individuals were manifested in behavior.

We examined ECoG recordings from 46 neurosurgical patients who were implanted with subdural electrode arrays and depth electrodes during presurgical evaluation of a treatment for drug-resistant epilepsy. We analyzed the recorded ECoG signals in terms of specific time-varying oscillatory components of neural activity (Fig. 4.1). Previous research has implicated oscillatory activity at a variety of frequencies during both memory encoding and retrieval (Fell et al., 2001; Paller & Wagner, 2002; Osipova et al., 2006; Sederberg et al., 2007) and in the represen-

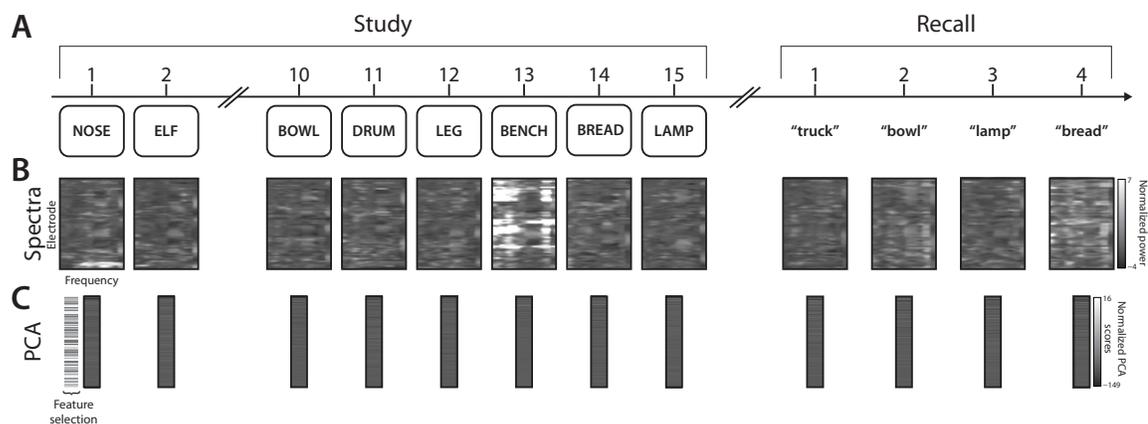


Figure 4.1. Illustration of behavioral and electrophysiological methods. **A.** After studying a list of 15 words and performing a brief distraction task, a participant recalls as many words as he can remember, in any order. **B.** During each study presentation, and just prior to each recall event, we calculate the z-transformed oscillatory power at each recording electrode for each of 50 log-spaced frequencies between 2 and 100 Hz ($2 \times 10^{0.0347x}$ for $x \in \{0, \dots, 49\}$). **C.** We use principal components analysis (PCA) to find a smaller number of orthogonal dimensions that jointly account for a large proportion of the variation in the data shown in panel B. We select principal components that vary with the meanings of the presented words for further analysis (see *Supplemental materials*).

tation of individual letters (J. Jacobs & Kahana, 2009). As the ECoG signals were recorded, patients volunteered to participate in a *free recall* memory experiment, in which they studied lists of common nouns and then attempted to recall them verbally in any order following a brief delay (see *Methods*). Over several sessions, each participant studied and recalled words from dozens of different word lists.

Our general approach is illustrated in Figure 4.1. For each study and recall event, we used Morlet wavelets to construct an N -dimensional vector containing, for each electrode, estimates of mean oscillatory power at 50 log-spaced frequencies between 2 and 100 Hz ($2 \times 10^{0.0347x}$ for $x \in \{0, \dots, 49\}$) during each study event (200 – 1,600 ms relative to the word’s appearance on screen) and recall event (-1,000 – 0 ms relative to the start of vocalization). We then used principal components analysis (PCA) to distill these highly correlated features into a smaller number of orthogonal components. We used the Kaiser criterion to choose, for each participant, the principal components that explained a substantial proportion of the variance in the

original N -dimensional vectors (Kaiser, 1960).

Based on the aforementioned neural and behavioral studies, we hypothesized that some components of neural activity might vary systematically with the semantic properties of studied words. Indeed, we found evidence for such components in all 46 participants that we examined (Tab. 4.2). If these semantic features (henceforth feature vectors) truly represent the identities of studied items, one should observe a positive correlation between neural and semantic similarity each time the participants think of the studied words — not only during study (when the feature vectors were identified), but also during recall. We thus selected the feature vectors for further study. For each participant, we computed the cosine similarity (normalized dot product) between the feature vectors from each pair of recalls they made (if a participant made n recalls in a given list, then we computed $n^2 - n$ cosine similarities for that set of recalls). We also computed the semantic similarity (LSA $\cos\theta$) between each pair of recalled words. We regressed the neural similarities on the semantic similarities and obtained a t -value for the β coefficient of the regression for each participant. A t -test on the distribution of t -values showed that the relation between neural and semantic similarity was reliably maintained during the 1 second interval leading up to vocalization of a recalled word ($t(45) = 2.28, p = 0.03$), even though the words were neither being displayed onscreen nor vocalized by the participant during the interval being measured. Rather, the neural patterns recorded just prior to recall reflect internally generated semantic representations.

A number of studies have shown that certain brain regions play a particularly strong role in representing semantic information during cognition. For example, Broca's and Wernicke's areas in the inferior frontal and superior temporal lobe, respectively, are well-known for their role in semantic processing (Demonet et al., 1992); the prefrontal cortex and temporal lobe have been found to support

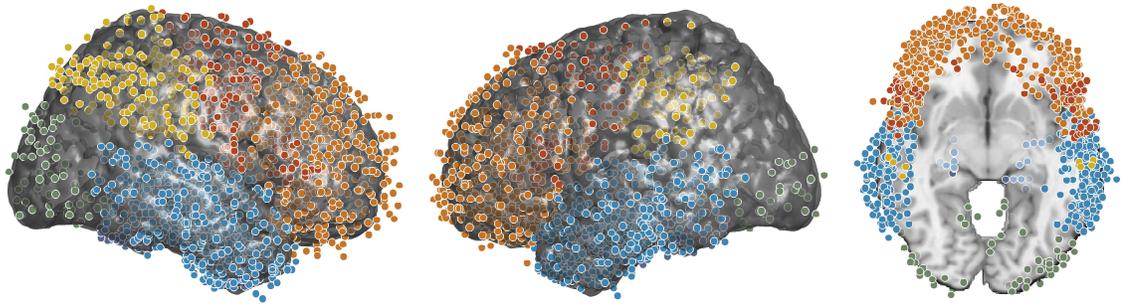


Figure 4.2. Regions of interest (ROIs). Each dot marks the location of a single electrode from our dataset in Montreal Neurological Institute (MNI) space. We divided our dataset into seven ROIs: temporal lobe (Temp; 1,358 electrodes), medial temporal lobe (MTL; 524 electrodes), hippocampus (Hippo; 130 electrodes), frontal lobe (FR; 1,369 electrodes), prefrontal cortex (PFC; 984 electrodes), parietal lobe (Par; 336 electrodes), and occipital lobe (Occ; 102 electrodes).

semantic clustering during free recall (Long, Öztekin, & Badre, 2010); and the occipital lobe is activated during visual recall (Le Bihan et al., 1993). To test whether the semantic information we measured was specific to particular brain regions, we repeated our semantic feature selection process for electrodes in each of the following brain regions: temporal lobe, medial temporal lobe, hippocampus, frontal lobe, prefrontal cortex, parietal lobe, and occipital lobe (Fig. 4.2). We found that semantic features (selected during encoding and measured just prior to recall) varied with the meanings of recalled words in the temporal lobe ($t(45) = 2.88, p = 0.006$), frontal lobe ($t(41) = 3.56, p = 0.001$), prefrontal cortex ($t(41) = 3.33, p = 0.002$), and occipital lobe ($t(26) = 3.64, p = 0.001$). Neither the medial temporal lobe ($t(41) = -0.17, p = 0.86$), hippocampus ($t(18) = 0.47, p = 0.64$), nor the parietal lobe ($t(31) = -0.16, p = 0.88$) showed such patterns.

To localize the semantic representations, we performed, for each pair of regions, paired t -tests between the distributions of t -values from the regressions of neural on semantic similarity, for participants with electrodes in both regions. When compared with either medial temporal or hippocampal electrodes, we found that temporal, frontal, and occipital electrodes all showed reliably stronger semantic effects

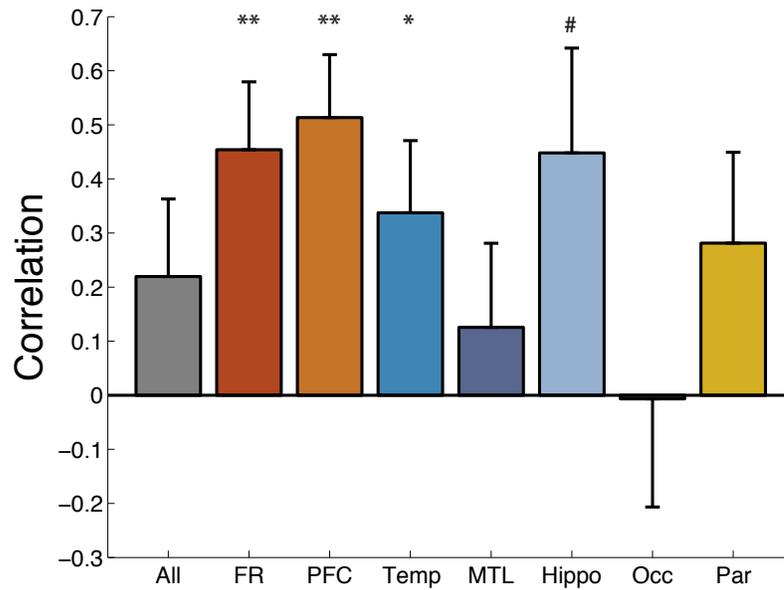


Figure 4.3. Neural activity predicts clustering during recall. The height of each bar indicates the correlation, across participants, between semantic clustering during recall (see *Supplemental materials*) and the correlation between neural and semantic similarity. Each bar corresponds to a single ROI (Fig. 4.2). The bar marked “All” indicates this correlation for all electrodes in the dataset. Error bars denote \pm SEM, estimated as $\frac{1-r^2}{\sqrt{n-1}}$. The symbols indicate the p -values for each correlation: # denotes $p < 0.1$, * denotes $p < 0.05$, and ** denotes $p < 0.01$. See Figure 4.4 for further details.

(p 's < 0.05 for all tests). Note that although the temporal lobe electrodes we analyzed were a superset of the medial temporal lobe electrodes (which were, in turn, a superset of the hippocampal electrodes), the temporal lobe as a whole nonetheless exhibited stronger effects than either the medial temporal lobe or hippocampus alone. We also found that prefrontal electrodes exhibited stronger semantic effects than parietal electrodes ($t(28) = 2.28, p = 0.03$). Taken together, our results suggest that semantic information is most strongly represented by temporal, prefrontal, and occipital networks, and less so by medial temporal and parietal networks.

Having identified a set of candidate semantic representations, we hypothesized that subtle differences in the structure of these neural patterns would be reflected in recall behavior. For example, suppose that participant *A* organized her conceptual memories primarily based on the meanings of the studied words, whereas participant *B* organized his conceptual memories based on the appearances of the objects represented by the words. One might expect participant *A* (but not *B*) to successively recall words such as CHERRY and BANANA, if both words had appeared on the studied list. By contrast, participant *B* would be more likely to successively recall APPLE and BALL. According to the distributed memory hypothesis, participant *A*'s concept representations should be comprised primarily of features related to the meanings of words, whereas participant *B*'s should be comprised primarily of features related to the appearance of the objects the words represent. More generally, we hypothesized that participants exhibiting strong correlations between neural and semantic similarity would also semantically cluster their recalls. To test this hypothesis, we examined the t -values from the regressions of neural on semantic similarity (one t -value per participant). We also computed the degree to which each participant clustered their recalls by semantic similarity using their semantic factor (see *Supplemental materials*). We then computed the correlation between these

measures of neural and behavioral organization separately for electrodes in each region we examined (Fig. 4.3). We found significant correlations between neural and behavioral clustering in the temporal lobe ($r = 0.33, p = 0.02$), frontal lobe ($r = 0.42, p = 0.006$), and prefrontal cortex ($r = 0.49, p = 0.001$). Although hippocampal electrodes showed no evidence of a semantic representation when we combined data from all of the participants, participants whose hippocampal activity varied with the meanings of recalled words also tended to cluster their recalls by their meanings ($r = 0.45, p = 0.06$). On the other hand, although the occipital lobe showed strong evidence of a semantic representation we found no evidence that this representation had any influence on recall order ($r = 0.00, p = 0.99$). This indicates that occipital lobe structures represent the meanings of the studied items but do not organize these conceptual representations in memory.

Given that the analyses we have described thus far rely on a single measure of semantic similarity, our finding that participants whose neural activity varies with the meanings of words also tend to semantically cluster their recalls is consistent with two interpretations. We have suggested above that some participants might organize conceptual memories semantically whereas others might use different organizational schema. However, an alternative interpretation is that all participants organize conceptual memories according to their meanings — but that the particular semantic similarity measure we used in our analyses, LSA, might reflect the true semantic representations for some participants more accurately than for others. To distinguish between these two interpretations, we utilized a second measure of semantic similarity based on word association spaces (WAS; Steyvers, Shiffrin, & Nelson, 2004). Whereas LSA similarity is derived from an automated analysis of a large text corpus (Landauer & Dumais, 1997), WAS similarity is derived from a series of free association experiments in which participants were given a cue item and

responded with the first word that came to mind (Nelson, McEvoy, & Schreiber, 2004; Steyvers et al., 2004). For the words presented to participants in our study, these two measures of semantic similarity are reliably, though weakly, correlated ($r = 0.26, p < 10^{-3}$; Spearman's $\rho = 0.18, p < 10^{-3}$). We found that the degree to which participants clustered their recalls by LSA and WAS similarity (semantic factor; see *Supplemental materials*) were reliably correlated ($r = 0.30, p = 0.04$). Further, the degree to which participants exhibited correlations between their neural and LSA similarity also predicted the degree to which their neural and WAS similarity were correlated ($r = 0.40, p = 0.006$). These analyses suggest that the neural signatures of semantic organization we observed reflect semantic information in general rather than a single measure of semantic similarity.

We have focused in the present study on the degree to which participants semantically organize their memories. However, other behavioral studies suggest that participants also organize remembered information according to the temporal or spatial contexts in which the information was learned (Howard & Kahana, 1999; Smith & Vela, 2001). One could easily adapt our general approach to study other forms of conceptual organization by adjusting the way in which conceptual features were selected (Fig. 4.1C) and by measuring the extent to which individual participants relied on the organizational scheme of interest (analogous to semantic factor in the present study). One such analysis is presented in Chapter 5.

The preceding analyses advance our understanding of how our brains represent, store, and retrieve conceptual information in two critical areas. First, in support of the distributed memory hypothesis, we demonstrate that thinking about semantically related words evokes similar patterns of neural activity in temporal, prefrontal, and occipital cortex. This indicates that concepts are represented by distributed networks in these brain areas. Second, whereas previous studies have

shown that the neural representations of individual concepts are conserved across individuals, we found that individual differences in how concepts are represented in temporal and frontal (but not occipital) cortex can be used to predict the order in which an individual will recall a list of words. This indicates that temporal and frontal networks organize conceptual information by representing relationships amongst stored concepts.

4.3 Methods

4.3.1 Participants

We tested 46 native-English-speaking patients with drug-resistant epilepsy who had arrays of subdural and/or depth electrodes surgically implanted for one to four weeks to localize the sites of seizure onset (see Tab. 4.2). The clinical team determined the placement of these electrodes with the goal of localizing suspected epileptogenic foci and identifying functional regions to be avoided in surgery. Our research protocol was approved by the appropriate institutional review boards and informed consent was obtained from the participants and their guardians. Data were collected as part of a long-term multicenter study with previously published articles describing separate analyses conducted on subsets of these data.

4.3.2 Behavioral methods

Participants studied lists of 15 or 20 high frequency nouns for a delayed free recall task. Following a fixation cue, the computer displayed each word for 1,600 ms followed by a 800–1,200 ms blank inter-stimulus interval. Each word was displayed at most once within a single testing session. For 18 s following list presentation,

participants solved a series of single-digit addition problems of the form $A + B + C = X$. Participants were then given 45 s to recall list items in any order. Vocal responses, digitally recorded during the trial, were scored for subsequent analysis. Participants recalled $22.1 \pm 1.1\%$ (mean \pm SEM) of the studied words. Repetitions and incorrect recalls ($28.0 \pm 3.0\%$ of all responses) were excluded from our analyses, as were responses that occurred within 1 s of a prior vocalization.

4.3.3 Recording methods

Subdural grids or depth electrodes (Ad-Tech Inc., Racine, Wisconsin) were implanted by neurosurgical teams solely for clinical purposes. The locations of the electrodes were determined by means of co-registered post-operative computed tomography and pre-operative magnetic resonance imaging (MRI), or from post-operative MRIs, by an indirect stereotactic technique and converted into Montreal Neurological Institute (MNI) coordinates. Electrocorticographic (ECoG) signals were recorded referentially using a Bio-Logic, XLTek, Neurofile, or Nicolet EEG digital video-EEG system. Depending on the amplifier, signals were sampled at 200, 256, 500, 512, or 1,024 Hz. Several hospitals applied bandpass filters to the recorded signals prior to writing to disk (Tab. 4.1). Where applicable, frequencies outside of the filtered range were excluded from further analysis. Data were subsequently notch-filtered with a Butterworth filter with zero phase distortion at 60 Hz to eliminate electrical line and equipment noise. ECoG signals and behavioral events were aligned using synchronization pulses sent from the testing computer (mean precision < 4 ms).

4.4 Supplemental materials

4.4.1 Overview

This section is comprised of three main parts. First, we include a table summarizing details about the recording setup at each of our collaborating hospitals. Next, we include additional details on our analytic methods not included in the main text. Finally, we include a table summarizing information about the individual patients.

4.4.2 Recording methods

HOSPITAL	HOSP	LOWER	UPPER
Brigham & Women's Hospital, Boston, MA	BW	0.5 Hz	60 Hz
Children's Hospital Boston, Boston, MA	CH	0.3 Hz	50 Hz
Thomas Jefferson University Hospital, Philadelphia, PA	TJ	$-\infty$	∞
Hospital of the University of Pennsylvania, Philadelphia, PA	UP	$-\infty$	∞

Table 4.1. Bandpass filters employed by our collaborating hospitals. Hospital codes (HOSP) are referenced in Table 4.2. The LOWER and UPPER columns denote the lower and upper limits of the bandpass filters, respectively. Frequencies outside of the bandpassed range were excluded from further analysis.

4.4.3 Analysis methods

Quantifying the degree of semantic clustering

The *Narrative* and Figure 4.4 reference a measure of the degree to which individual participants clustered their recalls by LSA similarity, termed the semantic factor. This measure has been described previously (Polyn, Norman, & Kahana, 2009), and is calculated as follows.

For each recall transition we create a distribution of semantic similarity values (LSA $\cos\theta$; Landauer & Dumais, 1997) between the just-recalled word and the set of words that have not yet been recalled. We next generate a percentile score by comparing the semantic similarity value corresponding to the next item in the recall sequence with the rest of the distribution. Specifically, we calculate the proportion of the possible similarity values that the observed value is greater than, since strong semantic clustering will cause the observed similarity values to be larger than average. When there is a tie, we score this as the percentile falling halfway between the two items. If the participant always chose the closest semantic associate, then their semantic factor would be 1. A semantic factor of 0.5 indicates no effect of semantic clustering. Each patient was assigned a semantic factor by taking the average of the percentile scores across all observed recall transitions.

Selecting candidate semantic features

To identify candidate components of the semantic representation for a given recording session, we selected PCA-derived features of the neural representation (Fig. 4.1C) as follows. We first generated the set of all possible pairings of presented words (for a list of n items, this set of all pairs contains $n^2 - n$ items). For each pair, we computed the LSA $\cos\theta$, a measure of semantic similarity, between those two words. We generated an $(n^2 - n)$ -dimensional vector, \vec{a} , containing these semantic similarity values. Separately for each feature, x , we computed the absolute difference in the value of x during each pair of presentations. We next defined \vec{b}_x as the $(n^2 - n)$ -dimensional vector containing the neural difference values for feature x . We computed the Pearson's correlation between \vec{a} and $-\vec{b}_x$ to obtain a single correlation coefficient, r_x , and an associated p -value for each list. We then combined

these correlation coefficients across lists into a summary correlation measure, \bar{r}_x :

$$\bar{r}_x = F^{-1}\left(\sum_{i=1}^L F(r_{x,i})\right),$$

where $r_{x,i}$ was the Pearson's correlation coefficient for feature x during list i , $F()$ was the Fisher z-prime transformation:

$$F(r) = \frac{\ln(1+r) - \ln(1-r)}{2},$$

and $F^{-1}()$ was the inverse of $F()$:

$$F^{-1}(z) = \frac{e^{2z-1}}{e^{2z+1}}.$$

In this way, if $r_{x,i}$ has large positive values across all lists, \bar{r}_x will have a large positive value. Similarly, if $r_{x,i}$ is negative across all lists, \bar{r}_x will have a large negative value. If $r_{x,i}$ is sometimes positive and sometimes negative (with approximately equal probability), \bar{r}_x will take on a value near zero. (Note that $-1 \leq r_{x,i}, \bar{r}_x \leq 1$.)

We also obtained a p -value, \bar{p}_x , associated with \bar{r}_x by applying the inverse Normal transformation to the p -values associated with the correlation coefficient for each list. We then summed across the transformed p -values and evaluated the cumulative Normal distribution function at this sum to obtain \bar{p}_x . We selected features with $\bar{r}_x > 0$ and $\bar{p}_x < 0.05$ for further analysis (see *Narrative*).

Detailed region of interest results

Figure 4.3 summarizes the degree to which neural activity in these brain regions reflects the order in which words are recalled. Figure 4.4 provides a more in-depth look at these effects in each region of interest.

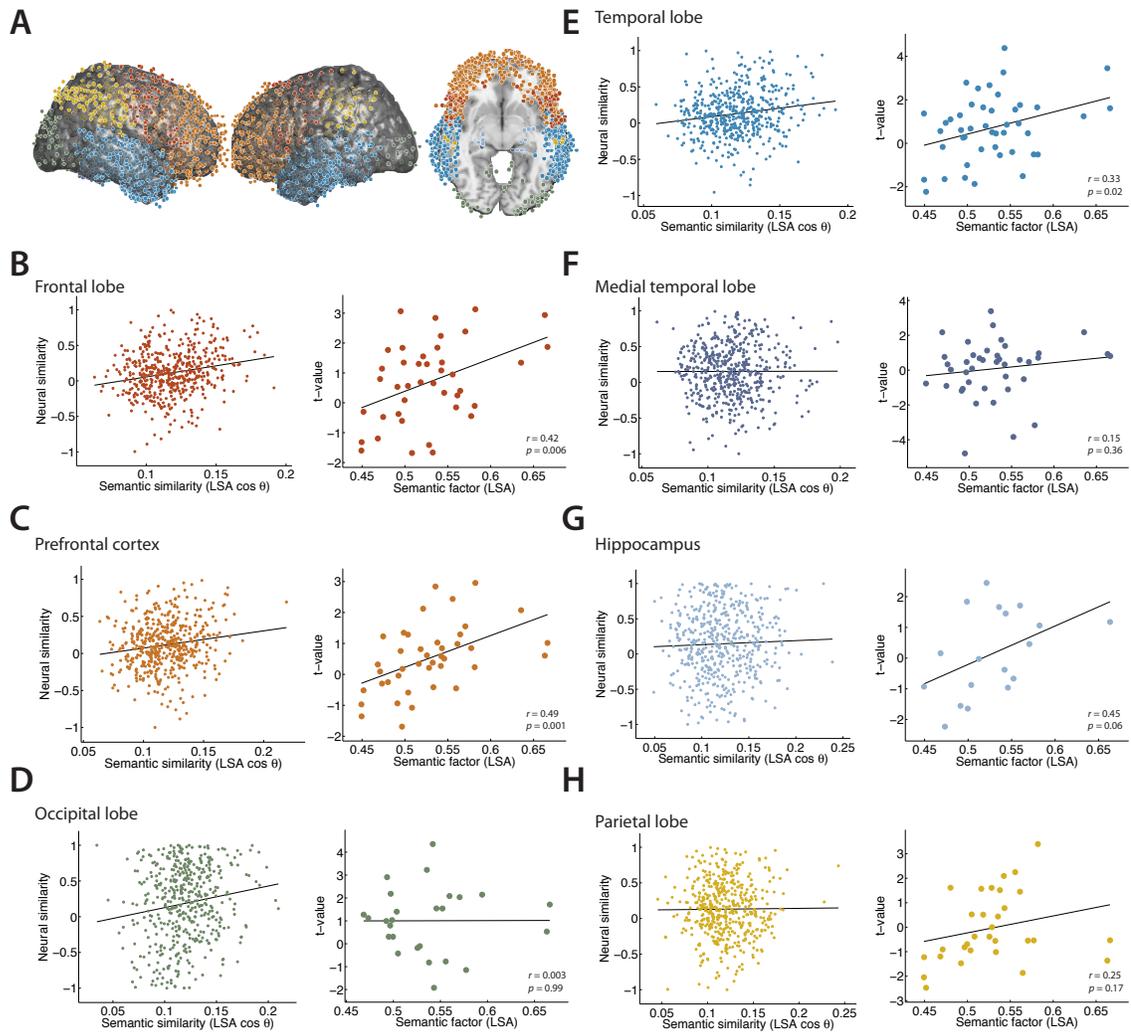


Figure 4.4. Detailed region of interest results. **A.** Each dot marks the location of a single electrode from our dataset in MNI space. (This panel is reproduced from Fig. 4.2 in the main text, for reference.) **B – H.** Each panel corresponds to a single region of interest. The left subpanels show the relation between neural and semantic similarity during recall. For each list, we computed the LSA $\cos \theta$ between each pair of recalled words (semantic similarity, x -axis). We also computed the cosine similarity between the neural semantic feature vectors for these same words (neural similarity, y -axis). Here we show the mean neural similarity in each of 500 equally sized bins comprising the full semantic similarity distribution. The right subpanels show the relation between neural clustering (t -value) and semantic clustering (semantic factor); each dot represents a single participant.

4.4.4 Supplemental patient information

ID	HOSP	AGE	SEX	HAND	ELC	FEA	L LEN	SES	LST	REC	REP	PLI	ELI
1	BW	33	F	R	64	59	20	1	15	55	2	26	12
2	BW	51	F	R	40	89	20	1	15	66	2	8	1
3	BW	32	M	R	32	48	15	3	39	212	5	9	4
4	BW	40	M	R	96	62	15	2	20	82	5	38	28
5	BW	44	M	R	16	24.5	15	2	20	58	1	12	19
6	BW	27	M	R	64	21	15	2	20	76	49	12	3
7	BW	38	M	R	104	72.67	15	3	30	136	3	30	15
8	CH	13	F	R	64	114	20	1	12	59	0	2	1
9	CH	12	F	R	104	146	20	1	15	39	0	0	2
10	CH	15	M	L	128	85.67	20	3	30	90	1	12	8
11	CH	17	M	R	64	92	20	3	45	178	20	27	17
12	CH	15	M	R	123	157	20	1	15	86	3	6	3
13	CH	11	M	R	104	148	20	2	30	104	3	2	3
14	CH	14	F	R	72	131	20	1	15	104	2	7	5
15	CH	8	F	R	86	137	20	2	30	159	5	18	10
16	CH	17	M	R	84	123	20	1	14	30	2	9	12
17	CH	17	M	L	124	110.5	20	4	60	116	2	104	13
18	CH	20	F	R	128	97	15	2	24	114	2	9	4
19	CH	14	M	R	94	70	15	3	30	94	0	18	13
20	CH	17	M	L	80	42.5	15	2	20	14	0	10	22
21	CH	19	F	R	125	54.5	15	2	17	47	2	10	1
22	CH	16	M	R	156	126	15	1	16	76	1	4	1
23	CH	12	M	L	84	67.5	15	2	20	52	9	12	25
24	CH	13	M	R	72	65.25	15	4	40	200	3	4	2
25	TJ	25	M	R	62	111.33	15	3	48	232	3	6	1
26	TJ	40	F	R	94	99.75	15	4	64	164	8	54	43
27	TJ	39	M	L	56	79	15	1	16	53	1	8	20
28	TJ	34	F	R	112	113.2	15	10	154	513	7	110	24
29	TJ	44	M	R	126	74	15	1	13	31	1	7	6
30	TJ	43	M	R	80	81.75	15	4	64	232	139	71	29
31	TJ	21	M	R	122	138	15	3	48	145	1	32	47
32	TJ	56	M	R	50	67	15	2	48	120	5	98	43
33	TJ	20	M	R	160	99	15	3	42	167	0	13	18
34	TJ	41	M	R	98	125	15	2	32	100	2	18	98
35	TJ	34	F	R	90	89.25	15	4	51	204	49	34	207
36	UP	38	M	R	62	53	15	4	40	135	3	68	24
37	UP	30	M	R	86	65	15	2	20	54	5	24	21
38	UP	43	M	R	66	36.33	15	3	18	31	22	12	33
39	UP	36	M	R	88	78.75	15	4	40	70	6	114	50
40	UP	25	M	R	62	60.25	15	4	40	135	2	1	2
41	UP	18	F	R	76	77.33	15	3	30	104	5	6	3
42	UP	27	F	R	48	75.5	15	2	32	104	2	43	20
43	UP	55	F	L	80	106	15	2	32	81	11	61	24
44	UP	18	M	A	100	90.67	15	3	48	253	7	8	3
45	UP	38	F	R	86	98	15	1	16	48	14	3	73
46	UP	40	M	R	58	93.75	15	4	64	304	1	14	9

Table 4.2. Patient and task information. This table provides the hospital (HOSP) at which each patient’s data were collected (see Tab. 4.1), as well as each patient’s age (AGE), gender (SEX), handedness or language mapping (HAND), number of implanted electrodes (ELC), and mean number of features selected for analysis across all sessions for that patient (FEA). Information about the task includes the list length (L LEN) used for each participant, number of testing sessions (SES), and the number of lists each participant encountered across all sessions (LST). Performance information includes the total number of correct recalls across all lists (REC), the total number of repeated recalls (REP), and the total number of incorrect recalls, which includes recalls of previously presented items (prior list intrusions, or PLIs) and recalls of items which were never presented (extra-list intrusions, or ELIs). In total the 46 patients contributed 3,970 electrodes and 4,055 selected features, studying 24,760 items presented in 1,552 lists.

Chapter 5

Oscillatory patterns in temporal lobe reveal context reinstatement during memory search

Jeremy R. Manning, Sean M. Polyn, Gordon Baltuch, Brian Litt, &
Michael J. Kahana

5.1 Abstract

Psychological theories of memory posit that when people recall a past event, they not only recover the features of the event itself, but they also recover contextual features representing information about neighboring events and the thoughts they evoke. The ability to associate items with their temporal context and to recover contextual information during recall has been termed episodic memory. We sought to determine whether contextual reinstatement in human memory may be observed in electrical signals recorded from the human brain. By analyzing electrocortico-

graphic recordings taken as 69 neurosurgical patients studied and recalled lists of words, we were able to uncover a neural signature of context reinstatement. Upon recalling a studied item, we found that the pattern of brain activity was not only similar to the pattern observed when the item was studied, but was also similar to neighboring list items, with similarity decreasing reliably with positional distance. The degree to which individual patients displayed this neural signature of context reinstatement was correlated with their tendency to successively recall neighboring list items. These effects were particularly strong in temporal lobe recordings. Our findings show that recalling a past episode evokes a neural signature of the temporal context in which the episode occurred, thus pointing to a neural basis for episodic memory.

5.2 Introduction

The pivotal distinction between memory for facts (semantic memory) and memory for episodes or experiences (episodic memory) has been argued to reflect, at least in part, the reinstatement of a gradually changing context representation that reflects not only external conditions, but also an ever-changing internal context state (Bower, 1972; Tulving, 1983). According to this view, the unique quality of episodic memory is that, in remembering an episode, we partially recover its associated mental context, and that this context information conveys some sense of when the experience took place in terms of its relative position along our autobiographical timeline.

A number of laboratory memory tasks rely on episodic memory, including experimenter-cued tasks (e.g. item recognition and cued recall) and self-cued tasks (e.g. free recall). Performing these episodic memory tasks requires distinguishing

the current list item from the rest of one's experience. According to early theories of episodic memory (e.g. Estes, 1955; Anderson & Bower, 1972) context representations are composed of many features that fluctuate from moment to moment, gradually drifting through a multidimensional feature space. During recall, this representation forms part of the retrieval cue, enabling us to distinguish list items from non-list items. Understanding the role of context in memory processes is particularly important in tasks such as free recall, where the retrieval cue is "context" itself.

Recent neurocomputational models of episodic memory (Sederberg, Howard, & Kahana, 2008; Polyn et al., 2009) suggest that contextual reinstatement underlies the *contiguity effect* — people's tendency to successively recall items that were presented in nearby positions on a studied list (Kahana, 1996). Behavioral studies of memory show that, for a given class of memories, the contiguity effect can span many other intervening memories (Howard & Kahana, 1999; Howard, Youker, & Venkatadass, 2008). This result is difficult to explain according to the view that contiguity arises from direct item-to-item associations that are established within a few seconds, as suggested by other classes of psychological and neurobiological theories (Raaijmakers & Shiffrin, 1981; Jensen & Lisman, 1998; Lisman, 1999; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Rolls & Kesner, 2006). The contiguity effect is an example of temporal clustering, which is perhaps the dominant form of organization in free recall.

Although this behavioral evidence provides indirect support for context-based theories of memory, there is no direct neurophysiological evidence for contextual reinstatement. To test the context reinstatement hypothesis, we studied 69 neurosurgical patients who were implanted with subdural electrode arrays and depth electrodes during treatment for drug-resistant epilepsy. As electrocorticographic

(ECoG) signals were recorded, the patients volunteered to participate in a free recall memory experiment, in which they studied lists of common nouns and then attempted to recall them verbally in any order following a brief delay.

5.3 Results

The recorded ECoG signals simultaneously sample local field potentials throughout the brain, and can be analyzed in terms of specific time-varying oscillatory components of neural activity. Such components have been implicated in memory encoding and retrieval processes (Fell et al., 2001; Paller & Wagner, 2002; Buzsáki, 2006; Osipova et al., 2006; Sederberg et al., 2007) and in the representations of individual stimuli (J. Jacobs & Kahana, 2009). For each study and recall event, we analyzed these oscillatory components across all recording electrodes (Fig. 5.1A,B). We constructed a matrix containing, for each electrode, measurements of mean oscillatory power in five frequency bands (δ : 2 – 4 Hz, θ : 4 – 8 Hz, α : 8 – 12 Hz, β : 12 – 30 Hz, and γ : 30 – 99 Hz) during each study event (200 to 1,600 ms relative to each word’s appearance on screen) and recall event (-600 to 200 ms relative to vocalization). We then used principal components analysis (PCA) to distill these highly correlated features into a smaller number of orthogonal components (Fig. 5.1C).

Context-based models conceive of context as a representation that integrates incoming information with a long time constant (Polyn & Kahana, 2008), leading to the prediction that the representation of temporal context evolves gradually as the experiment progresses (Manns, Howard, & Eichenbaum, 2007). We asked whether the neural recordings supported a gradually changing representation of context by regressing, for each participant, the mean similarity between the principal component vectors on their positional distance in the studied list (Fig. 5.2). The

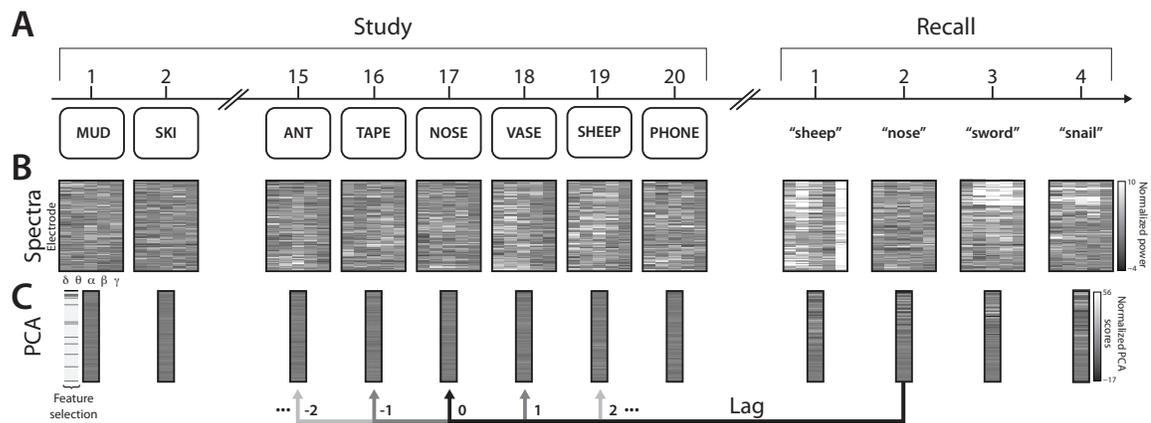


Figure 5.1. Illustration of behavioral and electrophysiological methods. **A.** After studying a list of 20 words and performing a brief distraction task, a participant recalls as many words as he can remember, in any order. **B.** During each study presentation and just prior to each recall event, we calculate the z-transformed oscillatory power at each recording electrode in each of five frequency bands (δ : 2 – 4 Hz, θ : 4 – 8 Hz, α : 8 – 12 Hz, β : 12 – 30 Hz, and γ : 30 – 99 Hz). **C.** We use principal components analysis (PCA) to find a smaller number of orthogonal dimensions that jointly account for a large proportion of the variation in the data shown in Panel B. We select those PCA components that show significant positive autocorrelation (a defining feature of temporal context) during the study phase of the experiment. We then compute the similarity (normalized dot product) between the feature vectors of each recall event (e.g., “nose”) and the feature vectors associated with the corresponding study event (lag = 0), as well as the similarity of the recall event to surrounding study events with varying lags.

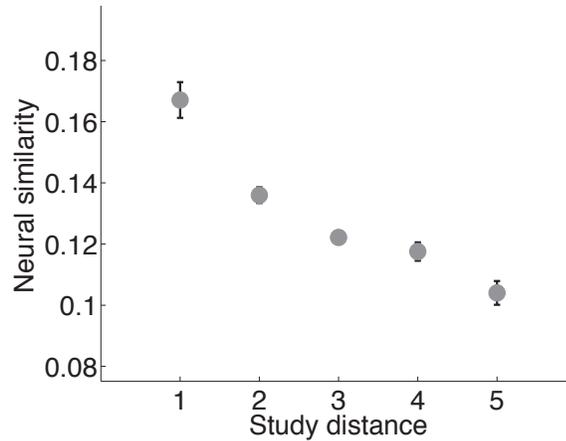


Figure 5.2. Evolution of ECoG activity as participants study lists of words. Mean neural similarity is shown as a function of study distance (difference in serial position) between pairs of presented words. Error bars denote \pm SEM.

similarity in recorded activity during each pair of word presentations decreased with the positional distance between the presentations (t -test on distribution of t -values from the regressions: $t(64) = -9.31, p < 10^{-10}$), indicating that the ECoG recordings evolve gradually over the course of the studied lists. Whereas this gradually changing neural representation is consistent with context-based models, such a result would also be expected to arise due to other autocorrelated neural processes that lack the rich dynamics implied by context-based theories of memory. To determine whether this gradually changing neural representation reflects the contexts in which list items were studied, we selected the autocorrelated PCA-derived features (henceforth referred to as feature vectors) for further analysis (see *Methods*).

To test whether the gradually changing neural representation is reinstated during recall, we compared feature vectors recorded during each study and recall event. First we identified the serial position (on the presentation list) of each correctly recalled word. If neural activity during study is reinstated during recall, then the neural activity recorded during a given recall event should be more sim-

ilar to activity recorded during the study event for the same word than during study events for other words (Figs. 5.3B,C). This finding would not be expected if the neural activity we measured did not contain content or context information (Fig. 5.3A). For each correctly recalled word (e.g., “nose” in Fig. 5.1A), we calculated the similarity between the feature vector associated with the recall event and the feature vectors associated with each of the studied items (e.g., ANT, TAPE, NOSE, VASE, SHEEP), where similarity is defined as the normalized dot product between the feature vectors (the vectors were normalized to have unit length before the dot product was performed). Each studied item was assigned a lag (positional distance) relative to the recalled item (e.g., VASE has a lag of +1 to “nose,” ANT has a lag of -2 to “nose,” and NOSE has a lag of 0 to “nose”). We found that the mean neural similarity at lag = 0 was significantly greater than the mean neural similarity at other lags (Fig. 5.4A, paired-sample *t*-test across 39 participants with at least 5 autocorrelated features; $t(38) = 3.10, p = 0.004$). This result would arise if the signal represents either content (the list words themselves) *or* context (the cues surrounding the items).

To distinguish between content and context reinstatement we compared the feature vectors associated with each recall event with the feature vectors associated with the neighbors of the recalled word in the study sequence. Context-based models predict that similarity between feature vectors should decrease as a function of absolute lag in both the forward (positive) and backward (negative) directions (Polyn & Kahana, 2008). For each participant we regressed the mean neural similarity between feature vectors on lag, separately for positive and negative lags (two regressions were performed for each participant). Each regression yielded a *t*-value associated with the slope (β coefficient) of the fitted line. Consistent with the context-reinstatement hypothesis, *t*-tests on the distributions of *t*-values

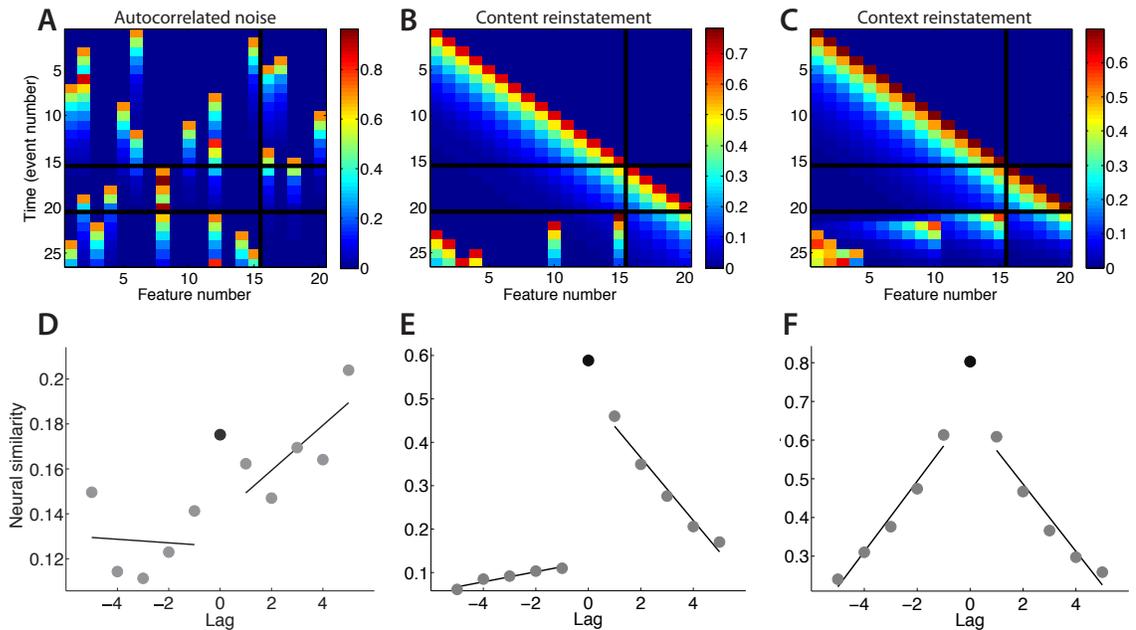


Figure 5.3. Predicted neural similarity as a function of lag according to three models. **A–C.** These panels show the pattern of activations for a simulated 20 neuron neural network as a 15 item list is studied. Events 1 – 15 of each matrix show activations after each item is presented. Events 16–20 show activations as distracting items are presented. Events 21–26 show activations as items 15, 10, 1, 2, 4, and 3 are recalled. In each simulation, a single neuron is activated during each experimental event. Once activated, a neuron’s activity decays gradually; thus multiple neurons may be active at a given time. **A.** For the *Autocorrelated noise* simulation, each experimental event activates a random neuron, irrespective of which item is being presented or recalled. **B.** For the *Content reinstatement* simulation, each neuron is activated by a single item or distractor (neurons 1–15 represent items; 16–20 represent distractors). Once activated, a neuron’s activity decays gradually; thus multiple neurons may be active at a given time. Only content information (specific to a single item) is reinstated during recall. **C.** The *Context reinstatement* simulation is similar to that shown in Panel B, but here we simulate context reinstatement during recall. **D–F.** These panels show the average expected neural similarity between the pattern of activity during study and recall as a function of lag. Each simulation used the same presented and recalled items that were included in our data analyses (Fig. 5.4). See *Supplemental materials* for further details on the simulations.

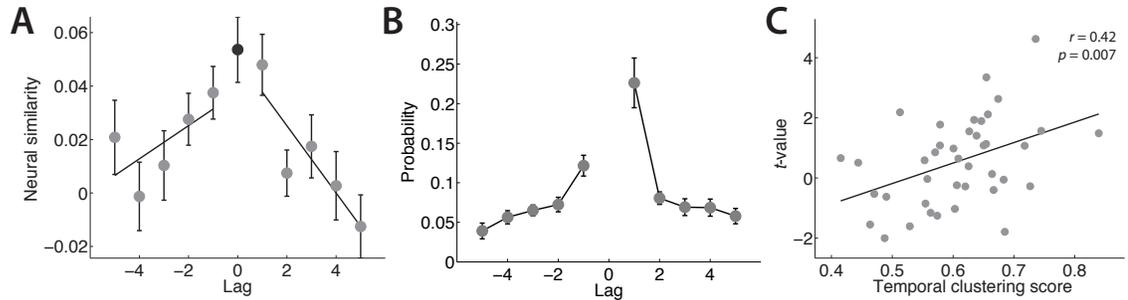


Figure 5.4. A neural signature of temporal context reinstatement. **A.** Neural similarity between the feature vector corresponding to recall of a word from serial position i and study of a word from serial position $i + \text{lag}$ (black dot denotes study and recall of the same word, i.e., lag = 0). **B.** Participants tend to successively recall neighboring study items (the contiguity effect). Here, we show the probability of recalling an item from serial position $i + \text{lag}$ immediately following an item from serial position i , conditional on the availability of an item in that list position for recall. Error bars in Panels A and B denote \pm SEM. **C.** Participants exhibiting greater context reinstatement also exhibited more pronounced contiguity effects. Here the t -value associated with the regressions in Panel A serves as a measure of the degree context reinstatement for each participant. (Only the regressions for negative lags were used, as the regressions for positive lags are not expected to distinguish between content and context reinstatement; Fig. 5.3.) Temporal clustering score measures the degree to which responses were clustered on the basis of their temporal contiguity at study (see *Supplemental analysis methods*).

across participants indicated that similarity decreased with absolute lag in both the positive ($t(38) = -3.63, p = 0.0008$) and negative ($t(38) = -2.42, p = 0.02$) directions. The decrease in similarity with absolute lag in both directions cannot be explained solely by content reinstatement or autocorrelated noise in the recordings (Figs. 5.3, 5.5). Mean similarity as a function of lag across participants is shown in Figure 5.4A.

The decrease in neural similarity with absolute lag elegantly mirrors the contiguity effect — people’s striking tendency to make transitions to neighboring items rather than remote ones, as seen in behavioral data for the same participants (Fig. 5.4B). Consistent with the hypothesis that the contiguity effect arises due to reinstatement of context (Howard & Kahana, 2002; Sederberg et al., 2008; Polyn et al., 2009), participants with stronger neural signatures of context reinstatement exhibited more pronounced contiguity effects than did participants with weaker re-

instatement effects ($r = 0.42, p = 0.007$, Fig. 5.4C). In recalling a list item, people not only reinstate that item's representation, as has been recently documented (Polyn et al., 2005; Gelbard-Sagiv, Mukamel, Harel, Malach, & Fried, 2008) but they also revive the brain activity associated with neighboring items. Further, the degree of this neural context reinstatement effect predicts the tendency of an individual participant to recall neighboring list items successively during memory search.

Having identified a neural signature of context reinstatement, we next asked whether this phenomenon could be localized to one or more brain regions. For example, recent work has given rise to the hypothesis that the medial temporal lobe (Skaggs, McNaughton, Wilson, & Barnes, 1996; Manns et al., 2007; Pastalkova, Itskov, Amarasingham, & Buzsáki, 2008; Jenkins & Ranganath, 2010; Howard, Viskontas, Shankar, & Fried, Submitted) and prefrontal cortex (Schacter, 1987; Polyn & Kahana, 2008; Jenkins & Ranganath, 2010) are critically involved in the maintenance and updating of temporal context. To test for regional specificity of context reinstatement, we repeated our test for neural context reinstatement using electrodes from each of the following regions of interest: temporal lobe (including the hippocampus and medial temporal lobe), frontal lobe (including prefrontal cortex), parietal lobe, and occipital lobe (Fig. 5.6A). We found that neural activity recorded from temporal lobe electrodes exhibited a decrease in similarity with increasing absolute lag in both the positive and negative directions (Fig. 5.6B; positive: $t(20) = -2.20, p = 0.04$; negative: $t(20) = -2.82, p = 0.01$). As in the whole brain analysis, the neural signature of context reinstatement in the temporal lobe was significantly correlated with the temporal clustering of participants' recalls (Fig. 5.6C; $r = 0.48, p = 0.03$). The frontal lobe exhibited a weak neural signature of context reinstatement that trended towards significance (positive: $t(20) = -2.85, p = 0.01$; negative: $t(20) = -1.54, p = 0.14$). However, this frontal signature of context re-

instatement was not correlated with temporal clustering of participants' recalls ($r = -0.08, p = 0.73$). Our findings in the parietal and occipital lobes were inconclusive due to insufficient data.

5.4 Discussion

The preceding analyses demonstrate that when recalling an item, the pattern of neural activity exhibits graded similarity to the neural activity measured during the encoding of items studied in neighboring list positions. Furthermore, the strength of this neural similarity effect tracks the behavioral contiguity effect in free recall: participants who exhibit a stronger tendency to make transitions among neighboring items during recall also exhibit a stronger relation between neural similarity and absolute lag. This pattern of results is exactly what one would predict on the basis of retrieved context theories of episodic memory (Bower, 1972; Howard & Kahana, 2002; Sederberg et al., 2008; Polyn et al., 2009). These theories posit that a gradually changing contextual state becomes associated with each experienced event, and that recalling an event revives the contextual state associated with the original experience. This retrieved context, in turn, activates other memories that were associated with similar contexts, producing the contiguity effect seen in recall tasks (Fig. 5.4B). The present findings provide critical neurobiological evidence in support of context reinstatement by showing that remembering an item reinstates the patterns of distributed oscillatory activity associated with surrounding (contextual) items from the original study episode. This neural signature of context reinstatement was observed both for the whole brain analysis and for recordings taken only from the temporal lobe.

What we have called a context representation might also be attributed to a

rehearsal-based model where participants engage in short bouts of covert recall during the study period. According to rehearsal-based models, words are rehearsed after they are presented, and more recently presented items are more likely to be rehearsed than distant items. Such models have been shown to be difficult to distinguish from context-based models (Laming, 2006; Davelaar, Usher, Haarmann, & Goshen-Gottstein, 2008), likely because a context-based mechanism is necessary to explain the pattern of rehearsals made in a free-recall task. To distinguish between these two classes of models we performed an analysis of the neural correlates of the primacy effect in our data. It has been well established that rehearsal is associated with enhanced recall for early list items (i.e., the primacy effect; Welch & Burnett, 1924; Postman & Phillips, 1965; Glanzer & Cunitz, 1966; Atkinson & Shiffrin, 1968; Rundus, 1971). Thus if our basic findings were driven by rehearsal, one would expect that participants exhibiting strong neural context reinstatement should also show a strong primacy effect. However, we observed no significant correlation between primacy and neural context reinstatement ($r = 0.12, p = 0.45$; see *Supplemental analysis methods*), indicating that rehearsal during study is unlikely to account for our findings.

Modern psychological and neuroscientific investigations are still grappling with basic questions regarding how the human brain establishes continuity in a rapidly changing environment, and how our memory system revives prior states of the world. Recent neurocomputational models of human memory (Bower, 1972; Sederberg et al., 2008; Polyn et al., 2009) posit that continuity is provided by a context representation that changes gradually over time as a consequence of the integration of present and past events. The current state of context is assumed to become associated with each newly experienced event, such that reminders of the event retrieve the event's associated context. This notion is consistent with Tulving's

contention that episodic memory retrieval is like mental time travel, in that when we remember the past, many details of the prior experience are retrieved along with the desired material (Tulving, 1983). By showing that a component of the neural activity retrieved during memory search shows graded similarity to the brain states observed during the study of neighboring stimuli, we provide neural evidence for temporal context reinstatement in humans.

5.5 Methods

5.5.1 Participants

We tested 69 patients with drug-resistant epilepsy who had arrays of subdural and/or depth electrodes surgically implanted for one to four weeks to localize the sites of seizure onset (see Tab. 5.1). The clinical team determined the placement of these electrodes with the goal of localizing suspected epileptogenic foci and identifying functional regions to be avoided in surgery (see *Supplemental recording methods* for details on recording methods). Our research protocol was approved by the appropriate institutional review boards and informed consent was obtained from the participants and their guardians. Data were collected as part of a long-term multicenter study with previously published articles describing separate analyses conducted on subsets of these data (Sederberg, Kahana, Howard, Donner, & Madsen, 2003; Sederberg et al., 2007).

5.5.2 Behavioral methods

Participants studied lists of 15 or 20 high frequency nouns for a delayed free-recall task. Following a fixation cue, the computer displayed each word for 1,600 ms

followed by a 800–1,200 ms blank inter-stimulus interval. Each word was displayed at most once within a single testing session. For 18 s following list presentation, participants solved a series of single-digit addition problems of the form $A + B + C = X$. Participants were then given 45 s to recall list items in any order. Vocal responses, digitally recorded during the trial, were scored for subsequent analysis. Participants recalled $22.7 \pm 1.0\%$ (mean \pm SEM) of the studied words. Repetitions and incorrect recalls ($32.4 \pm 2.6\%$ of all responses) were excluded from our analyses, as were responses that occurred within 1 s of a prior vocalization.

5.5.3 Data analysis

We measured oscillatory power in the ECoG recordings by applying a Hilbert transform to the Butterworth-bandpassed signal in each of five frequency bands (see *Results*). To reduce edge artifacts, we computed power at each frequency for the entire recording session before parsing the recordings into experimental events. Before applying PCA to the frequency \times electrode matrices (Fig. 5.1B) we z-transformed power values relative to the distribution of all events in the recording session (the z-transformation was performed independently for each frequency-electrode pair). We used the Kaiser criterion to choose, for each participant, the principal components that explained a substantial proportion of the variance (Kaiser, 1960). We next sought to identify principal components that changed gradually during the study period. Defining the value of component k for events at serial positions i and j as $v_k(i)$ and $v_k(j)$, we computed the correlation, r , between the absolute difference in the components ($|v_k(i) - v_k(j)|$) and the serial position lag ($|i - j|$) for $lag = 1 \dots (l - 1)$, where l was the length of the presented list. Components with $r > 0$ and $p < 0.1$ were selected for further analysis (see *Supplemental analysis methods*). Features that met these criteria were identified in

132 (out of 144) recording sessions. We further excluded an additional 68 sessions in which fewer than five candidate context features were identified. This threshold was chosen to balance two factors: first, we wanted to ensure that the feature vectors were of high enough dimensionality that it would be possible to observe neural signatures of context reinstatement (Fig. 5.4A) for each participant; second, we wanted to maximize the amount of data included in our analysis. We report the mean number of features selected for each participant in Table 5.1. To prevent selection bias, recall events were not used in the feature selection process.

5.6 Supplemental materials

5.6.1 Overview

This section is comprised of three main parts. First we include a table containing basic information about each neurosurgical patient whose data we analyzed. We next provide a table summarizing several details of the recording setup for each of our collaborating hospitals, as well as a description of the recording methods used in our study. Finally, we provide additional details on several analyses referenced in the *Results* and *Discussion* sections of this chapter.

5.6.2 Supplemental patient information

ID	HOSP	AGE	SEX	HAND	ELC	FEA	L LEN	SES	LST	REC	REP	PLI	ELI
1	BW	33	F	R	64	3	20	1	15	55	2	26	12
2	BW	51	F	R	40	5	20	1	15	66	2	8	1
3	BW	32	M	R	32	2	15	3	39	212	5	9	4
4	BW	40	M	R	96	5.5	15	2	20	82	5	38	28
5	BW	44	M	R	16	1	15	2	20	58	1	12	19
6	BW	27	M	R	64	1.5	15	2	20	76	49	12	3
7	BW	38	M	R	104	9.33	15	3	30	136	3	30	15
8	CH	13	F	R	64	9	20	1	12	59	0	2	1

Continued on next page . . .

ID	HOSP	AGE	SEX	HAND	ELC	FEA	L LEN	SES	LST	REC	REP	PLI	ELI
9	CH	12	F	R	104	19	20	1	15	39	0	0	2
10	CH	15	M	L	128	15.67	20	3	30	90	1	12	8
11	CH	17	M	R	64	4	20	3	45	178	20	27	17
12	CH	15	M	R	123	13	20	1	15	86	3	6	3
13	CH	11	M	R	104	0	20	2	30	104	3	2	3
14	CH	14	F	R	72	0	20	1	15	104	2	7	5
15	CH	8	F	R	86	6.5	20	2	30	159	5	18	10
16	CH	17	M	R	84	12	20	1	14	30	2	9	12
17	CH	17	M	L	124	10.5	20	4	60	116	2	104	13
18	CH	20	F	R	128	8.5	15	2	24	114	2	9	4
19	CH	14	M	R	94	6.67	15	3	30	94	0	18	13
20	CH	17	M	L	80	7	15	2	20	14	0	10	22
21	CH	19	F	R	125	8.5	15	2	17	47	2	10	1
22	CH	16	M	R	156	13	15	1	16	76	1	4	1
23	CH	12	M	L	83	5	15	2	20	52	9	12	25
24	CH	13	M	R	72	4.75	15	4	40	200	3	4	2
25	FR	33	M	R	98	8	20	1	9	43	0	0	0
26	FR	25	M	R	85	21	20	1	9	45	18	1	2
27	FR	31	M	L	56	4	20	1	9	24	0	4	0
28	FR	41	F	R	63	9	20	1	7	23	0	9	6
29	FR	34	F	L	40	3	20	1	7	38	3	5	3
30	FR	45	F	L	100	10	20	1	8	27	0	20	3
31	FR	46	F	L	14	0	20	1	1	4	3	0	5
32	FR	20	M	R	84	4	20	1	15	42	6	6	2
33	FR	53	F	L	41	5	20	1	15	49	26	21	21
34	FR	50	M	R	68	4	20	2	30	116	5	36	15
35	FR	28	M	L	112	11	20	1	15	37	1	1	1
36	FR	30	F	R	60	7	20	1	15	67	9	10	2
37	FR	37	F	L	30	6	15	1	20	65	84	73	42
38	FR	18	M	L	30	2	15	1	20	121	7	12	13
39	FR	23	M	L	58	3	15	4	56	281	98	32	29
40	FR	21	M	L	93	4	15	1	10	49	2	10	3
41	FR	28	F	R	86	7	15	1	10	36	4	4	6
42	FR	35	F	L	122	3	15	2	20	54	0	6	6
43	FR	37	F	L	52	2.5	15	4	35	161	38	29	28
44	FR	19	M	L	74	2	15	2	30	148	19	14	35
45	FR	41	F	R	30	3	15	1	15	15	0	3	38
46	FR	21	F	R	64	6	15	1	15	50	0	3	7
47	FR	43	F	R	56	0	15	1	15	23	5	11	59
48	FR	19	M	R	30	3	15	2	25	120	2	5	27
49	FR	21	M	R	70	5	15	5	53	408	41	8	194
50	FR	35	F	R	62	6	15	1	15	44	24	16	114
51	FR	25	M	R	84	5	15	2	30	145	5	3	96
52	FR	47	M	L	82	4	15	1	4	13	1	3	35
53	FR	45	F	R	88	4	15	1	10	43	11	3	0
54	TJ	25	M	R	62	0.67	15	3	48	232	3	6	1
55	TJ	40	F	R	94	3.25	15	4	64	164	8	54	43
56	TJ	39	M	L	56	2	15	1	16	53	1	8	20
57	TJ	34	F	R	111	4.3	15	10	154	513	7	110	24
58	TJ	44	M	R	125	7	15	1	13	31	1	7	6
59	UP	38	M	R	62	4.75	15	4	40	135	3	68	24
60	UP	30	M	R	86	3	15	2	20	54	5	24	21
61	UP	43	M	R	66	2.33	15	3	18	31	22	12	33
62	UP	36	M	R	88	5.75	15	4	40	70	6	114	50
63	UP	25	M	R	62	3	15	4	40	135	2	1	2
64	UP	18	F	R	76	7.33	15	3	30	104	5	6	3
65	UP	27	F	R	48	2	15	2	32	104	2	43	20
66	UP	55	F	L	80	1.5	15	2	32	81	11	61	24
67	UP	18	M	A	100	2.33	15	3	48	253	7	8	3
68	UP	38	F	R	86	6	15	1	16	48	14	3	73

Continued on next page ...

ID	HOSP	AGE	SEX	HAND	ELC	FEA	L LEN	SES	LST	REC	REP	PLI	ELI
69	UP	40	M	R	58	5.75	15	4	64	304	1	14	9

Table 5.1. Patient and task information. This table provides the hospital (HOSP) at which each patient’s data were collected (see Tab. 5.2), as well as each patient’s age (AGE), gender (SEX), handedness or language mapping (HAND), number of implanted electrodes (ELC), and mean number of features selected for analysis across all sessions for that patient (FEA). Information about the task includes the list length (L LEN) used for each participant, number of testing sessions (SES), and the number of lists each participant encountered across all sessions (LST). Performance information includes the total number of correct recalls across all lists (REC), the total number of repeated recalls (REP), and the total number of incorrect recalls, which includes recalls of previously presented items (prior list intrusions, or PLIs) and recalls of items which were never presented (extra-list intrusions, or ELIs). In total the 69 patients contributed 5,299 electrodes and 739 selected features, studying 29,030 items presented in 1,790 lists.

5.6.3 Supplemental recording methods

Subdural grids or depth electrodes (Ad-Tech Inc., Racine, Wisconsin) were implanted by neurosurgical teams solely for clinical purposes. The locations of the electrodes were determined by means of co-registered post-operative computed tomography and pre-operative magnetic resonance imaging (MRI), or from post-operative MRIs, by an indirect stereotactic technique and converted into Montreal Neurological Institute (MNI) coordinates. Electrocorticographic (ECoG) signals were recorded referentially using a Bio-Logic, XLTek, Neurofile, or Nicolet EEG digital video-EEG system. Depending on the amplifier, signals were sampled at 200, 256, 500, 512, or 1,024 Hz. Several hospitals applied bandpass filters to the recorded signals prior to writing to disk (Tab. 5.2). Where applicable, frequencies outside of the filtered range were excluded from further analysis. Data were subsequently notch-filtered with a Butterworth filter with zero phase distortion at 50 or 60 Hz to eliminate electrical line and equipment noise. ECoG signals and behavioral events were aligned using synchronization pulses sent from the testing computer (mean precision < 4 ms).

HOSPITAL	HOSP	LOWER	UPPER
Brigham & Women's Hospital, Boston, MA	BW	0.5 Hz	60 Hz
Children's Hospital Boston, Boston, MA	CH	0.3 Hz	50 Hz
University Hospital of Freiburg, Freiburg, Germany	FR	0.1 Hz	100 Hz
Thomas Jefferson University Hospital, Philadelphia, PA	TJ	$-\infty$	∞
Hospital of the University of Pennsylvania, Philadelphia, PA	UP	$-\infty$	∞

Table 5.2. Bandpass filters employed by our collaborating hospitals. Hospital codes (HOSP) are referenced in Table 5.1. The LOWER and UPPER columns denote the lower and upper limits of the bandpass filters, respectively. Frequencies outside of the bandpassed range were excluded from further analysis.

5.6.4 Supplemental analysis methods

Quantifying the contiguity effect

Figure 5.4C depicts an analysis relating the neural reinstatement effect to the recall behavior of the participants. Specifically, we show that participants showing stronger neural reinstatement effects tend to exhibit a stronger contiguity effect (whereby neighboring list items tend to be recalled successively). The contiguity effect is measured using the temporal clustering score, an analysis technique described previously (Polyn et al., 2009). The temporal clustering score is calculated as follows.

For each recall transition we create a distribution of temporal distances between the just-recalled word and the set of words that have not yet been recalled. These distances are simply the absolute value of the difference between the serial position of the just-recalled word and the set of not-yet-recalled words. A percentile score is generated by comparing the temporal distance value corresponding to the next item in the recall sequence with the rest of the distribution. Specifically, we calculate the proportion of the possible distances that the observed value is less than, since strong temporal clustering will cause observed lags to be smaller than average. As is often

the case, when there is a tie, we score this as the percentile falling halfway between the two items. If the participant always chose the closest temporal associate (which is only possible for pure serial recall in the forward or backward direction), then the temporal clustering score would yield a value of 1 (as there would never be an opportunity for a tie). A value of 0.5 indicates no effect of temporal clustering. Each patient was assigned a temporal clustering score by taking the average of the percentile scores across all observed recall transitions.

Quantifying the primacy effect

To test whether our main finding depicted in Figure 5.4A reflects rehearsal (rather than context reinstatement) we measured, for each participant, the strength of the primacy effect, a well-established correlate of rehearsal (Welch & Burnett, 1924; Postman & Phillips, 1965; Glanzer & Cunitz, 1966; Atkinson & Shiffrin, 1968; Rundus, 1971). The primacy and recency effects refer to an enhancement in memory for early and late list items, respectively, as compared with memory for intermediate list items (Deese & Kaufman, 1957; Murdock, 1962). The number of items that show a boost in memorability due to primacy or recency is relatively invariant to changes in list length; the primacy effect generally affects the first few items, while the recency effect generally affects the last six or so items (Murdock, 1962). In order to measure the strength of the primacy effect, we labeled the first three serial positions on each list as primacy positions and the last six serial positions as recency positions. The remaining positions were labeled as intermediate list positions (i.e., items 4 – 9 for 15 word lists, or items 4–14 for 20 word lists). We then measured the strength of the primacy effect for each participant by dividing their mean probability of recalling items from primacy positions by their mean probability of recalling items from intermediate list positions.

Simulations

To determine whether neural context reinstatement (e.g. Fig. 5.4A) could be attributed to non-context-based factors, we conducted three neural network simulations (Figs. 5.3, 5.5) that predict the expected outcome of our test for context reinstatement under various model assumptions. The *autocorrelated noise* model has neural activity evolve randomly over time, irrespective of what is happening in the experiment. The *content reinstatement* model has each neuron represent a different word; a neuron is activated if its associated word is presented or recalled. The *context reinstatement* model also has each neuron represent a different word. We simulate context reinstatement by activating not only the neuron associated with the word being recalled, but also other neurons that were active at the time the recalled word was studied.

For all three simulations we define an activity vector, \mathbf{f} , that defines the pattern of activation across the network. Each neuron in the network takes on a value between 0 (inactive) and 1 (maximally active). Let \mathbf{f}_i denote the state of \mathbf{f} after the i^{th} experimental event (i.e., a study presentation, distracting task, or recall). Our main analysis entails selecting autocorrelated components of neural activity as the candidate context representation (see *Results*). After this feature selection, the feature vectors we analyze are autocorrelated — a property we need to take into account in our simulations. In particular,

$$\mathbf{f}_i = \rho_i \mathbf{f}_{i-1} + \beta \mathbf{w}_i,$$

where β is a constant; ρ_i is a function of \mathbf{f}_{i-1} , \mathbf{w}_i , and β (with $0 \leq \rho_i, \beta \leq 1$); and \mathbf{w}_i is the pattern of neural activity specifically evoked by the i^{th} experimental event (for details, see Polyn & Kahana, 2008). In this way, the neural activity

measured after a given experimental event (e.g. presentation of the 5th list item) is a recency-weighted blend of the activity evoked by previous experimental events (e.g. activity evoked by presentations of items 5, 4, 3, 2, and 1). We initialize \mathbf{f}_0 by setting the activation of the first neuron to 1 and the activations of the other neurons to 0. We then simulate different experimental events by adjusting \mathbf{w}_i according to the particular rules of each model. We ensure that \mathbf{f}_i is always of unit length by setting

$$\rho_i = \sqrt{1 + \beta^2[(\mathbf{f}_{i-1} \cdot \mathbf{w}_i)^2 - 1]} - \beta(\mathbf{f}_{i-1} \cdot \mathbf{w}_i).$$

For the autocorrelated noise model, each \mathbf{w}_i is set to a vector of 0's, plus a 1 in a single random position. In this way, each \mathbf{w}_i activates one of the neurons in the network at random. As shown in Figure 5.5A, for $\beta < 0.5$, similarity between \mathbf{f}_i during presentation and \mathbf{f}_j during recall increases as a function of i . This is because, by definition, an autocorrelated signal measured at times t and $t + \Delta$ becomes more similar as $\Delta \rightarrow 0$. For $\beta > 0.5$, similarity as a function of lag flattens out, since as \mathbf{f}_i is dominated by \mathbf{w}_i the average similarity between \mathbf{f}_i and \mathbf{f}_j approaches the expected similarity between two independent draws of \mathbf{w}_i .

For the content reinstatement model, \mathbf{w} is set differently depending on the type of experimental event. In this model, each neuron is assigned a different word or distractor. During presentation of study items or distractors, \mathbf{w}_i is set to a vector of all 0's, plus a 1 in the position of the neuron representing the item or distractor being presented. During recall of the j^{th} presented item, we set $\mathbf{w}_i = \mathbf{w}_j$. As shown in Figure 5.5B, for $\beta < 0.5$, similarity increases as a function of lag. Since β is small, \mathbf{f}_i is dominated by \mathbf{f}_{i-1} rather than \mathbf{w}_i . Since the specifics of the experimental event contribute only minimally to \mathbf{f} , the simulation approximates the autocorrelated noise simulation. For $0.5 < \beta < 1$, neural similarity is roughly constant as a

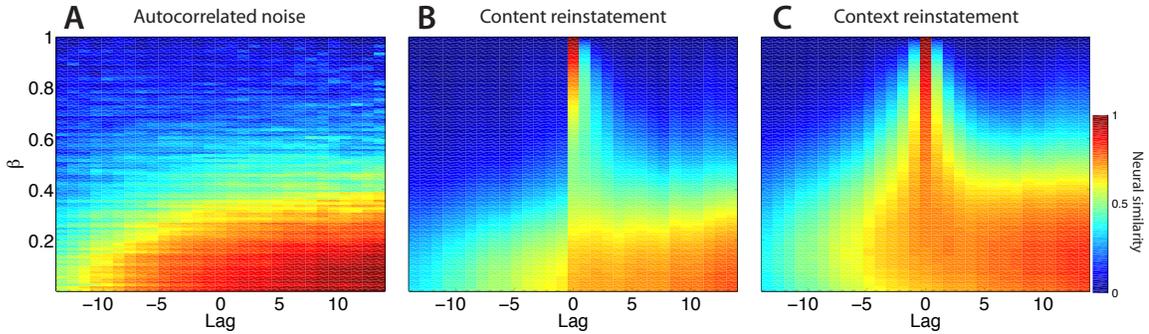


Figure 5.5. Simulated neural similarity as a function of lag and drift rate (β) given A. no content or context information in the neural recordings (Figs. 5.3A,D), B. content reinstatement without context reinstatement (Figs. 5.3B,E), and C. context reinstatement (Figs. 5.3C,F). Similarity is computed as the normalized dot product between the simulated feature vector after the recall of the i^{th} word and the feature vector corresponding to presentation of word $i + \text{lag}$. The first dimension (initialized to 1 prior to the start of the simulation) was ignored for the similarity calculations. Simulation results in Figure 5.3 in the main text used $\beta = 0.7$ (this choice was motivated by previously reported simulation results; Polyn et al., 2009).

function of lag for negative lags, but decreases as a function of lag for positive lags. This is because the pattern of activation during the i^{th} presentation will only contain traces of \mathbf{w}_j if $i > j$. Finally, for $\beta = 1$, similarity is 1 when lag = 0 and is 0 everywhere else. This is due to the fact that when $\beta = 1$, $\mathbf{f}_i = \mathbf{w}_i$, and so the neural activity evoked by the i^{th} item will be present only during its presentation or recall.

The context reinstatement model is identical to the content reinstatement model during the presentation of study items and distractors. To simulate context reinstatement during recall of the j^{th} presented item, we set $\mathbf{w}_i = \mathbf{f}_j$ (recall that \mathbf{f}_j will contain a recency-weighted average of the activations associated with the previously presented items). As shown in Figure 5.5C, for $\beta < 0.5$, similarity increases as a function of lag, just as in the other simulations. Importantly, for $0.5 < \beta < 1$, neural similarity decreases with absolute lag in *both* the positive and negative directions, as seen in the neural data (Fig. 5.4A). Finally, as in the content reinstatement simulation, for $\beta = 1$ similarity is 1 when lag = 0 and is 0 everywhere else.

These simulations show that regardless of the precise rate at which neural ac-

tivity evolves over time, the simplest model consistent with our neural results (Fig. 5.4A) is one in which the temporal context in which an item is studied is reinstated when the item is recalled. While we have not ruled out every possible model that does not include some form of context reinstatement, neither autocorrelated noise (Fig. 5.5A) nor content reinstatement alone (Fig. 5.5B) can account for the neural signature of context reinstatement we observed in our ECoG recordings.

Selecting autocorrelated features

Context-based theories of memory posit the existence of a gradually changing pattern of neural activity that becomes associated with each studied word during study, and is reinstated during recall. To identify candidate components of the context representation for a given recording session, we selected autocorrelated PCA-derived features of the neural representation (Fig. 5.1C) as follows. Separately for each feature x , we computed the Pearson's lag 1 autocorrelation coefficient (r) and associated p -value for the values of x within each list. We then combined the autocorrelation coefficients across lists into a summary autocorrelation measure, \bar{r} :

$$\bar{r} = F^{-1}\left(\sum_{i=1}^L F(r_i)\right),$$

where r_i was the Pearson's lag 1 autocorrelation coefficient for the values of x measured during list i , $F()$ was the Fisher z -prime transformation:

$$F(r) = \frac{\ln(1+r) - \ln(1-r)}{2},$$

and $F^{-1}()$ was the inverse of $F()$:

$$F^{-1}(z) = \frac{e^{2z-1}}{e^{2z+1}}.$$

In this way, if r_i has large positive values across all lists, \bar{r} will have a large positive value. Similarly, if r_i is negative across all lists, \bar{r} will have a large negative value. If r_i is sometimes positive and sometimes negative (with approximately equal probability), \bar{r} will take on a value near zero. (Note that $-1 \leq r_i, \bar{r} \leq 1$.)

We also obtained a p -value, \bar{p} , associated with \bar{r} by applying the inverse Normal transformation to the p -values associated with the Pearson's lag 1 autocorrelation coefficients for each list. We then summed across the transformed p -values and evaluated the cumulative Normal distribution function at this sum to obtain \bar{p} . We selected features with $\bar{r} > 0$ and $\bar{p} < 0.1$ for further analysis (see *Results*).

Identifying the time interval of the recall event

Our main analysis (Fig. 5.4A) compares the neural activity elicited by a studied word to the neural activity elicited by a word's retrieval during the recall period. We restrict our analysis of the study period to ECoG activity beginning 200 ms after the appearance of a word and ending when the word disappears from the screen. Here the 200 ms delay was meant to account for the lag between the word's appearance onscreen and the processing of the word by the participant.

In order to search for the optimal time interval for the recall event, we tested for context reinstatement while varying both the duration and onset of the time interval for the recall event. We tested time intervals ranging in duration from 100 to 1,000 ms (in increments of 100 ms), and onsets ranging from -1,000 to -200 ms (in increments of 100 ms) relative to the time the participant began their vocalized

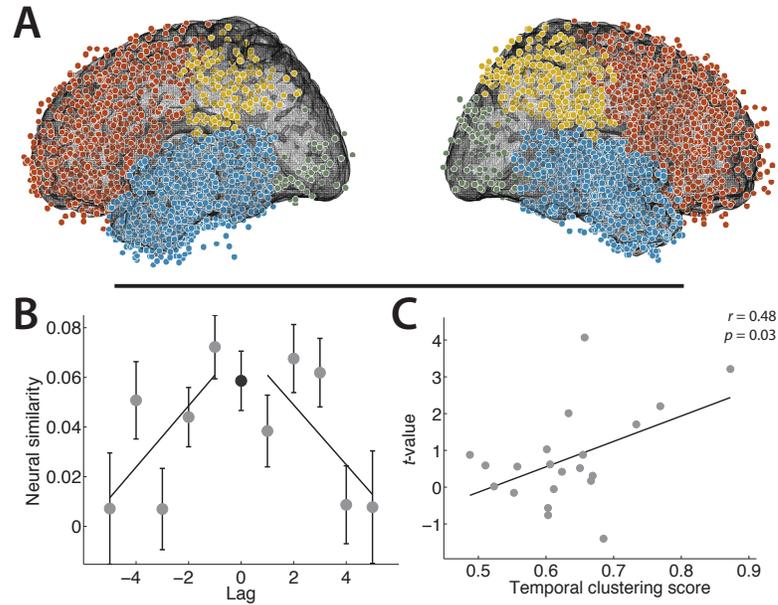


Figure 5.6. Evidence for context reinstatement in the temporal lobe. **A.** Each dot marks the location of a single electrode from our dataset in MNI space. We divided our dataset into four regions of interest: temporal lobe (blue; 1,815 electrodes), frontal lobe (red; 1,737 electrodes), parietal lobe (yellow; 512 electrodes), and occipital lobe (green; 138 electrodes). **B – C.** These panels are in the same format as Figures 5.4A and C, but reflect data from temporal lobe electrodes only.

recall. This analysis indicates that the context reinstatement effect is strongest for the recall interval ranging from -600 to 200 ms relative to vocalization.

To account for the possibility that different brain regions reinstate context at different times relative to vocalization, we repeated this optimization analysis separately for each region of interest. The best time interval for the temporal lobe was from -400 to -300 ms (Figs. 5.6B,C). The time interval that gave the strongest frontal lobe effect was from -900 to -400 ms; however, the frontal effect was not statistically reliable (see *Results*).

Contributions to the context representation by brain region and frequency band

Figure 5.6 provides additional details on our region of interest analysis reported in *Results*. In addition to asking whether specific brain regions contribute to the

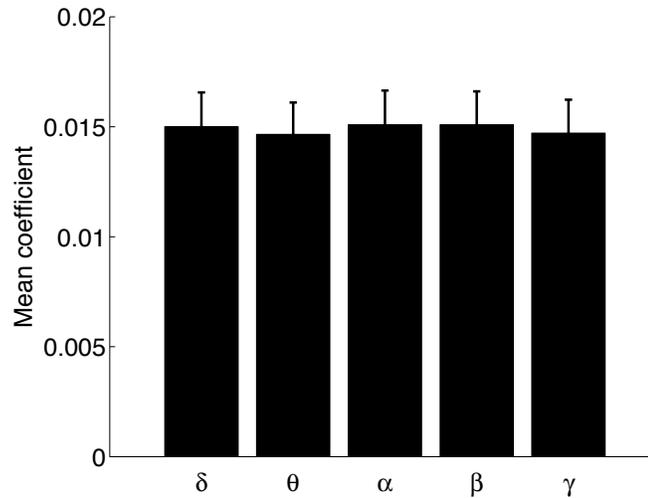


Figure 5.7. Mean contributions of each frequency band to selected features. Error bars indicate 95% confidence intervals (Loftus & Masson, 1994).

representation of context, a natural question is whether the principal components comprising the feature vectors tend to weight particular oscillatory components of ECoG activity more heavily than others. Because principal components analysis (PCA) performs a linear mapping from the n -dimensional space of the original set of activity vectors onto the m -dimensional PCA space (where $m \leq n$), we can use the PCA coefficients to perform the inverse mapping of the feature vectors back onto the original n -dimensional space. The PCA coefficients tell us how much each of the elements in the original PC vectors contributes to each of the principal components in the feature vectors. This allowed us to determine the degree to which each oscillatory component recorded from each electrode contributes to each element of the feature vectors. For a given frequency band, we assessed the degree to which that frequency band contributed to the feature vectors across all study and recall events by examining the distribution of PCA coefficients assigned to that frequency band across all participants. An analysis of PCA coefficients across frequency bands revealed no significant differences between frequency bands (repeated measures

analysis of variance (ANOVA): $F(4, 37) = 0.57, MSE = 10^{-5}, p = 0.69$). This finding suggests that the selected features are comprised of oscillatory activity at a broad range of frequencies.

Chapter 6

General discussion

In this chapter, I first summarize how the body of work presented in the preceding chapters advances our current understanding of how the brain acquires, stores, and retrieves information. I then suggest future directions for this work.

6.1 Contributions of this dissertation

The role of the brain can be broken down into three general tasks: acquiring information about its environment (i.e., input; Chapter 2), performing computations on newly acquired and previously acquired information (i.e., processing and storage; Chapters 3 and 4), and using the processed and stored information to generate behaviors that change or manipulate the environment (i.e., retrieval and output; Chapter 5). The continuum ranging from input to output is illustrated in Figure 6.1; a complete understanding of the brain entails an integrated understanding of many discrete points along this continuum. The tick marks in Figure 6.1 mark the contributions of each chapter of this dissertation.

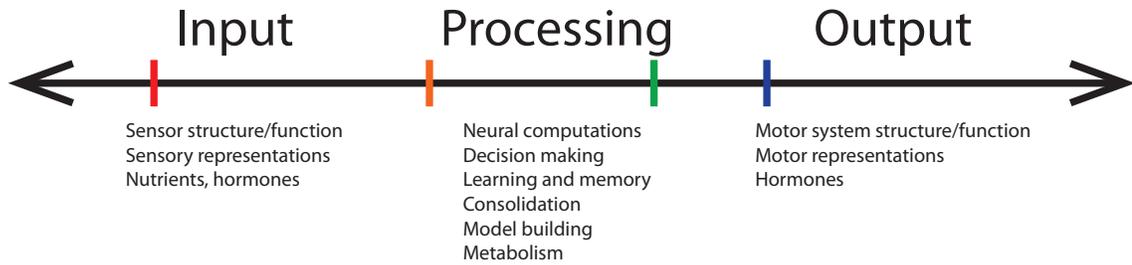


Figure 6.1. Putting this dissertation into context. The continuum ranging from input to output comprises a full specification of the role of the brain. The contributions of each chapter of this dissertation are marked by colored tick marks. Chapter 2 (red) examines how arrays of sensors should be designed so as to make accurate inferences about the environment. Chapter 3 (orange) examines how the firing of individual neurons relates to population activity, which has implications for how the brain performs computations. Chapter 4 (green) examines how the brain represents individual words and organizes these representations in memory. Chapter 5 (blue) examines how the brain recollects prior experiences. Taken together, this body of work spans a wide range of this continuum. A complete understanding of the brain requires filling in the gaps.

6.2 Future directions: Chapter 2

6.2.1 Learning receptor types from receptor responses

The model system presented in Chapter 2 provides a means of evaluating an arbitrary retinal mosaic's ability to reconstruct visual stimuli from photoreceptor responses. Two critical assumptions of the model are that the neural machinery has accurate knowledge of the receptors' response properties (e.g. the tuning curves of each receptor) and of the statistical properties of the stimulus distribution. However, it is not clear *a priori* whether such knowledge must be "built in" (i.e., genetically encoded), or whether these components might be learned over time by observing photoreceptor responses.

Doi, Inui, Lee, Wachtler, and Sejnowski (2003) showed that unsupervised learning could produce L/M cone opponent neurons within a neural population when the visual input was typical of natural viewing. Critically, however, they do not address the question of whether such learning could identify which neurons in the population carried the opponency. Preliminary work by Manning, Hurst, and

Brainard (2011) suggests that observed receptor responses contain sufficient information to simultaneously label receptors in the mosaic and estimate the parameters of the stimulus distribution via unsupervised learning. An important goal of future work will be to gain a better understanding of how the receptor responses drive the learning process.

6.3 Future directions: Chapter 3

6.3.1 Behaviorally relevant broadband shifts in local field potential power spectra

The primary finding presented in Chapter 3 is that broadband shifts in local field potential (LFP) power spectra are correlated with the firing rates of the underlying neurons. This work leaves open the question of whether these spectral changes are related to behavior and cognition (for review, see Crone, Korzeniewska, & Franaszczuk, 2011). To test the hypothesis that broadband changes are behaviorally relevant, one could replicate many of the previously reported analyses of oscillatory components of neural activity (e.g., Fell et al., 2001; Paller & Wagner, 2002; Osipova et al., 2006; Sederberg et al., 2007; J. Jacobs & Kahana, 2009) using broadband power as a factor.

Preliminary work by Pagan, Lega, Jacobs, and Kahana (in preparation) and Burke, Zaghoul, Jacobs, and Kahana (in preparation) suggests that broadband changes in LFP power spectra accompany memory formation. Both of these studies examined the ECoG free recall dataset analyzed in Chapters 4 and 5. In an analysis similar to that of Sederberg et al. (2007), Pagan et al. (in preparation) examined the spectral pattern at each electrode as each word was studied. They then

asked whether there were differences in the patterns of activity during study of words that were subsequently remembered, as compared with words that were forgotten (i.e., subsequent memory effects (SMEs); Paller & Wagner, 2002). This analysis revealed strong broadband decreases in power during study of words that were subsequently remembered (i.e., negative broadband SMEs). This effect was most apparent at temporal electrodes. The negative broadband SMEs occurred simultaneously with the narrowband SMEs observed by Sederberg et al. (2007) and others.

Burke et al. (in preparation) also examined the pattern of activity observed during study of each word. Here the authors were primarily interested in whether the positive gamma SMEs reported by Sederberg et al. (2007) reflected changes in connectivity between distributed networks in the brain. Their analysis used coherence, which measures shared oscillatory power with a constant phase offset, as a proxy for inter-electrode connectivity. The analysis revealed that successful memory encoding is accompanied by a decrease in gamma coherence between temporal electrodes and those in other brain regions. The authors then performed post-hoc analyses to clarify the nature of this phenomenon. These analyses indicated that the positive gamma SMEs reflected broadband changes in spectral power analogous to those reported in Chapter 3 and by Pagan et al. (in preparation), rather than true oscillations (e.g., Fries et al., 2007). However, in addition to changes in the overall height of the power spectrum reported in these studies, Burke et al. (in preparation) found that successful memory encoding was characterized by changes in the *slope* of the power spectrum. This is indicative of simultaneous high frequency increases and low frequency decreases, similar to those reported by K. Miller, Leuthardt, et al. (2007).

Taken together, these preliminary studies show that, in addition to the nar-

rowband changes that accompany memory formation, several brain areas exhibit broadband changes in power. If the brain's ability to subsequently recall studied words reflects the degree to which the study events become associated with their temporal context (e.g., Howard & Kahana, 2002), the neural signature of temporal context reported in Chapter 5 suggests a potential mechanism underlying the broadband SMEs observed by Pagan et al. (in preparation) and Burke et al. (in preparation). In particular, the neural signature of temporal context reinstatement reported in Chapter 5 is characterized by a broad range of frequencies (Fig. 5.7). Thus it is possible that the degree to which the studied words are bound to their temporal context appears in the neural data as a broadband change in spectral power.

6.3.2 Changes in the firing rate–LFP relation as a function of firing rate

Most analyses of the relation between single-neuron spiking and spectral components of the LFP, including those reported in Chapter 3, assume that this relation remains constant over time. However, several recent studies have reported that the correlation between firing rate and gamma power changes with firing rate (Nir et al., 2007; Mazzone, Whittingstall, Brunel, Logothetis, & Panzeri, 2010). Preliminary work by Ramayya, Manning, Jacobs, and Kahana (2010) suggests that the correlation between firing rate and broadband power is positively correlated with firing rate. An important goal of future work will be to incorporate these findings into neural decoders.

6.4 Future directions: Chapters 4 and 5

6.4.1 A PCA-based framework for identifying behaviorally relevant components of neural activity

The methods used in Chapters 4 and 5 to identify relevant components of neural activity are quite similar. The general approach taken for both studies can be summarized as follows:

1. Create feature vectors for neural activity recorded during each word presentation (study event) and just prior to each recall (recall event).
2. Use principal components analysis (PCA) to map the full set of features onto orthogonal components.
3. Discard components with coefficients (i.e., eigenvalues) less than 1 (Kaiser, 1960).
4. Use study events to select components of interest (e.g., components that vary with the meanings of words as in Chapter 4; autocorrelated components as in Chapter 5; etc.) for further analysis.
5. Perform analyses on selected components from recall events.

Applying feature selection to the principal components rather than the original data has a number of useful properties that make this approach particularly well-suited to analyzing neural data.

The use of PCA in analyses of neural data is widespread, both as a tool for dimensionality reduction and as a means of teasing apart multivariate signals (e.g., Andersen, Gash, & Avison, 1999; Viviani, Grön, & Spitzer, 2005; K. Miller et al.,

2009; Zhong et al., 2009). By definition, PCA combines features that vary in similar ways; each principal component is a weighted sum of correlated features in the original data. Correlations across features of the neural data can arise if the sources of those features are directly connected in the brain (e.g., Fries et al., 2007), if the sources receive shared input (e.g., Steriade, Amzica, & Contreras, 1996; Llinas & Steriade, 2006), if the sources are nearby in space (Pernier, Perrin, & Bertrand, 1988) or frequency (Buzsáki, 2006), or due to other aspects of the brain's physiology (e.g., K. Miller et al., 2009).

In the analyses I present in Chapters 4 and 5, I assume that neural features that represent the same internal variables (e.g. semantic features in Chapter 4; temporal context features in Chapter 5) will be correlated. If so, applying PCA to the data combines neural features that represent the same internal variables. In this way, PCA can be conceived of as a noise-reduction technique. Thus, it should be easier to identify neural correlates of these internal variables in the principal component vectors than in the original data.

Figure 6.2 provides a simple example that illustrates this notion. Consider neural recordings from n noisy sources in the brain. Here a source can mean the voltage recorded at a particular electrode, the power of an oscillation at a particular frequency, BOLD activation of a voxel, or any other relevant measure of neural activity. Suppose that each of the n sources receives the common sinusoidal input signal shown in Panel A. Finally, suppose that the connection between the common input and each source is either excitatory (i.e., positive) or inhibitory (i.e., negative), with equal probability of either designation. In this way, the signal measured from each source i is given by:

$$f_i(t) = w_i s(t) + \epsilon_i(t),$$

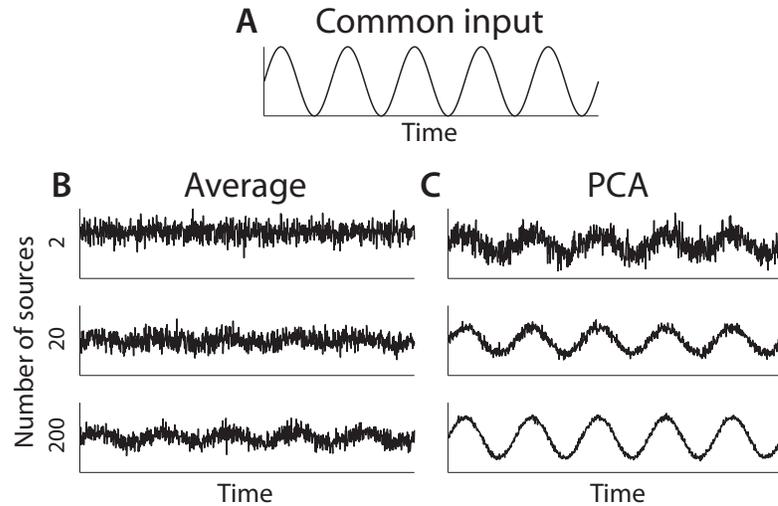


Figure 6.2. Applying feature selection to principal components. Data in all panels are synthetic. **A.** The sinusoidal input contributes equally to each measured neural source. Each source is noisy, and is connected to the common input via either an excitatory or an inhibitory synapse. **B.** Averaging the signals from the n sources fails to recover the common input. The subpanels show the average signal from 2, 20, and 200 sources. **C.** The PCA-based decoder recovers the common input. The subpanels show the first principal component derived from 2, 20, and 200 sources.

where w_i is set to 1 if the connection between source i and the common input is excitatory, and -1 otherwise; $s(t)$ is the contribution from the common sinusoidal input at time t ; and $\epsilon_i(t)$ is a draw from $N(0, 1)$ taken independently for each source at each time t . We can then examine the extent to which various decoding algorithms recover the input signal for various choices of n (here I use $n \in \{2, 20, 200\}$). I consider two decoding algorithms: the first averages the signal measured from each source (Fig. 6.2B), and the second examines the principal component with the largest eigenvalue (Fig. 6.2C). Visually comparing panels B and C shows that the PCA-based decoding algorithm best recovers the common input signal. The intuition here is simple. Because w_i is set to 1 or -1 with equal probability, as $n \rightarrow \infty$ the input will exactly cancel out in the average signal across all of the sources. By contrast, the PCA-based decoder will effectively flip the sign of w_i to the extent that doing so will better explain correlations amongst the sources.

The success of this approach in identifying the neural representations of studied items (Chapter 4) and temporal context (Chapter 5) suggests that the analysis framework could be adapted to detect other neural patterns as well. For example, one could select neural components that distinguish between tasks associated with each study event (e.g., Polyn et al., 2009) or features of the environmental context associated with each studied item (e.g., Isarida & Isarida, 2007). This general framework could also be adapted to a wide variety of tasks beyond free recall; for example, one could select components of neural activity that vary with spatial location during virtual navigation tasks (e.g., Ekstrom et al., 2003; Newman et al., 2007).

6.4.2 A model-based framework for interpreting neural data

From behavioral data alone, it is impossible to determine whether failure to recall a given item is due to an error in encoding or recall (or both). Behavioral models make predictions about encoding and recall errors by making assumptions about how these processes work (which may or may not be supported by neural recordings). Neural recordings can provide information about whether encoding or recall has succeeded (Sederberg et al., 2007) and about the dynamics of the recall phase (Polyn et al., 2005). To date, however, most analyses of neural recordings during free recall have ignored the well-characterized stereotyped trends identified in the behavioral literature including the primacy, recency, and contiguity effects (e.g., Murdock, 1962; Kahana, 1996; Kahana, Howard, & Polyn, 2008). (One exception is the analysis presented in Chapter 5.)

Because patterns identified in neural recordings are typically confounded with behavioral patterns, it has been difficult to gain a deep understanding of the neural processes underlying encoding and recall. To address this confound, one could go

beyond the analysis presented in Chapter 5 by using the Temporal Context Model (TCM; Howard & Kahana, 2002), an established context-based memory model, as a framework for interpreting neural recordings. Here I do not assume that TCM (or any model, for that matter) is an accurate description of how the brain retrieves episodic memories (although the analysis in Chapter 5 provides good support for the context reinstatement hypothesis). Instead, I propose to use TCM's ability to fit behavioral trends in the data as a means of interpreting neural recordings while taking behavioral trends into account.

One potential application of this approach would be to implement a model-based SME analysis. For example, one could fit TCM's parameters to a single participant's behavioral data. Running TCM many times using the best-fitting parameters from that participant would yield an estimated probability of recalling each studied word. One could then use the PCA-based framework described in Section 6.4.1 to identify components of neural activity measured during study that predict probability of subsequent recall (according to TCM's predictions). Whereas standard SME analyses (e.g., Paller & Wagner, 2002) implicitly assume that each studied word is either perfectly encoded (if it is subsequently recalled) or perfectly ignored (if it is not subsequently recalled), the increased sensitivity of the model-based SME analysis may lead to the identification of new neural signatures of successful encoding. In this way, a model-based framework for interpreting neural data may provide novel insights into the neural mechanisms underlying the encoding and recall of episodic memories.

6.4.3 A model-based framework for tracking the neural representation of temporal context

An important question raised by the analysis in Chapter 5 concerns the detailed nature of the neural representation of temporal context. Neither the analysis in Chapter 5 nor the behavioral data distinguish between a context representation that drifts at random through a multidimensional space (e.g., Estes, 1955), or whether contextual drift is driven by inputs, such as studied list items (e.g., Howard & Kahana, 2002; Sederberg et al., 2008; Polyn et al., 2009; Socher et al., 2009). One method of getting at this question would be to select neural features that both vary with the meanings of studied items (as in Chapter 4) and exhibit temporal autocorrelations (as in Chapter 5). If these neural features exhibited a neural signature of context reinstatement (Fig. 5.4), this would indicate that semantic features of the studied words contribute to the temporal context representation.

The hierarchical Bayesian modeling framework developed by Socher et al. (2009) suggests an interesting alternative means of studying the context representation in finer detail. In particular, their framework could be extended to incorporate neural data (e.g. ECoG, fMRI) collected as participants study and recall words, which would allow the neural signature of temporal context to be tracked over the course of the experiment. Whereas the analysis presented in Chapter 5 identified a neural signature of temporal context reinstatement in data averaged over many lists, the hierarchical Bayesian-based approach would provide a means of tracking the temporal context representation as it evolved during each individual word presentation and recall. The ability to track the temporal context representation at this level of detail would have far-reaching implications for fields ranging from cognitive neuroscience to education to law enforcement.

6.5 Concluding remarks

To date, no theoretical model provides a complete explanation of how the brain supports acquisition, storage, and retrieval. Similarly, no known recording device allows researchers to record all brain signals (e.g. membrane potentials from all neurons) relevant to these processes. Nonetheless, both theoretical models and neural recordings can be informative. To the extent that models and neural recordings contain independent information, each approach can inform the other. In this way, analyses that are simultaneously informed by model predictions and recorded neural data seem most promising at elucidating the neural machinery underlying cognition.

References

- Aguilar, M., & Stiles, W. S. (1954). Saturation of the rod mechanism of the retina at high light levels. *Journal of Modern Optics*, *1*, 59 - 65.
- Ala-Laurila, P., Donner, K., Crouch, R. K., & Cornwall, M. C. (2007). Chromophore switch from 11-cis-dehydroretinal (A2) to 11-cis-retinal (A1) decreases dark noise in salamander red rods. *Journal of Physiology*, *585*, 57 – 74.
- Ala-Laurila, P., Pahlberg, J., Koskelainen, A., & Donner, K. (2004). On the relation between the photoactivation energy and the absorbance spectrum of visual pigments. *Vision Research*, *44*, 2153 – 2158.
- Andersen, A. H., Gash, D. M., & Avison, M. J. (1999). Principle component analysis of the dynamic response measured by fMRI: a generalized linear systems framework. *Magnetic Resonance Imaging*, *17*(6), 795 – 815.
- Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*(2), 97-123.
- Annibaldi, S. V., & Hopcraft, K. I. (2002). Random walks with power-law fluctuations in the number of steps. *Journal of Physics A: Mathematical and General*, *35*, 8635 - 8645.
- Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing. *Network: Computation in Neural Systems*, *3*(2), 213-251.
- Atick, J. J., Li, Z. P., & Redlich, A. N. (1992). Understanding retinal color coding

- from first principles. *Neural Computation*, 4(4), 559-572.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, p. 89-105). New York: Academic Press.
- Balasubramanian, V., & Berry, M. J. (2002). A test of metabolically efficient coding in the retina. *Network*, 13, 531 - 552.
- Balasubramanian, V., Kimber, D., & Berry, M. J. (2001). Metabolically efficient information processing. *Neural Computation*, 13, 799 - 815.
- Ball, T., Demandt, E., Mutschler, I., Neitzel, E., Mehring, C., Vogt, K., et al. (2008). Movement related activity in the high gamma range of the human EEG. *Neuroimage*.
- Barlow, H. B. (1956). Retinal noise and absolute threshold. *Journal of the Optical Society of America*, 46, 634 - 639.
- Barlow, H. B. (1957). Purkinje shift and retinal noise. *Nature*, 179, 255 - 256.
- Baylor, D. A., Lamb, T. D., & Yau, K. W. (1979). Responses of retinal rods to single photons. *Journal of Physiology (London)*, 288, 613 - 634.
- Baylor, D. A., Nunn, B. J., & Schnapf, J. L. (1984). The photocurrent, noise, and spectral sensitivity of rods of the monkey macaca fascicularis. *Journal of Physiology (London)*, 357, 575 - 607.
- Belitski, A., Gretton, A., Magri, C., Murayama, Y., Montemurro, M., & Logothetis, N. (2008). Local field potentials and spiking activity in primary visual cortex convey independent information about natural visual stimuli. *J. Neurosci*.
- Berger, T. O. (1985). *Statistical decision theory and bayesian analysis*. New York: Springer-Verlag.
- Blackwell, D., & Girschick, M. A. (1954). *Theory of games and statistical decisions*. New York: Wiley.

- Bossomaier, T. R. J., Snyder, A. W., & Hughes, A. (1985). Irregularity and aliasing: solution? *Vision Research*, 25, 145 - 147.
- Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (p. 85-121). New York: John Wiley and Sons.
- Bowmaker, J. K. (1991). The evolution of vertebrate visual pigments and photoreceptors. In J. R. Cronly-Dillon & R. L. Gregory (Eds.), *Evolution of the eye and visual system* (p. 63 - 81). CRC Press.
- Bowmaker, J. K., & Kunz, Y. W. (1987). Ultraviolet receptors, tetrachromatic colour vision and retinal mosaics in the brown trout (*salmo trutta*): age-dependent changes. *Vision Research*, 27, 2101 - 2108.
- Brainard, D. H. (1993). Preceptual variability as a fundamental axiom of perceptual science: comments on ashby and lee. In S. C. Masin (Ed.), *Foundations of perceptual theory*. Elsevier Science Publishers.
- Brainard, D. H., Williams, D. R., & Hofer, H. (2008). Trichromatic reconstruction from the interleaved cone mosaic: Bayesian model and the color appearance of small spots. *Journal of Vision*, 8(5), 1 - 21.
- Brenner, N., Bialek, W., & de Ruyter van Steveninck, R. (2000). Adaptive rescaling maximizes information transmission. *Neuron*, 26, 695 - 702.
- Buchsbaum, G., & Gottschalk, A. (1983). Trichromacy, opponent colour coding and optimum colour information transmission in the retina. *Proceedings of the Royal Society of London B*, 220, 89-113.
- Burke, J. F., Zaghoul, K. A., Jacobs, J., & Kahana, M. J. (in preparation). Gamma coherence decreases and broadband spectral changes mark successful memory encoding.
- Burns, M., & Arshavsky, V. (2005). Beyond counting photons: trials and trends in

- vertebrate visual transduction. *Neuron*, 48, 387 - 401.
- Burton, G. J., & Moorehead, I. R. (1987). Color and spatial structure in natural images. *Applied Optics*, 26, 157-170.
- Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nature Neuroscience*, 7, 446–451.
- Buzsáki, G. (2006). *Rhythms of the brain*. New York: Oxford University Press.
- Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., et al. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, 313(5793), 1626–1628.
- Caplan, J. B., Madsen, J. R., Schulze-Bonhage, A., Aschenbrenner-Scheibe, R., Newman, E. L., & Kahana, M. J. (2003). Human theta oscillations related to sensorimotor integration and spatial learning. *Journal of Neuroscience*, 23, 4726–4736.
- Chen, C., Lakatos, P., Shah, A. S., Mehta, A. D., Givre, S. J., Javitt, D. C., et al. (2007). Functional anatomy and interaction of fast and slow visual pathways in macaque monkeys. *Cerebral Cortex*, 17, 1561 - 1569.
- Chittka, L., & Menzel, R. (1992). The evolutionary adaptation of flower colours and the insect pollinators' colour vision. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 171, 171 - 181.
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: John Wiley & Sons.
- Cox, D., & Savoy, R. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, 19, 261 - 270.
- Crone, N. E., Korzeniewska, A., & Franaszczuk, P. (2011, Jan). Cortical gamma responses: searching high and low. *International Journal of Psychophysiology*, 79(1), 9-15.

- Crone, N. E., Miglioetti, D. L., Gordon, B., Sieracki, J. M., Wilson, M. T., Uematsu, S., et al. (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. I. Alpha and beta event-related desynchronization. *Brain*, *121*, 2271 - 2299.
- Crone, N. E., Miglioretti, D. L., Gordon, B., & Lesser, R. P. (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain*, *121*(12), 2301.
- Crone, N. E., Sinai, A., & Korzeniewska, A. (2006). High-frequency gamma oscillations and human brain mapping with electrocorticography. *Progress in Brain Research*, *159*, 275 - 295.
- Cruikshank, S. J., Rose, H. J., & Metherate, R. (2001). Auditory thalamocortical synaptic transmission in vitro. *Journal of neurophysiology*, *87*, 361 - 384.
- Cummings, M. E. (2004). Modelling divergence in luminance and chromatic detection performance across measured divergence in surfperch (embiotocidae) habitats. *Vision Research*, *44*, 1127 - 1145.
- Curcio, C. A., Allen, K., Sloan, K., Lerea, C., Klock, I., & Milam, A. (1991). Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *Journal of Comparative Neurology*, *312*, 610 - 624.
- D'Agostino, P. R. (1969). The blocked-random effect in recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, *8*, 815-820.
- Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, *112*, 3-42.
- Davelaar, E. J., Usher, M., Haarmann, H. J., & Goshen-Gottstein, Y. (2008). Postscript: Through TCM, STM shines bright. *Psychological Review*, *115*(4),

1116-1118.

- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17-22.
- Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology*, *54*, 180-187.
- Demonet, J., Chollet, F., S., R., Cardebat, D., Nespoulous, J., Wise, R., et al. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, *116*(6), 1753 – 1768.
- Demontis, G. C., Bisti, S., & Cervetto, L. (1993). Light sensitivity, adaptation and saturation in mammalian rods. *Prog Brain Research*, *95*, 15 - 24.
- DeVries, S., Qi, X., Smith, R. G., Makous, W., & Sterling, P. (2002). Electrical coupling between mammalian cones. *Current Biology*, *12*, 1900 – 1907.
- Doi, E., Inui, T., Lee, T., Wachtler, T., & Sejnowski, T. J. (2003). Spatiochromatic receptive field properties derived from information-theoretic analysis of cone mosaic responses to natural scenes. *Neural Computation*, *15*(2), 397 – 417.
- Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S., & Knight, R. T. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *Journal of Neurophysiology*, *94*, 4269 - 4280.
- Ekstrom, A. D., Caplan, J., Ho, E., Shattuck, K., Fried, I., & Kahana, M. (2005). Human hippocampal theta activity during virtual navigation. *Hippocampus*, *15*, 881–889.
- Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., et al. (2003). Cellular networks underlying human spatial navigation. *Nature*, *425*, 184–187.
- Ekstrom, A. D., Viskontas, I., Kahana, M. J., Jacobs, J., Upchurch, K., Bookheimer,

- S., et al. (2007). Contrasting roles of neural firing rate and local field potentials in human memory. *Hippocampus*, 17(8), 606–17.
- Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, 62, 145-154.
- Fell, J., Klaver, P., Lehnertz, K., Grunwald, T., Schaller, C., Elger, C. E., et al. (2001). Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling. *Nature Neuroscience*, 4(12), 1259-1264.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4, 2379-2394.
- Fried, I., Wilson, C., Maidment, N., Engel, J. J., Behnke, E., Fields, T., et al. (1999). Cerebral microdialysis combined with single-neuron and electroencephalographic recording in neurosurgical patients. *Journal of Neurosurgery*, 91, 697–705.
- Fries, P., Nikolić, D., & Singer, W. (2007). The gamma cycle. *Trends in Neurosciences*, 30(7), 309–316.
- Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291(5508), 1560–1563.
- Fu, Y., Kefalov, V., Luo, D. G., Xue, T., & Yau, K. W. (2008). Quantal noise from human red cone pigment. *Nature Neuroscience*, 11, 565 - 571.
- Garrigan, P., Ratliff, C., Klein, J. M., Sterling, P., Brainard, D. H., & Balasubramanian, V. (2006, March). *Structure of the primate cone mosaic and the statistics of color in nocturnal images*. Salt Lake City, UT.
- Garrigan, P., Ratliff, C. P., Klein, J. M., Sterling, P., Brainard, D. H., & Balasubramanian, V. (2010). Design of a trichromatic cone array. *PLoS Comput Biol*, 6(2),

e1000677.

- Geisler, W. S. (1987). The ideal observer concept as a modeling tool. In T. C. on Vision (Ed.), *Frontiers of visual science, n* (p. 17-31). Washington, D.C.: National Academy Press.
- Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., & Fried, I. (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. *Science, 3*, 96-101.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton: Chapman & Hall/CRC.
- Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5*, 351-360.
- Glanzer, M., Koppelaar, L., & Nelson, R. (1972). Effects of relations between words on short-term storage and long-term storage. *Journal of Verbal Learning and Verbal Behavior, 11*, 403-416.
- Graham, C. H., & Hsia, Y. (1969). Saturation and the foveal achromatic interval. *Journal of the Optical Society of America, 59*, 993.
- Guthrie, E. R. (1935). *The psychology of learning*. Harper.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*, 2425-2429.
- Henrie, J. A., & Shapley, R. (2005). LFP power spectra in V1 cortex: the graded effect of stimulus contrast. *Journal of Neurophysiology, 94*, 479 - 490.
- Hofer, H., Carroll, J., Neitz, J., Neitz, M., & Williams, D. R. (2005). Organization of the human trichromatic cone mosaic. *Journal of Neuroscience, 25*, 9669-9679.
- Holland, P. W., & Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Communications in Statistics: Theory and Methods,*

A6, 813 - 827.

- Hooke, R. (1969). *The posthumous works of Robert Hooke: With a new introduction by Richard S. Westfall*. Johnson Reprint Corp.
- Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 923–941.
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 269-299.
- Howard, M. W., Viskontas, I. V., Shankar, K. H., & Fried, I. (Submitted). Human medial temporal lobe ensembles maintain and retrieve a gradually-changing state of temporal context. *Submitted*.
- Howard, M. W., Youker, T. E., & Venkatadass, V. (2008). The persistence of memory: Contiguity effects across several minutes. *Psychonomic Bulletin & Review*, 15, 58–63.
- Hsu, A., Smith, R. G., Buchsbaum, G., & Sterling, P. (2000). Cost of cone coupling to trichromacy in primate fovea. *Journal of the Optical Society of America*, 17, 635 – 640.
- Isarida, T., & Isarida, T. K. (2007). Environmental context effects of background color in free recall. *Memory & Cognition*, 35(7), 1620 – 1629.
- Jacobs, G. H. (1981). *Comparative color vision*. New York: Academic Press.
- Jacobs, G. H. (1996). Primate photopigments and primate color vision. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 577-581.
- Jacobs, G. H., & Rowe, M. P. (2004). Evolution of vertebrate colour vision. *Clinical and Experimental Optometry*, 87, 206-216.
- Jacobs, J., & Kahana, M. J. (2009). Neural representations of individual stimuli in humans revealed by gamma-band ECoG activity. *Journal of Neuroscience*,

29(33), 10203–10214.

- Jacobs, J., Kahana, M. J., Ekstrom, A. D., & Fried, I. (2007). Brain oscillations control timing of single-neuron activity in humans. *Journal of Neuroscience*, 27(14), 3839–3844.
- Jenkins, L. J., & Ranganath, C. (2010). Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. *Journal of Neuroscience*, *In press*.
- Jensen, O., Kaiser, J., & Lachaux, J. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, 30(7), 317–324.
- Jensen, O., & Lisman, J. E. (1998). An oscillatory short-term memory buffer model can account for data on the Sternberg task. *J. Neuroscience*, 18, 10688–10699.
- Jerbi, K., Freyermuth, S., Minotti, L., Kahane, P., Berthoz, A., & Lachaux, J. (2009). Watching Brain TV and Playing Brain Ball: Exploring Novel BCI Strategies using Real-Time Analysis of Human Intracranial Data. *International Review of Neurobiology*, 159–168.
- Juergens, E., Guettler, A., & Eckhorn, R. (1999). Visual stimulation elicits locked and induced gamma oscillations in monkey intracortical- and EEG-potentials, but not in human EEG. *Experimental Brain Research*, 129, 247 - 259.
- Just, M. A., Cherkassky, V. L., Aryal, S., & Mitchell, T. M. (2011). A neurosemantic theory of concrete noun representation based on underlying brain codes. *PLoS One*, 5(1), e8622.
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24, 103–109.
- Kahana, M. J. (2006). The cognitive correlates of human brain oscillations. *Journal of Neuroscience*, 26(6), 1669–1672.
- Kahana, M. J., Howard, M. W., & Polyn, S. M. (2008). Associative retrieval processes in episodic memory. In H. L. Roediger III (Ed.), *Cognitive psychology of memory*.

- Vol. 2 of Learning and memory: A comprehensive reference, 4 vols. (J. Byrne, Editor).*
Oxford: Elsevier.
- Kahana, M. J., Seelig, D., & Madsen, J. R. (2001). Theta returns. *Current Opinion in Neurobiology, 11*, 739–744.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*(1), 141 - 151.
- Kandel, E., Schwartz, J., & Jessell, T. (2000). *Principles of Neural Science*. Appleton & Lange.
- Kanisza, G. (1979). *The organization of vision*. New York: Praeger.
- Kaur, S., Lazar, R., & Metherate, R. (2004). Intracortical pathways determine breadth of subthreshold frequency receptive fields in primary auditory cortex. *Journal of Neurophysiology, 91*, 2551 - 2567.
- Kelber, A., & Roth, L. S. (2006). Nocturnal colour vision - not as rare as we might think. *Journal of Experimental Biology, 209*, 281-288.
- Koch, K., McLean, J., Berry, M., Sterling, P., Balasubramanian, V., & Freed, M. (2004). Efficiency of information transmission by retinal ganglion cells. *Current Biology, 14*, 1523-1530.
- Koch, K., McLean, J., Segev, R., Freed, M. A., Berry, M. J., Balasubramanian, V., et al. (2006). How much the eye tells the brain. *Current Biology, 16*, 1428 - 1434.
- Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace and World.
- Köhler, W. (1947). *Gestalt psychology*. New York: Liveright.
- Kreiman, G. (2007). Brain science: from the very small to the very large. *Current Biology, 17*(17), 768 - 770.
- Kreiman, G., Hung, C., Kraskov, A., Quiroga, R., Poggio, T., & DiCarlo, J. (2006). Object selectivity of local field potentials and spikes in the macaque inferior

- temporal cortex. *Neuron*, 49(3), 433–445.
- Lachaux, J. P., George, N., Tallon-Baudry, C., Martinerie, J., Hugueville, L., Minotti, L., et al. (2005). The many faces of the gamma band response to complex visual stimuli. *NeuroImage*, 25, 491-501.
- Lachaux, J. P., Jerbi, K., Bertrand, O., Minotti, L., Hoffman, D., Schoendorff, B., et al. (2007). A blueprint for real-time functional mapping via human intracranial recordings. *PLoS One*, 2(10), e1094.
- Lachaux, J. P., Jung, J., Mainy, N., Dreher, J. C., Bertrand, O., Baciou, M., et al. (2008). Silence is golden: Transient neural deactivation in the prefrontal cortex during attentive reading. *Cerebral Cortex*, 18(2), 443 - 450.
- Laming, D. (2006). Predicting free recalls. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5), 1146-1163.
- Land, M., & Osorio, D. (2003). Colour vision: colouring the dark. *Current Biology*, 13, R83 - R85.
- Landauer, T. K., & Dumais, S. T. (1997). Solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- Laughlin, S. B. (2001). Energy as a constraint on the coding and processing of sensory information. *Current Opinion in Neurobiology*, 11, 475 - 480.
- Laughlin, S. B., & Hardie, R. C. (1978). Common strategies for light adaptation in the peripheral visual systems of fly and dragonfly. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 128, 319-340.
- Laughlin, S. B., & Sejnowski, T. J. (2003). Communication in neuronal networks. *Science*, 301, 1879-1874.
- Le Bihan, D., Turner, R., Zeffiro, T., Cuenod, C., Jezard, P., & Bonnerot, V. (1993).

- Activation of human primary visual cortex during visual recall: a magnetic resonance imaging study. *Proc Natl Acad Sci U S A*, 90, 11802 – 11805.
- Levin, A., Durand, F., & Freeman, W. T. (2008). *Understanding camera trade-offs through a bayesian analysis of light field projections*. Cambridge, MA: MIT.
- Lewis, A., & Zhaoping, L. (2006). Are cone sensitivities determined by natural color statistics? *Journal of Vision*, 6, 285 – 302.
- Lisman, J. (1999). Relating hippocampal circuitry to function: recall of memory sequences by reciprocal Dentate-CA3 interactions. *Neuron*, 22, 233-242.
- Llinas, R. R., & Steriade, M. (2006). Bursting of thalamic neurons and states of vigilance. *J Neurophysiol*, 95, 3297 – 3308.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1, 476-490.
- Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *Journal of Neuroscience*, 23(10), 3963 - 3971.
- Long, N. M., Öztekin, I., & Badre, D. (2010). Seperable prefrontal cortex contributions to free recall. *J Neurosci*, 30(33), 10967 - 10976.
- Luce, R. D. (2003). Whatever happened to information theory in psychology? *Review of General Psychology*, 7, 183 - 188.
- Lythgoe, J. N., & Partridge, J. C. (1989). Visual pigments and the acquisition of visual information. *Journal of Experimental Biology*, 146, 1 - 20.
- Manning, J. R., Hurst, B., & Brainard, D. H. (2011, February). *Learning receptor types from receptor responses*. Salt Lake City, UT.
- Manns, J. R., Howard, M. W., & Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron*, 56(3), 530–540.
- Marner, L., Nyengaard, J. R., Tang, Y., & Pakkenberg, B. (2003). Marked loss of

- myelinated nerve fibers in the human brain with age. *J Comp Neurol*, 462(2), 139 - 43.
- Massof, R. W. (1977). A quantum fluctuation model for foveal color thresholds. *Vision Research*, 17, 565 – 570.
- Mazzoni, N., Whittingstall, K., Brunel, N., Logothetis, N. K., & Panzeri, S. (2010, September). Understanding the relationships between spike rate and delta/gamma frequency bands of lfps and eegs using a local cortical network model. *NeuroImage*, 52(3), 956-72.
- Miller, K., den Nijs, M., Shenoy, P., Miller, J., Rao, R., & Ojemann, J. (2007). Real-time functional brain mapping using electrocorticography. *Neuroimage*, 37(2), 504–507.
- Miller, K., Leuthardt, E. C., Schalk, G., Rao, R. P. N., Anderson, N. R., Moran, D. W., et al. (2007). Spectral Changes in Cortical Surface Potentials during Motor Movement. *Journal of Neuroscience*, 27(9), 2424.
- Miller, K., Zanos, S., Fetz, E. E., den Nijs, M., & Ojemann, J. (2009). Decoupling the cortical power spectrum reveals real-time representation of individual finger movements in humans. *Journal of Neuroscience*, 29(10), 3132 - 3137.
- Miller, K. J., Shenoy, P., den Nijs, M., Sorensen, L. B., Rao, R. P. N., & Ojemann, J. G. (2008). Beyond the gamma band: the role of high-frequency features in movement classification. *IEEE Transactions on Biomedical Engineering*, 55(5), 1634 - 1637.
- Milstein, J., Mormann, F., Fried, I., & Koch, C. (2009). Neuronal shot noise and Brownian $1/f^2$ behavior in the local field potential. *PLoS One*, 4(2), e4338.
- Mitchell, T., Shinkareva, S., Carlson, A., Chang, K., Malave, V., Mason, R., et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880), 1191.

- Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiological Review*, 65(1), 37 - 100.
- Mollon, J. D., & Bowmaker, J. K. (1992). The spatial arrangement of cones in the primate fovea. *Nature*, 360, 677 - 679.
- Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., & Malach, R. (2005, Aug). Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science*, 309(5736), 951–954.
- Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64, 482-488.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments and Computers*, 36(3), 402-407.
- Newman, E. L., Caplan, J. B., Kirschen, M. P., Korolev, I. O., Sekuler, R., & Kahana, M. J. (2007). Learning your way around town: How virtual taxicab drivers learn to use both layout and landmark information. *Cognition*, 104(2), 231–253.
- Nir, Y., Fisch, L., Mukamel, R., Gelbard-Sagiv, H., Arieli, A., Fried, I., et al. (2007). Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. *Current Biology*, 17(15), 1275–1285.
- Nunez, P. L., Srinivasan, R., Westdorp, A. F., Wijesinghe, R. S., Tucker, D. M., Silberstein, R. B., et al. (1997). EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalography and Clinical Neurophysiology*, 103(5), 499–515.
- Osipova, D., Takashima, A., Oostenveld, R., Fernandez, G., Maris, E., & Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of

- declarative memory. *J Neurosci*, 26(28), 7523–7531.
- Osorio, D., & Vorobyev, M. (1996). Colour vision as an adaptation to frugivory in primates. *Proceedings of the Royal Society of London B*, 263, 593 - 599.
- Pagan, M., Lega, B. C., Jacobs, J., & Kahana, M. J. (in preparation). The influence on memory formation of narrowband and broadband shifts in intracranial EEG data.
- Pal, N. R., & Pal, S. K. (1993). A review of image segmentation techniques. *Pattern Recognition Letters*, 26, 1277 - 1294.
- Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, 6(2), 93-102.
- Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, 321, 1322 - 1327.
- Percival, D. B., & Walden, A. T. (1993). *Spectral analysis for physical applications*. Cambridge, U. K.: Cambridge University Press.
- Pernier, J., Perrin, F., & Bertrand, O. (1988). Scalp current density fields: concepts and properties. *Electroencephalogr Clin Neurophysiol*, 69, 385 – 389.
- Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., & Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nature Neuroscience*, 5, 805–811.
- Pfurtscheller, G., & Cooper, R. (1975). Frequency dependence of the transmission of the EEG from cortex to scalp. *Electroencephalography and Clinical Neurophysiology*, 38, 93–96.
- Platt, J. R. (1956). Wavelength dependence of radiation-noise limits on sensitivity of infrared photodetectors. *Journal of the Optical Society of America*, 46, 609 – 610.

- Plaut, D. C., & McClelland, J. L. (2010). Locating object knowledge in the brain: Comments on Bowers's (2009) attempt to revive the grandmother cell hypothesis. *Psychol Rev*, *117*(1), 284 - 290.
- Pollio, H. R., Kasschau, R. A., & DeNise, H. E. (1968). Associative structure and the temporal characteristics of free recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 190-197.
- Polyn, S. M., & Kahana, M. J. (2008). Memory search and the neural representation of context. *Trends in Cognitive Sciences*, *12*, 24–30.
- Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, *310*, 1963-1966.
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156.
- Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, *17*, 132-138.
- Pratt, W. K. (1978). *Digital image processing*. New York: John Wiley & Sons.
- Quiroga, R. Q., Nadasdy, Z., & Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, *16*, 1661–1687.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93-134.
- Ramayya, A. G., Manning, J. R., Jacobs, J., & Kahana, M. J. (2010). *The firing rate-LFP relation changes as a function of firing rate in humans*. San Diego, CA.
- Rasch, M. J., Gretton, A., Murayama, Y., Maass, W., & Logothetis, N. K. (2008). Inferring spike trains from local field potentials. *Journal of Neurophysiology*,

99, 1461-1476.

- Regan, B. C., Julliot, C., Simmen, B., Vienot, F., Charles-Dominique, P., & Mollon, J. D. (2001). Fruits, foliage and the evolution of primate color vision. *Philosophical Transactions: Biological Sciences*, 356, 229 - 283.
- Ribak, C. E., & Dashtipour, K. (2002). Neuroplasticity in the damaged dentate gyrus of the epileptic brain. *Prog Brain Res*, 136, 319 – 328.
- Rieke, F., & Baylor, D. A. (2000). Origin and functional impact of dark noise in retinal cones. *Neuron*, 26, 181-186.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 803-814.
- Rolls, E. T., & Kesner, R. P. (2006). A computational theory of hippocampal function, and empirical tests of the theory. *Progress in Neurobiology*, 79, 1 - 48.
- Ruderman, D. L., Cronin, T. W., & Chiao, C. C. (1998). Statistics of cone responses to natural images: implications for visual coding. *Journal of the Optical Society of America A*, 15, 2036-2045.
- Rundus, D. (1971). An analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 89, 63-77.
- Rushton, W. A. H. (1962). Visual pigments in man. *Scientific American*, 207, 120 - 132.
- Schacter, D. L. (1987). Memory, amnesia, and frontal lobe dysfunction. *Psychobiology*, 15, 21-36.
- Schnapf, J. L., Nunn, B. J., Meister, M., & Baylor, D. A. (1990). Visual transduction in cones of the monkey macaca fascicularis. *Journal of Physiology (London)*, 427, 681-713.
- Schneeweis, D. M., & Schnapf, J. L. (1999). The photovoltage of macaque cone

- photoreceptors: adaptation, noise, and kinetics. *Journal of Neuroscience*, 19, 1203 - 1216.
- Schneeweis, D. M., & Schnapf, J. L. (2000). Noise and light adaptation in rods of the macaque monkey. *Visual Neuroscience*, 17, 659 - 666.
- Schneider, D. (1969). Insect olfaction: deciphering system for chemical messages. *Science*, 163, 1031 - 1037.
- Scholes, J. H. (1975). Colour receptors and their synaptic connexions in the retina of a cyprinid fish. *Philosophical Transactions of the Royal Society of London B*, 270, 61-118.
- Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, 115, 893-912.
- Sederberg, P. B., Kahana, M. J., Howard, M. W., Donner, E. J., & Madsen, J. R. (2003). Theta and gamma oscillations during encoding predict subsequent recall. *Journal of Neuroscience*, 23(34), 10809–10814.
- Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., Litt, B., Brandt, A., et al. (2007). Gamma oscillations distinguish true from false memories. *Psychological Science*, 18(11), 927–932.
- Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., McCarthy, D. C., Brandt, A., et al. (2007). Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cerebral Cortex*, 17(5), 1190–1196.
- Sellick, P. M., Patuzzi, R., & Johnstone, B. M. (1982). Measurement of basilar membrane motion in the guinea pig using the mossbauer technique. *Journal of the Acoustic Society of America*, 72, 131 - 141.
- Semon, R. W. (1923). *Mnemonic psychology* (b. duffy, trans.). London: George Allen & Unwin (Original work published 1909).
- Shinkareva, S. V., Mason, R. A., Malave, V. L., Wang, W., Mitchell, T. M., & Just, M. A.

- (2008). Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS One*, *e1394*, 1-9.
- Simoncelli, E. P. (2005). Statistical modeling of photographic images. In A. Bovik (Ed.), *Handbook of image and video processing* (p. 431-441). Academic Press.
- Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci*, *18*, 555–586.
- Skaggs, W. E., McNaughton, B. L., Wilson, M. A., & Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, *6*, 149–172.
- Skorupski, P., & Chittka, L. (2008). Towards a cognitive definition of colour vision. Available from *Nature Precedings*, <http://hdl.handle.net/10101/npre.2008.1766.1>.
- Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: a review and meta-analysis. *Psychonomic Bulletin and Review*, *8*(2), 203-220.
- Snyder, A. W., Stavenga, D. G., & Laughlin, S. B. (1977). Spatial information capacity of compound eyes. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, *116*, 183-207.
- Socher, R., Gershman, S., Perotte, A., Sederberg, P., Blei, D., & Norman, K. (2009). A Bayesian analysis of dynamics in free recall. *Advances in Neural Information Processing Systems*, *22*.
- Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B*, *216*, 427 - 459.
- Steriade, M., Amzica, F., & Contreras, D. (1996). Synchronization of fast (30 – 40 Hz) spontaneous cortical rhythms during brain activation. *Journal of Neuroscience*, *16*, 392 – 417.
- Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for

- predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*. Washington, DC: American Psychological Association.
- Tamura, T., Nakatani, K., & Yau, K. W. (1989). Light adaptation in cat retinal rods. *Science*, *245*, 755 - 757.
- Tanji, K., Suzuki, K., Delorme, A., Shamoto, H., & Nakasato, N. (2005). High-frequency γ -band activity in the basal temporal cortex during picture-naming and lexical-decision tasks. *Journal of Neuroscience*, *25*(13), 3287–3293.
- Thomson, E. E., & Kristan, W. B. (2005). Quantifying stimulus discriminability: a comparison of information theory and ideal observer analysis. *Neural Computation*, *17*, 741 - 778.
- Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford.
- Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*, 381-391.
- Twer, T. von der, & MacLeod, D. I. A. (2001). Optimal nonlinear codes for the perception of natural colors. *Network: Computation in Neural Systems*, *12*, 395-407.
- van Hateren, J. H. (1992). A theory of maximizing sensory information. *Biological Cybernetics*, *68*, 23-29.
- van Hateren, J. H. (1993). Spatial, temporal and spectral pre-processing for colour vision. *Proceedings of the Royal Society of London Series B - Biological Sciences*, *251*, 61-68.
- Viviani, R., Grön, G., & Spitzer, M. (2005). Functional principal component analysis of fMRI data. *Human Brain Mapping*, *24*, 109 – 129.
- von der Malsburg, C. (1981). *The correlation theory of brain function* (Internal Re-

- port No. 81-2). Göttingen, Germany: Max-Planck-Institute for Biophysical Chemistry.
- Walls, G. L. (1942). *The vertebrate eye and its adaptive radiation*. New York.
- Walraven, J., Enroth-Cugell, C., Hood, D. C., MacLeod, D. I. A., & Schnapf, J. L. (1990). The control of visual sensitivity: receptor and postreceptor processes. In L. Spillman & J. S. Werner (Eds.), *Visual perception: the neurophysiological foundations* (p. 53-101). San Diego: Academic Press.
- Walraven, P. L. (1962). *On the mechanisms of colour vision*. Soesterberg, The Netherlands: TNO.
- Wassle, H., & Riemann, H. J. (1978). The mosaic of nerve cells in the mammalian retina. *Proceedings of the Royal Society of London Series B*, 200, 441-461.
- Watson, A. B. (1987). The ideal observer concept as a modeling tool. In T. C. on Vision (Ed.), *Frontiers of visual science, n* (p. 32-37). Washington, D.C.: National Academy Press.
- Welch, G. B., & Burnett, C. T. (1924). Is primacy a factor in association-formation. *American Journal of Psychology*, 35, 396-401.
- Witter, M. (2002). The parahippocampal region: past, present and future. In W. M. & F. Wouterlood (Eds.), *The parahippocampal region: Organization and role in cognitive functions* (pp. 3-19). Oxford: Oxford Univ. Press.
- Yee, E., Drucker, D., & Thompson-Schill, S. (2010). fMRI-adaptation evidence of overlapping neural representations for objects related in function or manipulation. *NeuroImage*, 50, 753 - 763.
- Yellott, J. (1982). Spectral analysis of spatial sampling by photoreceptors: topological disorder prevents aliasing. *Vision Research*, 22, 1205 - 1210.
- Yellott, J. (1983). Spectral consequences of photoreceptor sampling in the rhesus monkey. *Science*, 221, 383 - 385.

- Yin, L., Smith, R. G., Sterling, P., & Brainard, D. H. (2006). Chromatic properties of horizontal and ganglion cell responses follow a dual gradient in cone opsin expression. *Journal of Neuroscience*, *26*, 12351 - 12361.
- Zhong, Y., Wang, H., Lu, G., Zhang, Z., Jiao, Q., & Liu, Y. (2009). Detecting functional connectivity in fMRI using PCA and regression analysis. *Brain Topography*, *22*, 134 – 144.