

THE RELATIONSHIP BETWEEN EFFICIENT CODING
OF NATURAL SCENES IN THE HUMAN VISUAL SYSTEM
AND STATISTICAL REGULARITIES IN ART

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Daniel Jacob Graham

May 2008

© 2008 Daniel Jacob Graham

THE RELATIONSHIP BETWEEN EFFICIENT CODING
OF NATURAL SCENES IN THE HUMAN VISUAL SYSTEM
AND STATISTICAL REGULARITIES IN ART

Daniel Jacob Graham, Ph. D.

Cornell University 2008

Natural images have been shown to exhibit predictable statistical properties, including consistent pairwise luminance statistics, spatiotemporal statistics, and contrast and intensity distributions. Since evolution has presumably selected neural coding strategies that are highly efficient with respect to representational, metabolic and developmental resources, statistical regularities can be used to predict why visual systems show the visual coding strategies they do. This dissertation begins with a new theoretical foundation of efficient visual coding and it describes a suite of studies aimed at testing and expanding specific hypotheses of efficient coding of natural images. It is specifically concerned with applying notions of coding efficiency deduced from statistical regularities of natural scenes to images created for human viewing, especially artworks. From the perspective of visual coding, paintings are a rich and largely unexplored class of images. As will be shown, the fact that paintings (even highly abstract ones) share the same basic statistical redundancies as natural scenes suggests that humans exploit efficient coding strategies used by the brain in order to make two dimensional representations that can be viewed by the eye. The findings presented in Chapter 2, which show a model of how retinal processing can be matched to regularities in the spatial frequency power spectrum of scenes, are especially relevant to art. I argue that because evolution appears to have chosen retinal

coding schemes that are efficient with respect to statistical regularity in natural scenes, art whose statistics deviate strongly from such regularity will be attempted rarely. The proposals regarding coding efficiency are employed to explain why art through the ages shows the statistical regularity it does. This “perceptibility hypothesis” argues against explanations for art’s statistical regularity that invoke universal aesthetics. However, artists must fundamentally alter some statistical properties of scenes in order to depict them in paint. A statistical model of how artists compress the large dynamic range of luminances in scenes into the far smaller range available in paint is also presented. I show that no single functional form can describe artists’ nonlinear luminance compression strategies, and I propose that each painter’s compression strategy may be characteristic to her work.

BIOGRAPHICAL SKETCH

Daniel Graham graduated from Middlebury College with a BA in physics and a minor in art history in 2001. He began the Ph.D. program in the physics department at Cornell that year as part of the National Science Foundation's interdisciplinary IGERT program, finishing his MS in 2004. He then transferred to the psychology department to work with Prof. David Field, whom he had met through the IGERT program. After acquiring a background in visual neuroscience, Graham drew upon his knowledge of statistical physics and art history in order to test new ideas about efficient coding in the human visual system. He will begin a post-doctoral fellowship at Dartmouth College with Prof. Dan Rockmore where he will apply many of these ideas to problems of art authentication and art historical scholarship. In his spare time, Graham enjoys traveling, cooking, playing ultimate Frisbee and hiking with his dog, Shelby.

ACKNOWLEDGMENTS

I wish to thank David Field for his many years of patient instruction and collaboration. I am deeply grateful for all the experience and wisdom he has imparted to me, and that he believed that I was worthy of gaining such profound knowledge. All young scientists should be so lucky as to have an advisor like David. Thanks also go to James Cutting, Barbara Finlay, Steve Strogatz and Bruce Lewenstein—they have aided generously in my education and I respect them all as scholars of the first order. Damon Chandler contributed greatly to this work as well and I thank him and the other members of the Field Lab for their support: Sherry Xian, Mark Albert and Jordan DeLong. I am indebted to Herbert F. Johnson Museum, Cornell University, which granted me unrestricted access to its entire digital collection of art. Franklin Robinson, Julie Magura, Elizabeth Emrich, Renee Grinnell and Meg Elliott each played an important role in my collaboration with the museum. The following funding sources supported the work collected in this dissertation: NIH EY015393 grant to DJG; NSF IGERT program in nonlinear dynamics institutional grant; Cornell Provost's Diversity Fellowship; NGA contract HM 1582-05-C-0007 to DJF; NSF LOCNET program. Finally, I thank Reanna—this would not have been possible without your love.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH	iii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	vii
CHAPTER 1: INTRODUCTION. A FRAMEWORK FOR MODELS OF EFFICIENT CODING IN THE VISUAL SYSTEM	1
<i>1.1. Background</i>	3
<i>1.1.1. Efficient for What Task?</i>	5
<i>1.1.2. Defining Efficiency</i>	7
1.2. Representational Efficiency	8
<i>1.2.1. Correlation and Decorrelation</i>	9
<i>1.2.2. Optimal Information Transfer</i>	11
<i>1.2.3. Beyond Correlations: Sparseness and Independence</i>	12
<i>1.2.4. Optimality with Nonlinear Systems</i>	15
1.3. Metabolic efficiency	17
<i>1.3.1. Spike Efficiency</i>	17
<i>1.3.2. Minimum Wiring</i>	19
1.4. Learning Efficiency	21
<i>1.4.1. Sparseness and Invariance</i>	21
<i>1.4.2. Overcompleteness</i>	23
<i>1.4.3. "Hard-Coded" Efficiency</i>	23
<i>1.4.4. Efficient Learning from the Environment</i>	24
<i>1.4.5. Hybrid Strategies: Efficient Innate Learning</i>	25
CHAPTER 2: CAN THE THEORY OF “WHITENING” EXPLAIN THE CENTER-SURROUND PROPERTIES OF RETINAL GANGLION CELL RECEPTIVE FIELDS?	28
2.1 Background	28
<i>2.1.1. “Whitening” and Decorrelation</i>	29
<i>2.1.2. Overcomplete Coding and Response Equalization</i>	33
<i>2.1.3. Response Equalization and Vector Length</i>	35
2.2. Study 1: Calculation of Vector Length Sensitivity of Ganglion Cells	37
<i>2.2.1. Methods</i>	37
<i>2.2.2. Results</i>	38
2.3. Study 2: Decorrelation and Sparseness in Model Ganglion Cells	40
<i>2.3.1. Methods</i>	41
<i>2.3.2. Results</i>	44
<i>2.3.3. Sparseness</i>	45
<i>2.3.4. Compressive Nonlinearities in the Retina</i>	45
2.4. Discussion	48
<i>2.4.1. Contrast Sensitivity and Vector Length</i>	49
<i>2.4.2. Further Constraints</i>	51

2.4.3. <i>Localization</i>	53
2.5. Conclusion	54
2.5.1. <i>Implications for the Study of Art</i>	55
CHAPTER 3: STATISTICAL REGULARITIES OF ART IMAGES AND NATURAL SCENES: SPECTRA, SPARSENESS, AND NONLINEARITIES	56
3.1 Background	56
3.2. Materials	60
3.3. Study 1: Statistical Regularities in Painted Art	61
3.3.1. <i>Methods</i>	61
3.3.2. <i>Results</i>	61
3.3.3. <i>Discussion</i>	63
3.4. Study 2: Sparseness and Nonlinearities	65
3.4.1. <i>Methods</i>	65
3.4.2. <i>Results</i>	65
3.4.3. <i>Discussion</i>	68
3.5. General Discussion	69
CHAPTER 4: VARIATIONS IN INTENSITY STATISTICS FOR REPRESENTATIONAL AND ABSTRACT ART, AND FOR ART FROM THE EASTERN AND WESTERN HEMISPHERES	76
4.1 Introduction	76
4.2. Study 1: Painting Provenance	79
4.2.1. <i>Methods</i>	79
4.2.2. <i>Results</i>	80
4.2.3. <i>Discussion</i>	83
4.3. Study 2: Painting Content	84
4.3.1. <i>Methods</i>	84
4.3.2. <i>Results</i>	85
4.3.3. <i>Discussion</i>	86
4.4. General Discussion	88
4.5. Conclusion	89
CHAPTER 5: CONCLUSION. TOWARDS A THEORY OF NATURAL ART: THE PERCEPTIBILITY HYPOTHESIS AND THE EFFICIENT ARTIST	91
5.1. Background	93
5.2. Efficiency of Regular Spatial Statistics: The Perceptibility Hypothesis	95
5.3. The Artist's Look-Up Table	101
5.3.1 <i>Study 1: Model of the Artist's Look-up Table</i>	105
5.3.2 <i>Study 2: Preference for Artist's Transforms</i>	110
5.3.3 <i>Discussion</i>	115
5.4. Conclusion	118
CHAPTER 6: GENERAL DISCUSSION AND FUTURE DIRECTIONS: PERCEPTIBILITY, FAMILIARITY AND PICTURE LIKING	120

6.1 General Discussion	120
6.2 Future Directions: Perceptibility and Picture Liking	123
<i>6.2.1 Familiarity and Novelty</i>	125
<i>6.2.2 Picture Liking</i>	128
6.3 Conclusion	135
REFERENCES	137

LIST OF FIGURES

Figure 1.1. A process that chooses pixel intensity randomly will produce white noise images.	4
Figure 1.2. Natural scenes adjusted to show amplitude spectrum slopes decreasing from $-.5$ to -1.75 .	9
Figure 1.3. Noise with a spatial frequency amplitude spectrum like that of natural scenes.	12
Figure. 2.1. Vector space representations of efficient coding strategies.	32
Figure 2.2. Comparison of an increasing vector length model of contrast sensitivity, and a constant vector length model.	36
Figure 2.3. Plot of the vector length sensitivity of cells.	40
Figure 2.4. Simulation of receptive field filtering of natural scene image.	43
Figure 2.5. Mean ratio across images of population sparseness for convolutions.	47
Figure 3.1. Log-log plot of spatial frequency versus amplitude for the paintings and natural scenes.	62
Figure 3.2. Histogram of the best-fit slopes for all art and natural scene images.	63
Figure 3.3. Bleeding Rain by Norman Bluhm (left) has an amplitude spectrum that is fit by a line whose slope is -1.14 in log-log coordinates.	64
Figure 3.4. Sparseness of the pixels and of the responses of a difference-of-Gaussians filter.	66
Figure 3.5. Plots showing response sparseness to wavelet filters across 3 scales.	67
Figure 3.6. Plot of median sparseness (kurtosis) values for natural scenes, paintings, and noise.	68
Figure 3.7. A natural scene from van Hateren's collection displayed with linear scaling in luminance and after application of a $\log(x)$ nonlinearity.	69
Figure 3.8. The painting with the steepest (most negative) amplitude spectrum slope in the sample.	73

Figure 4.1. The monotone boundaries shown here demonstrate that a fractal edge by itself can generate a $1/f$ amplitude spectrum.	78
Figure 4.2. Images as classified according to provenance.	81
Figure 4.3. Image categories as determined by the agreement of 6 judges using a forced-choice paradigm.	84
Figure 4.4. Plot of mean amplitude spectrum slope by image category.	86
Figure 5.1. Examples of images with spatial frequency amplitude spectrum slope of approximately 0, -1 and 3.	99
Figure 5.2. Linear scaling of natural scene luminances produces a very dark scene with a few highlights, but log luminance compression can generate an acceptable scene.	103
Figure 5.3. Luminance calibrated image of painting “Taughannock Falls” by Neil Berger.	108
Figure 5.4. Luminance calibrated scene.	108
Figure 5.5. “Artified” scene.	109
Figure 5.6. Luminance transform T mapping Taughannock Falls scene luminance histogram into “Taughannock Falls” painting luminances.	110
Figure 5.7. Example of display viewed by observers for scene number 9 in the test set transformed according to a number of histograms.	114
Figure 5.8. Histogram of preference rankings by transform summed over 3 observers.	114
Figure 6.1. Statistics can be independently varied, such that the image can have an amplitude spectrum that falls as $1/f^{0.5}$ and $1/f^{1.5}$ respectively or it can have a skew of roughly -1 and $+1$, respectively.	122
Figure 6.2. This figure shows a schematic picture of how familiarity and novelty produce movements within a state space of images that show natural scene-like statistics.	131

LIST OF TABLES

Table 4.1. Pixel intensity statistics for art of the Eastern and Western hemispheres, with standard error shown in parentheses.	82
Table 4.2. Table shows how amplitude spectrum slope varies with image content as determined by a forced-choice test.	85
Table 5.1. Preference data for natural scenes transformed so that each image histogram matched the mean histogram for each grouping of art	113

CHAPTER 1

INTRODUCTION: A FRAMEWORK FOR MODELS OF EFFICIENT CODING IN THE VISUAL SYSTEM

Artworks are artifacts of human perception. But they are a class of images that have rarely been considered to be relevant to the study of efficiency in the visual system. The major goal of the work described here is to demonstrate that a careful examination of statistical properties of artworks can illuminate visual coding in humans and it can also grant insights into trends, commonalities and variations across art history. Furthermore, it can unite these quite disparate fields of study.

What must a human artist do to a two-dimensional surface in order to produce an image, which is perceptible by other humans? Some authors propose that pictures that succeed in being perceived are the ones that are easy to “see” and to comprehend. Seeing pictures, in this view, requires some degree of “fluency” in terms of the affect-level response to an image (See e.g.. Winkielman *et al.* 2003). Some argue that highly informative pictures or scenes are like sedative drugs. (Biederman and Vessel 2006). Many researchers have noted that humans like images they have seen before (Zajonc, 1968) while others argue that humans and other creatures seek out novel pictures and scenes, even at the risk of injury to themselves (Berlyne, 1970). It has been proposed that picture-makers succeed when they are able to capture perceptual constancies to which the visual system is attuned (Zeki, 1998). All of these notions are supported by evidence and indeed they all help inform the hypothesis advanced here, though it is beyond the scope of this dissertation to examine each previous hypothesis in detail. However, none of these ideas addresses a more fundamental question: what is the minimum set of transformations or constraints that would permit art making in the first place? The existing hypotheses have not specified the basic statistical properties of a

hand-made image that would be necessary to achieve basic perception. Here I examine this question and offer the following proposal: Could it be that art is efficient with respect to the early visual system, and therefore that seeing a work of art is dependent at least in part on its efficiency relative to other possible types of images? Put another way, are the minimum necessary constraints for art also the most efficient ones with respect to the visual system?

The specific focus of this dissertation is to address the questions of whether and how much the process of art making can be considered “efficient.” Since an understanding of efficiency in the brain is crucial to the argument, the bulk of this chapter is devoted to defining this notion. Whereas many earlier studies detailed below relate to how the visual system has evolved processing strategies that are efficient with respect to the visual environment of humans, none has considered whether art making exploits these efficiencies.

Throughout our history as modern humans, artists have performed uncontrolled but purposeful experiments in vision. Such experiments reflect not only the statistical structure of the natural world (as will be shown), they also include information about how a visual scene must be depicted using pigment such that other humans can view it. The importance of natural scene-like statistical regularities in art (Chapters 3 and 4) should become clear after discussion of efficiency in neural coding presented in this chapter, and of the experiments regarding ganglion cell coding described in Chapter 2. It will also become apparent that human artists must alter the statistics of the visual world in order to represent the world. A form of nonlinear luminance compression will be shown to be a necessary feature of nearly all hand-made art. The specific form of this compression, which is termed the “artist’s look-up table” (or, loosely, the “artist’s gamma”) is not a simple log-like function. It may in fact be a transfer function, which is characteristic to an artist or to a category of art. Here it should be stressed that color

and local luminance modifications play an enormous role in the perceptual effect of a picture and will not be discussed herein, though these factors may also be amenable to study in terms of efficiency. Of course, composition, context and a host of factors related to affect also play a large role in the impression an image makes on a human viewer. In this dissertation, only basic statistical characteristics of art and low-level perceptual responses to it will be considered. Nevertheless, this approach grants significant insights into the relationship between human art making and the human visual system.

1.1 Background

Images and movies of the natural world are known to share a variety of statistical regularities. Such stimuli show consistent spatial statistics (e.g., Field, 1987, 1994; Burton and Moorhead, 1987), spatio-temporal statistics (van Hateren and Ruderman, 1998), contrast and intensity distributions (e.g., Brady and Field, 2000; Frazor and Geisler, 2006), and chromatic structure (e.g., Webster and Mollon, 1997; Hoyer and Hyvärinen, 2000). Consider the images shown in Figure 1. The first row shows white noise and the second shows natural scenes. A process that generates random white noise images will generate all possible images with equal probability. However, because of the statistical regularities of natural scenes, the possibility of seeing a natural scene generated from a white noise process is extremely low. For even a small 8x8 pixel patch, the entropy of the white noise patches with 8 bits of grey level is $8 \times 8 \times 8$ or 512 bits. This results in a total of 2^{512} , or about 10^{154} equally probable images. Natural scenes are considerably redundant (i.e., they have much lower entropy) with estimates that they contain approximately 40% of the entropy of white noise for small (8x8) patches, and still lower relative entropy for larger patches (Chandler and Field, 2007). This difference would in turn imply that one white noise

pattern out of 10^{90} would have the basic statistics of natural scenes (assuming a flat distribution). Although the precise ratio requires an estimate of the true distribution, the estimate does demonstrate just how redundant natural scenes are. Because of this redundancy, it is possible to build a visual system that is efficient if it is specifically dedicated to representing such an environment. A visual system can reduce the size of its problem space and focus its efforts on what it is likely to encounter in the world. Because of this redundancy, visual systems can be shaped by evolution and development to take advantage of the environment.

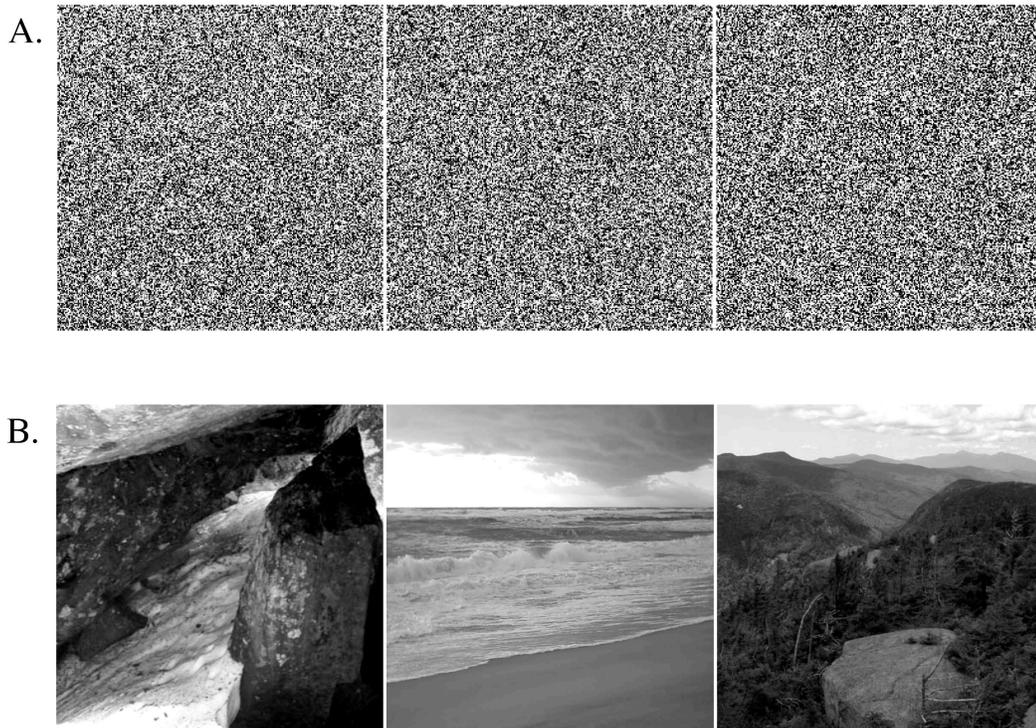


Figure 1.1. A process that chooses pixel intensity randomly will produce white noise images (A.). Images resembling natural scenes (B.) would almost never occur in such a process.

Using the supposition that the visual system employs codes that approach optimal efficiency with respect to predictable (i.e., redundant) structure, one can go some way

in explaining why visual neurons show the coding properties they do. A number of studies have shown success at predicting many of the basic linear properties of visual neurons (e.g., Field, 1987, 1994; Zetzsche and Röhrbein, 2001; Wiskott and Sejnowski, 2002; Simoncelli and Olshausen, 2001). In the following sections I will review some of this work. But it should be noted that this work and any extension of it depend on the appropriate definition of efficiency. In the classical definition of efficiency used in engineering, a system is most efficient if all of its work (within limits imposed by thermodynamics) is done in service of its task. In this view, efficiency presumes that the system has some well-defined tasks, and that the system's design is a direct reflection of the need to do as little work as possible to achieve those tasks. But how does one define these tasks for biological systems? The framework developed in this section underlies the argument developed later in this dissertation regarding the relationship between statistical regularities in art and efficient visual coding.

1.1.1 Efficient for What Task?

For any biological system, an account of why the system is structured as it is requires consideration of the interaction among three primary factors: (1) the uses (or goals) of the system; (2) the environment in which the system must function; and (3) the constraints that history and biology put on the design. Within these constraints, the forces of evolution and development shape the design of any neural system.

Many sensory systems, whose task or goal is often readily definable and measurable, have sensitivity that is near their physical limits (e.g., Squires, 2004; Sterling, 2004; Sundar *et al.*, 2003; Aizenberg *et al.*, 2001; Van Essen, 1997; Denk and Webb, 1989; Baylor *et al.*, 1979). With respect to the brain as a whole, however, the only clearly defined task is the most general one in biology: differential

reproductive success. But this goal tells us little about why a visual system would have the structure it does. Clearly, the visual system is involved in vision, but to treat it as an independent sub-system with some basic input/output relationship is too much of a simplification. Certainly, the output of the visual system cannot be reduced to a behavior. And although it may seem computationally reasonable, there is little reason to presume that the output of the visual system is simply some “object-detector”. There are three reasons for this: Object recognition is not a well localized processing task in cortex; no known congenital disorder obliterates this ability exclusively; and patients with visual agnosia caused by lesions typically show severe deficits in other visual faculties (Bloom, 2002; Farah, 2004). Even though object recognition is distributed, one might still propose that object recognition could be the collective goal of the visual system's many parts. But object recognition surely competes for neural real estate with areas whose goals include accurate spatial mapping, motion predictions, and a host of other tasks.

This difficulty in defining the visual system's general goal is likely to prove a significant hurdle for studies attempting to apply efficient coding theories to later stages of processing. In addition, our current ability to predict a neuron's behavior is significantly limited at these later stages. Much of the success of efficient coding techniques has been in the application to neurons relatively early in the visual pathway (retina, lateral geniculate nucleus, and V1). For neurons in these areas, the family of response properties is relatively well defined and it is possible to talk about how the information in the visual image is represented with the array of neurons in each of these areas (although see Olshausen and Field, 2005). This in turn allows us to address the question of whether the array of neurons is processing the information efficiently. However, the approach still requires a precise definition of efficiency.

1.1.2. Defining Efficiency

For those who have proposed theories of optimality in early visual coding, much of the discussion has centered on the metric of efficiency. Most papers in the field focus on what will be called representational efficiency. Such papers employ the tools of information theory, and they have explored the properties of neurons that are involved in representing the image. These papers have focused on issues relating to correlations and statistical independence in the firing rates of neurons. Much work has involved neural networks and computational models of visual areas. A second line of research has focused on metabolic efficiency. Several influential papers have investigated the metabolic costs of generating spikes, and others have argued that constraints that minimize wiring are important for explaining known neuronal properties. In this chapter, a third form of efficiency is proposed, called learning efficiency. An important consideration for any sensory system is the challenge of learning about relative probabilities of events in the world from a handful of samples.

To the extent that the visual system is optimally efficient at carrying out its multitude of tasks, these efficiency rules are likely to each contribute significantly to a description of why the visual system is designed as it is. Visual systems may approach optimality (in the engineering sense of the word) across each of these three dimensions but as will be argued, one of these dimensions—learning efficiency—is one that engineers rarely consider. These dimensions are not necessarily orthogonal to one another, nor do they currently all have well-defined units of measurement. These dimensions are mere sketches of the terrain over which the human visual system appears to have been optimized through evolution. Furthermore, these dimensions can find application in other modalities as well, though here the focus is on the visual system. Together, these efficiency dimensions may also suggest ways to design efficient artificial visual systems.

1.2 Representational Efficiency

Marcelja (1980) was the first to propose that neurons found in the primary visual cortex (V1) show a number of similarities to the mathematical functions described by Gabor's (1946) theory of communication. Later studies (Field and Tolhurst, 1986; Jones and Palmer, 1987) confirmed that Marcelja's Gaussian-modulated sinusoid model (i.e. the Gabor function model) provides a good first order approximation to the receptive field properties of these neurons. Some of the first computational models of the early visual system (e.g., Watson, 1983; Daugman, 1985) demonstrated how an array of neurons with these properties can represent a visual image. However, this work left unresolved the important question of *why* such a solution might evolve.

One approach suggests that an understanding of the why question requires consideration of the environment in which the system functions. This approach suggests that theories should be guided by an understanding of the statistics of natural scenes. Early television researchers, and later Field (1987) and Burton and Moorhead (1987), found that the Fourier spatial frequency spectra of natural scenes typically fall off as $1/f^k$ where f is spatial frequency and k is approximately 1.2 (Tolhurst *et al.*, 1992; Field 1993). See Figure 1.2 for an example of scenes that have been filtered to show an array of k values. This regularity expresses the same redundancy as the autocorrelation function of scenes, which measures how similar neighboring points are in terms of luminance. The $1/f$ structure follows from two properties: (1.) Neighboring points are correlated, and (2.) the images are roughly scale invariant (Field, 1987). A number of studies have shown that there are a wide variety of forms of redundancy found in natural scenes that go beyond the pairwise correlations and spatial frequency spectra, as described below. However, pairwise correlations have been the source of a number of theories regarding efficient coding and many researchers imply that these are the most relevant statistics, so a discussion of such correlations follows.

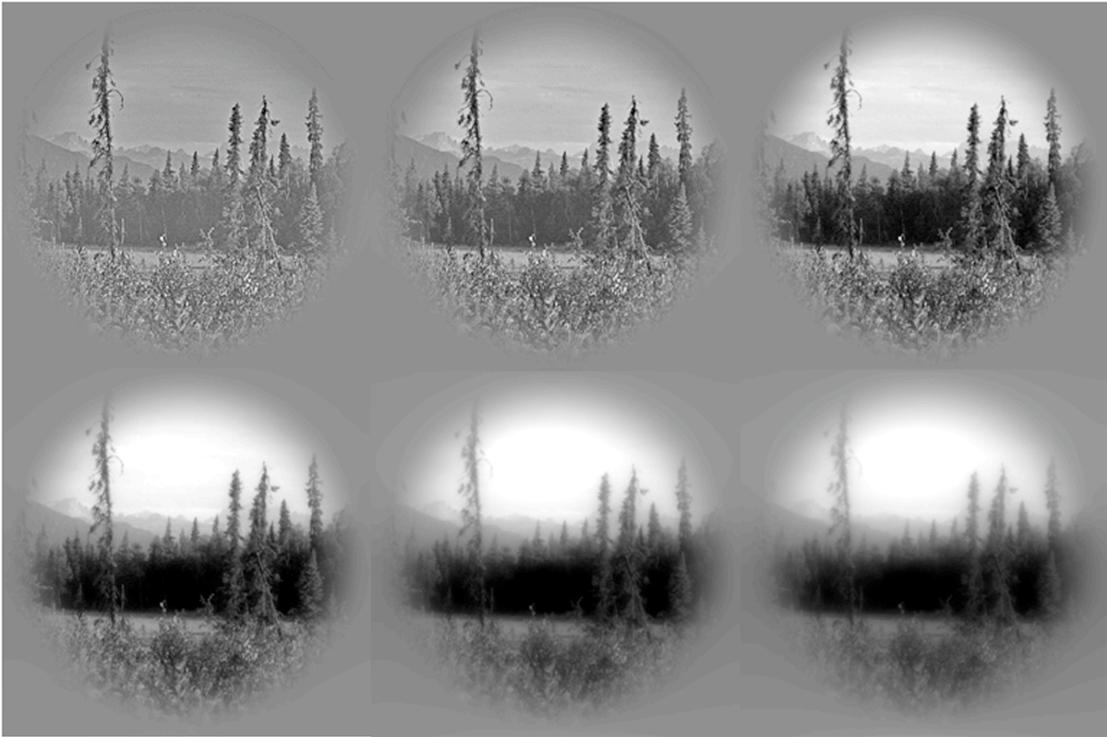


Figure 1.2. Natural scenes adjusted to show amplitude spectrum slopes decreasing from -0.5 to -1.75 by steps of $.25$ from left to right and bottom to top.

Note also that Field and Brady (1997) showed that a fractal edge with the right scaling (i.e., an edge with a length that increases in proportion to decreasing scale) can also produce $1/f$ structure in the 2-D amplitude spectrum. See Figure 4.1 for an example.

1.2.1 Correlation and Decorrelation

When the notion of stimulus redundancy is considered, most papers typically refer to the pairwise correlations in the data. Many of the earliest theories of efficient sensory coding (e.g., Attneave, 1954; Barlow, 1961) developed from the notion that if the neural responses to two stimuli are correlated, then an efficient system should strive to represent the data with reduced correlation. This is certainly one important form of redundancy. Chandler and Field (2007) argue that the pairwise correlations (as represented by the power spectra) account for approximately 40% of the total

redundancy in natural scenes (for 8x8 natural scene patches).

In theory, any representation with significant correlations implies that most of the signal lies in a subspace within the larger space of possible representations. By choosing a representation that codes for only that subspace, it is possible to represent the data with significantly reduced dimensionality (e.g., a reduced number of neurons, or a smaller dynamic range of responses). Srinivasan, Laughlin and Dubs (1982) showed that the center-surround structure of fly large monopolar cells is well matched to the correlations in a collection of scenes they considered relevant to the fly. For a given level of noise, they found that a weighted, linear sum over space transmits the greatest amount of information about their collection of natural inputs. Atick and Redlich (1992) continued this line of inquiry arguing that the amplitude spectra of natural scenes are effectively flattened once they reach retinal outputs. The argument is that the roughly linear rise in sensitivity with increasing spatial frequency is inversely related to the $1/f$ -distributed fall-off in natural scene spectra, thus allowing neighboring neurons to be uncorrelated (i.e., have a flat spectrum). Atick and Redlich (1992) argue that this decorrelation is reflected in a flattening of the spatial frequency response of retinal ganglion cells. However, this argument is dependent on appropriate spacing of retinal neurons. Moreover, the current evidence suggests that neighboring neurons are significantly correlated in the presence of natural scenes (Puchalla *et al.*, 2005; Nirenberg *et al.*, 2001).

As argued in Chapter 2, an independent goal of retinal coding is to achieve compression while maintaining equal response magnitude across the array of neurons of different sizes. This approach achieves a form of response spectrum flattening but it is quite independent of whether the neurons are tuned and spaced ideally to achieve decorrelation. This work suggests in order to explain the details of neural response

properties, one must acknowledge that decorrelation in the retina and LGN is far from complete and that efficient coding schemes must consider additional constraints.

1.2.2 Optimal Information Transfer

Another strategy attempts to utilize each level of response magnitude with equal frequency. To this end, Laughlin (1981) compared the distribution of local contrasts in the blowfly environment with contrast responses in large monopolar cells in the creature's eye. The intensity-response function measured for these cells is well matched to a distribution of natural contrasts, such that the range of possible responses approximated the information-theoretic ideal. In other words, contrast responses in the fly are distributed across the range of environmental contrasts in such a way as to minimize the number of "code words" necessary for transmitting information about the entire range of contrasts at a given fidelity. Later studies have shown that adaptation of visual neurons with respect to local luminance (van Hateren and Snippe, 2001; Schwartz and Simoncelli, 2001), orientation (Wainwright, 1999), and contrast (Yu and Lee, 2004) can also be addressed using this maximum information transfer (infomax) approach. A more elaborate extension of infomax models combines feedback of stored predictors from higher cortical areas with typical natural inputs (Mumford, 1994; Rao and Ballard, 1997). These schemes optimize over families of representations that minimize error between the input and a top-down representation. Feedback representation optimization models of this sort effectively produce a system where primary visual areas reduce their response as higher areas provide better descriptions of input "content," in line with recent imaging findings (e.g., Murray *et al.*, 2002). It should be noted that in many infomax models, optimal redundancy reduction (decorrelation) and maximum information transfer strategies are equivalent.

1.2.3 Beyond Correlations: Sparseness and Independence

The pairwise correlations found in natural scenes represent only one form of redundancy. Figure 3 shows two images—a natural scene and noise—with similar $1/f$ structure and therefore similar pairwise correlations. There are a number of ways to describe the differences between these two images. For example, they differ in their phase spectra. But the two images can also be described in terms of differences in their sparse structure. For the $1/f$ noise image, all linear representations produce response distributions that are Gaussian. However, for the natural scene, some projections of the data are non-Gaussian. That is, when the appropriate array of linear filters is used to represent a natural scene, the histogram of activity will be a non-Gaussian histogram.



A.

B.

Figure 1.3. Noise with a spatial frequency amplitude spectrum like that of natural scenes (A.) has the same pairwise correlations as the natural scene (B.) but lacks other statistical regularities of scenes.

As noted by Field (1994), a non-Gaussian histogram implies low entropy in the first-order responses and relatively high entropy in the higher-order relationships between the filters for a linear transform (i.e., more independent). In other words, a

system that produces maximally non-Gaussian histograms produces a representation where the neurons are maximally independent. This is the basic idea behind sparse coding algorithms (Olshausen and Field, 1996) and independent components analysis or ICA (e.g., Bell and Sejnowski, 1997).

In a code with maximal independence, the firing of each neuron provides maximal unique information (i.e., the sharing of information with other neurons has been minimized). If the data consist of an array of relatively rare, sparse events then matching the neurons to those events will produce activity that is sparse. The definitions of sparseness have varied in the literature. In general, "sparse" implies a relatively high probability of no activity across the population, and some proportion of relatively active neurons. In the computational literature, where one is often modeling neurons as linear operators, the kurtosis (the 4th statistical moment) of the response histogram can be used to describe relative sparseness. Other metrics such as the sparseness index (e.g., Rolls and Tovee, 1995) have proved more useful for spike trains.

Field (1987, 1994) demonstrated that arrays of linear neurons with properties like that found in primary visual cortex appear to maximize the sparse response to natural scenes. Olshausen and Field (1996) further demonstrated that a neural network that attempts to represent natural scenes and maximize sparseness will produce an array of neurons with spatial properties like those found in cortical simple cells. That is, a system that is forced to produce a faithful representation of the input using only a handful of neurons (each firing near its maximum response when it is active) gives simple-cell-like receptive fields. This suggests that at least at the level of primary visual cortex, visual system representations show evidence of being efficiently matched to the sparseness of natural scenes. Similar results have been found for spatio-chromatic stimuli (e.g. Hoyer and Hyvärinen, 2000; Taylor *et al.*, 2000;

Wachtler *et al.*, 2001; Lee *et al.*, 2002; Caywood, Willmore, Tolhurst, 2004) and spatio-temporal patterns (e.g., van Hateren and Ruderman, 1998; Olshausen, 2003b).

It must be emphasized that sparse outputs of these networks result from the sparse structure of the data. It would be relatively simple to produce a nonlinearity that forces a sparse output independent of the input, one that would not be an efficient coding strategy. The networks described above expressly search for the sparse structure that exists in the data. Moreover, recordings of primate visual neurons in response to natural scenes have also been shown to produce a sparse output (e.g., Vinje and Gallant, 2000, 2002, Willmore and Tolhurst, 2001; David, Vinje, and Gallant, 2004). Visual neurons show sparse responses in both early stages of cortical processing like V1 and in higher levels like inferotemporal cortex (Lamme, 1995; Baddeley *et al.*, 1997). Highly sparse firing is a widespread phenomenon in a variety of brain areas and species, including monkey association areas (Abeles *et al.*, 1990), rabbit motor areas (Beloozerova, Sirota and Swadlow, 2003), rat somatosensory areas (Brecht and Sakmann, 2002), rat auditory areas (DeWeese *et al.*, 2003), rat hippocampus (Thompson and Best, 1989) and centers believed to be involved in bird song generation (Hahnloser *et al.*, 2002).

However, the sparse response of these neurons is not direct evidence that these neurons are efficiently representing the environment. All of these neurons are nonlinear and higher-level neurons are very nonlinear. A proof that these codes are efficient would require a clear understanding of how the array of neurons represents the input image. We are not at this level of understanding—and we may never be. Most visual neurons beyond the retina can be modulated by higher levels of processing. Although a number of papers have implied that a decent theory of primary visual cortex is close at hand, Olshausen and Field (2005) have argued that we are still a long way from such a theory. Part of the argument comes from the work of the

Gallant lab (e.g., David *et al.* 2004). This work has made an effort to measure the responses of primate visual neurons to a set of natural stimuli and to predict from these responses how each neuron would respond to an arbitrary natural scene. They have found that even in V1, typically less than 50% of the response variance to new stimuli can be predicted. In higher levels of the visual system, that prediction accuracy is further reduced. In addition, as noted in Olshausen and Field (2005), recording studies select only a small portion of the neurons in any given area and researchers typically record only from large pyramidal neurons that produce large spikes. Other neuron types (e.g., granular neurons that produce smaller spikes) are often not surveyed in these studies, which has led to the interesting suggestion that there exists a neural "dark matter" problem (Shoham *et al.*, 2006).

1.2.4 Optimality with Nonlinear Systems

Despite the difficulties in providing a clear metric for defining efficiency in nonlinear systems, a number of efforts have been made to generalize efficiency arguments to account for the known nonlinear properties of visual neurons (e.g., Field, 1993; Schwartz and Simoncelli, 2001; Wainwright *et al.* 2002; Zetsche and Röhrbein, 2001; Prenger *et al.*, 2003). One particularly fruitful approach involves a technique called slow feature analysis (Wiskott and Sejnowski, 2002; Berkes and Wiskott, 2005). According to this model, a stimulus moving across an image results in quick changes to the neurons responding to the input (for a linear array of neurons). If one considers the histogram of neural activity integrated over some time period, one would find that the activity is less sparse than in a single time frame. However, the changes in images are not random and are often caused by objects or backgrounds showing consistent movement. Slow feature analysis attempts to take into account this redundancy in the movements of features. The technique attempts to find nonlinear

solutions that are capable of describing the moving images with a relatively consistent set of neurons, despite the changes that are occurring. Slow feature analysis has been demonstrated to produce nonlinear behavior like that shown with complex cells (Berkes and Wiskott, 2005). It has been argued that a linear version of slow feature analysis can be mathematically equivalent to certain forms of spatiotemporal ICA (Blaschke *et al.*, 2006). Given that object identities are relatively invariant over time, there is hope that these lines of work will eventually be capable of producing object-level representations like those found in inferotemporal cortex. However, we have not yet reached that point.

It should also be noted that there exist a variety of techniques that come under the general heading of nonlinear ICA (e.g., Schwartz and Simoncelli, 2001; Kayser *et al.*, 2003; Malo and Gutierrez, 2006). This area of research is too vast to be reviewed here. Many of these papers have generated some of the nonlinear properties of visual neurons. However, since we do not yet have a complete model of the nonlinearities in these neurons, we cannot yet argue that such techniques are capable of accounting for the full array of neural properties in any region of visual cortex.

Without a clear account of how information is represented in any given area, we are left without any kind of proof that the information is represented efficiently. Linear models that optimize sparseness and independence do produce simulated neurons with many properties like those found in V1. This certainly supports the notion that V1 is directed towards an efficient code for natural scenes. But as will be argued in the next sections, there are both metabolic and learning costs for high sparseness and therefore a more complete understanding of efficient coding requires examining factors beyond those considered in proposals of representational efficiency. Both the design of artificial visual system and the development of algorithms for representing natural scenes will likely benefit from an understanding of a larger range of evolutionary

constraints, like those described below. These additional biological constraints require a broadening of current notions of efficient design beyond representational efficiency. We may someday find that the best artificial systems look remarkably like systems one finds in biological systems, and it is likely that a full account of biological systems will require an understanding of these biological requirements.

1.3 Metabolic efficiency

Consideration of representational efficiency as discussed above leads to important insights into the design of visual systems. It provides a number of metrics of efficiency that are generally independent of the limitations of energy or of neural hardware. This section attempts to explain the properties of neurons from a consideration of metabolic constraints. Two broad hypotheses of efficient neural design: spike efficiency and neural wiring optimization are discussed.

1.3.1 Spike Efficiency

Any information processing strategy in the nervous system will incur metabolic costs. If one accepts that information is primarily transmitted through the use of spikes, then a relevant question is whether one of the constraints on information processing is the number of spikes. Two studies have made detailed efforts to estimate the metabolic cost of spikes and both have come to the conclusion that the high cost of spikes indeed results in an important constraint on neural processing. Attewell and Laughlin (2001) and Lennie (2003) both find that the total available metabolic resources and the cost of a spike limit the firing rate to less than 1 Hz—and probably less than 0.2 Hz. As noted by Olshausen and Field (2005) most studies have found firing rates significantly higher suggesting that the cells that are typically recorded have unusually high firing rates. Levy and Baxter (1996) argue that when the cost of

spiking is considered, maximum information transfer is attained when only 2-16% of the neurons are firing. Lennie (2003) estimates that the limited resources imply that at any given moment, only 1/50th of the population of cortical neurons will show high firing rates. Thus, in order to save energy individually and across populations, visual neurons must adopt highly sparse patterns of firing. Therefore, one can argue that from both a representational point of view and a metabolic point of view, sparse firing is efficient. However, both the goals of the system and its modifiable elements also play a role in determining the extent to which sparse firing can be considered efficient. For if the goal were simply to use as few metabolic resources as possible, neurons could, for instance, simply be tuned to have higher thresholds for contrast and thus lower average spike rates. Measures such as the Treves-Rolls metric (Treves *et al.*, 1999) are useful for the purpose of distinguishing low spike rates from sparse firing. In addition, as Lennie (2003) argues, the fact that neocortical neurons often show bursting activity (spiking at 100 Hz for short periods) means the cost of any particular spike is on average quite high. The properties of the system's modifiable elements (i.e., the spike rate of the neuron) thus contributes to why a sparse coding approach that employs coding units with such properties can be considered efficient.

A related debate centers on the optimal histogram of activity for a real neuron that can only produce a positive firing rate. If the goal of a neuron is to maximize the total information rate while minimizing the mean activity, the most efficient distribution is the exponential distribution (Levy and Baxter, 1996; Rieke *et al.*, 1997). There is certainly a similarity between the histograms of activity of visual neurons (in response to natural signals) and the exponential distribution (Baddeley *et al.*, 1997). However, on close inspection, the exponential model does not appear to fit. For example, Treves *et al.* (1999) found that for neurons in inferotemporal cortex, the exponential model could be rejected in 84% of cases. The lack of any particular successful model led

these authors to conclude that there was "no special optimization principle or purpose to the firing distributions found."

Moreover, there is a question of whether the goal should be to optimize the information rate. Consider a case where there are n possible causes or features in an image and there exist n neurons available to represent those features. One could argue from a representational viewpoint that one should match the neurons to the features. If each feature had a particular response probability, then matching that feature would produce a histogram of activity that matched the histogram of the feature, not one that necessarily matched the histogram corresponding to the optimal compression algorithm. Certainly one would not expect that there is a match between the number of features and the number of neurons. But this example is meant to show that if we expect neurons to provide an explicit representation of their environment, part of the impetus for that explicit representation may be to match the probability distributions of the environment—not simply to maximize information rates.

1.3.2 Minimum Wiring

In addition to metabolic constraints, any neural architecture will be dependent on the anatomical "wiring" available. This so-called wiring optimization principle (see, e.g., Koulakov and Chklovskii, 2001) dates back to Ramon y Cajal (see Laughlin and Sejnowski, 2003). This type of metabolic efficiency is typically measured by comparing the wiring volume and distribution of real neurons with calculations of the minimum volume needed to connect model cells and cortical areas to one another. Mitchison (1991) has argued that stripe patterns in cortex (e.g., ocular dominance columns) are efficient at minimizing the volume of dendritic wiring needed by areas receiving input from two separate sources, compared to alternative arrangements. Durbin and Mitchison (1990) found that a highly reduced model of cortical wiring

arrangement matched real wiring in certain ways and minimized a model of wiring cost. Along the same lines, Koulakov and Chklovskii (2001) found that pinwheel patterns of orientation preference in cortex allow efficient wiring. In a related study, neural wiring density and local signal delay have similarly reached an optimal compromise (Chklovskii *et al.* 2002). Together, these studies suggest that intracortical wiring optimization can help explain why cortical maps have the topography they do (Chklovskii and Koulakov, 2004). Cherniak *et al.* (2004) and others have shown that minimal wiring may also help explain why brain areas are placed and interconnected as they are. In an early study using a reduced anatomical model of the *C. Elegans* ganglia, Cherniak (1994) employed a connection-optimizing algorithm to search for connectivities that minimized total wiring. The actual wiring of the ganglia matched the optimal arrangement from a family of possible orderings.

Some have argued that brain evolution and development impose invariant connectivity ratios, which may tend to enforce optimal wiring. For example, Changizi and Shimojo (2005) suggest that across phyla the average number of synapses per neuron scales with the number of neurons per physiologically defined area, and that the number of connections among areas (per area) scales with the number of areas. However, this approach has several limitations. It should be noted that the number of ways that an efficient wiring system could strive to optimize the cost of adding cell volume, increasing cell metabolism, delaying and attenuating the signal and making projections during development is essentially unlimited. But when considered in concert with representational constraints, minimum wiring arguments can provide insights about efficient receptive field design (e.g., Vincent and Baddeley, 2003). I will return to the notion of efficient wiring in the retina in the next chapter.

1.4 Learning Efficiency

In addition to the fact that neural systems, unlike human-engineered ones, are difficult to assess in terms of their goals and costs, they also differ from traditionally engineered systems because they are produced through development and learning rather than through carefully planned assembly. Indeed part of the reason why the level of efficiency of a brain is hard to measure is that it becomes specialized to a variety of tasks—often concurrently—during development, the combination of which helps the adult creature reproduce. Therefore, in order to explain why the visual system of the adult has the structure it does given a proposal of efficient processing of natural scenes, one must consider how visual system representations may depend on efficient learning and development. This efficiency dimension in particular is an example of somewhere our intuition about the brain seeking desperately to conserve energy can lead us astray. Indeed, learning efficiency could be at odds with metabolic cost optimizations since the proportion of metabolic energy consumption in infant human brains (which are most subject to this constraint) is roughly three times that for adults (Hofman, 1983).

In discussions of efficiency in sensory systems, many studies have used various types of neural networks and other learning algorithms to generate early sensory processes. There are certain basic problems that any learning algorithm must overcome. One of these issues—invariance—is examined followed by evidence of contributions to learning efficiency from innate, purely learned and hybrid strategies, which span the well-known nature/nurture debate.

1.4.1 Sparseness and Invariance

Certain constraints are imposed on a system that must learn about its environment from a relatively small number of examples. In essence, efficient learning is dependent

on finding a balance between selectivity of neurons for specific features and invariance across examples whose features vary in irrelevant ways (e.g., lighting). This task is further complicated because each example is likely to be seen just once in a lifetime. A second instance is unlikely to be presented with the same lighting, the same position, the same size and the same orientation as the first. To allow any calculation of the probability of what is likely, it is critical that the system be capable of generalizing to multiple instances of that object.

Sparse coding, in the strict version, does not help with this problem. The typical argument against hierarchical sparse codes is that a system that develops a neuron for every object would require too many object detectors in order to function. This "grandmother cell" hypothesis has been critiqued elsewhere (see Gross, 2002). But it is worth noting that plausible models of object recognition (e.g., Riesenhuber and Poggio, 2000) suggest the need for grandmother cell-like coding for certain tasks (e.g., distinguishing between different faces) and sparse population codes for other tasks (e.g., categorization). That is, object recognition may require a variety of strategies that vary in their degree of sparseness.

It has also been argued that optimizing for sparseness and independence can assist a system in identifying novel inputs and in detecting new relationships (Baum *et al.* 1988; Field, 1994; Olshausen, 2003a). Consider the case where objects are uniquely represented by just one neuron versus the case where the object is identified by the relative activity of 100 neurons. In the single neuron case, to learn that two objects often co-occur requires a relatively simple algorithm (e.g., Hebbian learning). But when identification requires a particular activity profile among many neurons, the learning algorithm would require a far more complex association among the 100 units involved in the representation.

1.4.2 *Overcompleteness*

An important property of visual representations beyond the optic nerve is that they are highly overcomplete. Visual cortex in humans contains on the order of 1000 times more neurons than the two optic nerves (Barlow, 2001). In macaque V1 alone, there are approximately 50 times more output fibers than input fibers. These overcomplete codes must involve significant redundancy. Overcompleteness has been suggested as an optimally efficient way "to *model* the redundancy in images, not necessarily to reduce it" (Barlow, 2001). As Riesenhuber and Poggio (2000) point out, overcompleteness allows any particular signal to be represented with a higher degree of sparseness than is possible with complete codes, a property which is useful for generalization during learning, as described above. In order to achieve efficient learning, one must consider solutions that may not be optimally efficient from a strict representational efficiency point of view. That is, solutions that are representationally efficient (e.g., maximum information transfer arguments) may not help explain how the system achieves efficient learning.

1.4.3 *"Hard-Coded" Efficiency*

Though every part of the visual system undergoes developmental change, some properties of the adult cells are not thought to be influenced by the external environment. Maloney (1986) found that the distribution of natural chromaticities could be efficiently and accurately coded using just three wavelength vectors, and that the three cone spectral sensitivities in particular may optimally achieve color constancy based on typical surface spectral reflectances in the natural world (see also Buchsbaum and Gottschalk, 1983). Others (e.g., Osorio and Vorobyev, 1996) have proposed that the distribution of spectral sensitivities in primates is optimally weighted

so as to separate nutritious fruit from surrounding foliage, a genetic adaptation that could have aided propagation of trichromatic primates.

Moreover, at the level of the retina, there is a deep conservation of "virtually all functional and structural features of importance" and of the developmental program in the vertebrate retina (Finlay *et al.*, 2005). This may indicate that visual system development is optimized to allow successful elaboration, robustness and specialization (what is referred to as "evolvability") for a range of species that live in a variety of habitats using a single developmental and organizational scheme (Finlay *et al.*, 2005). It can therefore be seen as an efficiency which is innate, but which is connected to the types of variations present in a changing visual environment.

The development of classical receptive field organization in cortex does not appear to require visual input, and relatively little spatial refinement is needed to achieve adult level acuity. One area of debate centers on the question of the necessity of visual experience in order to achieve adult receptive field structures, which presumably contribute to efficient representations in the brain. Recent studies (e.g., Carrasco *et al.*, 2005) have suggested the prevailing view—which proposes that animals dark-reared during the critical period of development are left with unrefined receptive fields because visual experience had no chance to "prune" dendritic arbors—could be in need of reevaluation.

1.4.4 Efficient Learning from the Environment

In higher levels of cortex, learning leads to a marked decrease in mean response in prefrontal and inferotemporal cortex, which has been linked to the detection of novel stimuli. In particular, learning based on natural stimulus matching in monkeys led to systematically lower mean responses to learned natural stimuli in prefrontal cortex, and these responses did not vary with the addition of noise (Rainer and Miller, 2000;

Ranganath and Rainer, 2003). At lower levels of cortex, learning has a minor effect on classical response properties like orientation tuning and receptive field size in V1 cells (Ghose *et al.* 2002). Learning also has a relatively small effect on responses in V4 to noise-degraded natural stimuli (Rainer *et al.* 2004). Therefore, efficient learning strategies may be an important principle especially in higher levels of visual cortex.

1.4.5 Hybrid Strategies: Efficient Innate Learning

Finally, there is evidence that spontaneous activity in the visual system in a period before the creature's eyes open may help refine the organization of the visual system. There is good evidence that suppressing spontaneous activity in the retina (so-called retinal waves) can affect the refinement of retinal projections to the LGN (see, e.g., Wong, 1999; Butts, 2002). As some have noted, spontaneous, patterned activity is observed in cortex, hippocampus, thalamus, retina and spinal cord in a variety of creatures (Butts *et al.*, 1999). Simple programs of the sort that could theoretically produce spontaneous, patterned activity could well be coded into genes, and the "running" of these programs could produce the necessary statistical properties that a developing visual system needs to operate at a basic level once the eyes open (Albert, Schnabel and Field, 2008). This notion is referred to here as innate learning, since it employs elements of both learning and innateness. These programmed learning strategies are efficient in the sense that in the absence of external stimuli, they require far fewer genetic instructions in order to develop proper response properties compared to full genetic specificity.

Innate learning in the form of spontaneous retinal activity is not required for the initial formation of ocular dominance columns (Crowley and Katz, 2000) or of orientation selectivity. But there is strong evidence that retinotopic maps and ocular dominance patterns fail to refine properly in animals raised in altered visual

environments (e.g., strobe environments, binocular deprivation, environments with altered spatial statistics).

In short, learning efficiency may require solutions that would not be predicted by representational efficiency arguments or metabolic efficiency arguments alone. The difficulty of learning invariant properties of objects based on a handful of presentations is one reason why this is so. The degree to which the visual system is efficiently engineered to learn is dependent on innate properties, on learned associations, and on patterns of programmed spontaneous activity, the combination of which may approach "optimal" efficiency in concert with metabolic and representational constraints.

Visual systems have been argued to be operating near theoretical limits of optimality—not simply parsimoniously—for certain tasks. But given the profusion of ways that brain structure, cell development and function, and neural representation are believed to be optimal in some way, it is argued that visual system design represents a compromise among these many demands for efficiency. This chapter has attempted to elucidate a variety of ways in which visual systems can strive for efficiency. No single quantity (e.g., energy consumption) applies strict limits to brain design (except in the absolute sense), nor is the system optimized to perform a single, "optimizable" task. There is much to be learned from investigating the contribution from a variety of constraints on efficiency.

In the next chapter, the notions of representational efficiency and metabolic efficiency as defined here are applied in a novel way to spatial processing in the retina. The results of this application are then used to inform an expanded view of efficient coding, one that is relevant to the processing of human-made visual art. As will be argued, the notion of efficient coding of natural scenes has important implications in the study of statistical regularities in art. Models of representationally efficient codes

constructed according to framework developed herein—particularly those related to retinal ganglion cell coding—offer a compelling explanation of why art from around the world and throughout recent history displays regularity in its spatial statistics.

CHAPTER 2

CAN THE THEORY OF “WHITENING” EXPLAIN THE CENTER-SURROUND PROPERTIES OF RETINAL GANGLION CELL RECEPTIVE FIELDS?

2.1 Background

The previous chapter outlined the range of research on efficient neural design and it described the foundations of efficiency in the visual system. Here the focus is directed to an important question in efficient neural coding in the early visual system, namely, What is the goal of center-surround receptive field organization in retinal ganglion cells? The answer to this question will have implications for notions of efficient coding in other parts of the brain, and for the study of statistical regularities in art, as described in proceeding chapters.

Across mammalian species one finds that the layout, development, and structure of the retina is remarkably well conserved (Finlay, de Lima Silveira, and Reichenbach, 2004). Center-surround antagonism in particular is found in some form in the early visual systems of all vertebrates and invertebrates (e.g., Land, 1985). Although there have been numerous proposals regarding what is achieved by retinal processing—ranging from “edge enhancement” (Balboa and Grzywacz, 2000; Ratliff, 1965) to decorrelation—here it is argued that current models of retinal ganglion cells are insufficient to account for basic aspects of information processing in the retina. It is also important to emphasize that the account provided here is in no way complete. The varieties of nonlinearities and cell types will certainly require a much richer model. The emphasis, however, is on the basic center-surround organization of retinal ganglion cells.

2.1.1 “Whitening” and Decorrelation

Natural scenes typically have strong spatial pairwise correlations which can be expressed as a spatial frequency amplitude spectrum that falls as $1/\text{frequency}$ —or as a power spectrum that falls as $1/f^2$ (Burton and Moorhead, 1987; Field, 1987). The prevailing view of retinal coding was shaped by Srinivasan, Laughlin, and Dubs (1982), and by Atick and Redlich (1992), whose decorrelation hypothesis focuses on the relation between the ganglion cell tuning curves and the spectra of natural scenes. Srinivasan *et al.* (1982) showed that a group of detectors sampling over space will transmit the greatest amount of information given the presence of noise by taking a weighted linear sum over the spatial arrangement of a group of detectors. Atick and Redlich (1992) extended this line of thinking and proposed that the goal of retinal coding is to produce a decorrelated output in response to natural scenes. They noted that for a range of spatial frequencies, the tuning curve for ganglion cells increases with spatial frequency. Since natural scenes’ amplitude spectra fall with increasing spatial frequency, the multiplication of these two spectra (which corresponds to a spatial convolution) will result in a flat spectrum over this range of frequencies. Flattening of the spectrum is sometimes called “whitening” and it can result in neurons with decorrelated activity under appropriate sampling conditions. However, two quite separate ideas of whitening have been proposed. Both address the ways in which the early visual system handles the $1/f$ amplitude spectrum of natural scenes but each has different requirements and each achieves different objectives:

- The first theory of whitening, which is called here the *response equalization hypothesis*, was proposed by Field (1987) for cortical neurons, and extended by Brady and Field (1995, 2000) and Field and Brady (1997). In this account, the goal is to produce a representation where each neuron has roughly the same average

activity in the presence of natural scenes. Neurons tuned to high frequencies would need increased response gain to produce the same response as low frequency neurons.

- As described above, the *decorrelation hypothesis* of Atick and Redlich (1992) argues that the relationship between the spectrum of each individual ganglion cell and the $1/f$ spectrum of the input results in decorrelated responses. This decorrelation depends on both the relative spectra and the sampling density of neurons.

These two models are not incompatible with each other. Indeed, both can be independently correct or incorrect. The decorrelation hypothesis is appealing because it predicts spatial redundancy reduction at the retinal output. With appropriate retinal sampling, neighboring neurons will have no pairwise correlations in space. In the case of retinal neurons, the center-surround structure of the filters results in bandpass tuning curves for which a portion of the curve rises with frequency. As Atick and Redlich note, the increase in sensitivity with frequency has one important cost: it magnifies the noise at high frequencies. They provide a convincing argument that the reduction in sensitivity at the higher spatial frequencies, especially under low light conditions (i.e., high noise) provides an efficient strategy for coding natural scenes. This line of argument requires two important features: (1) the tuning curves must have the appropriate shape, and (2) the tiling of neurons must be appropriate. For the decorrelation model, the peak of the tuning curve determines the sampling distance that is required to achieve decorrelation.¹ If the response (i.e., the tuning curve

¹ If the tuning curve of the neuron increases with frequency out to some peak frequency P and falls off sufficiently fast past this point, then sampling at a frequency of $2 \cdot P$ will produce uncorrelated responses in the presence of an image with a $1/f^2$ power spectrum. Sampling at higher frequencies ($>2 \cdot P$) and therefore at smaller distances will result in correlated firing.

multiplied with the image spectrum) is not flat, or if the sampling rate of the ganglion cell mosaic is too low, neighboring neurons will be correlated. Multineuron recording studies testing the independence of ganglion cell responses find that nearby ganglion cells of similar functional classes have significantly correlated firing patterns across species (Arnett, 1978; Arnett and Sparker, 1981; DeVries, 1999; Johnsen and Levine, 1983; Mastronarde, 1989; Meister, 1996; Meister, Lagnado, and Baylor, 1995). There is also evidence that in development correlated firing is important for retinal neurons to innervate correctly: synchronized firing has been proposed as a mechanism that helps coordinate the proper development of neural wiring (see Wong, 2000). Therefore, if one wishes to argue that the primary goal of retinal coding is to produce a representation with uncorrelated responses, one must consider the evidence that the retina has not been fully successful.

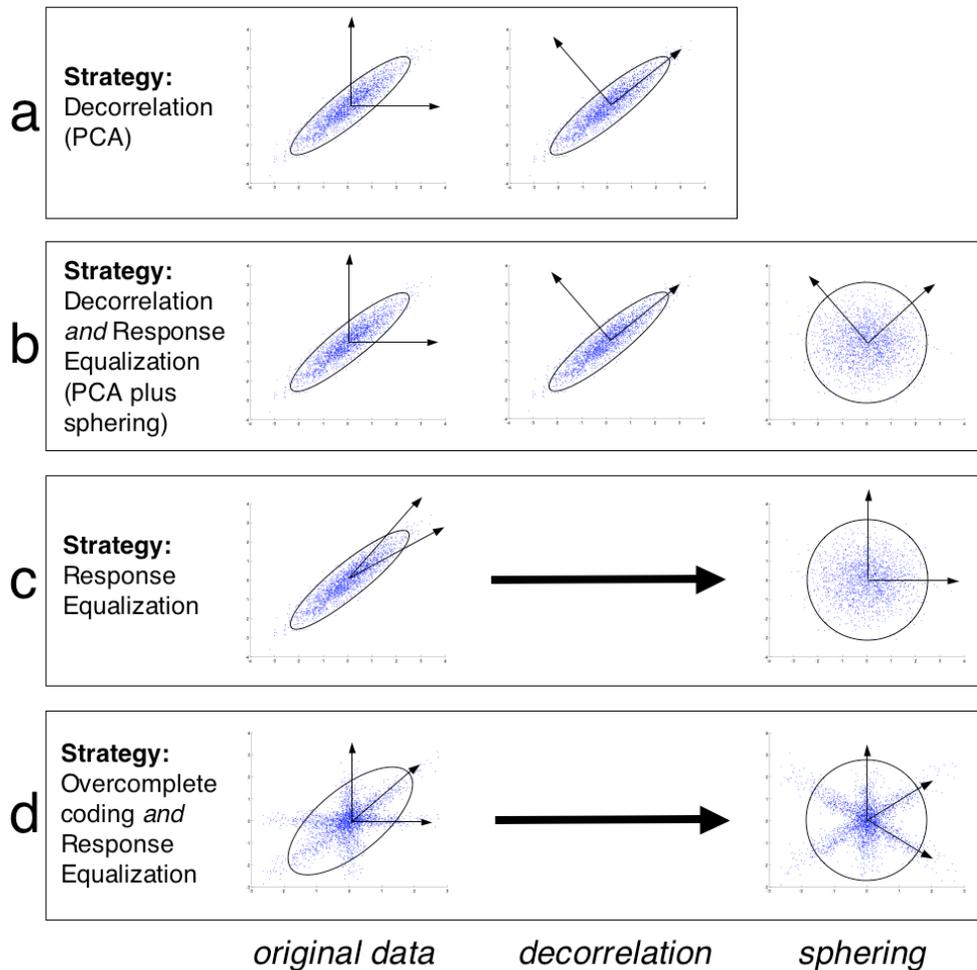


Figure. 2.1. Vector space representations of efficient coding strategies. (A) Shows a two-dimensional PCA transform on Gaussian data. The axes are rotated so that they are aligned with the principal components of the data. In (B), the principal component vectors are normalized or “sphered” (B, right panel) such that the variance along each of the basis vectors is normalized. (C) shows an example of a response equalization coding strategy without a change in vector length. The two vectors shown in the untransformed space (C, left panel) are not orthogonal. The correct choice of vectors results in response equalization. The outcomes of the transforms in both (B and C) are part of a family of rotations in the sphered space. The data in (D) lie in a two-dimensional space but in order to discover the six-pointed star-shape of the data, three non-orthogonal basis vectors are required. The superimposed ellipse in (D) has its major and minor axes aligned with the two principal axes of these data. An overcomplete coding strategy that includes response equalization aligns the representation’s basis vectors with the causes of the data (D, right panel) but these vectors are not uncorrelated (although they are sparse), even in the sphered space.

It can be helpful to think of the two hypotheses of retinal coding—decorrelation and response equalization—in terms of vector spaces. First consider the case of two-dimensional Gaussian data with a strong correlation between two orthogonal vectors (Figure 2.1), for example, two pixels. One method for generating a decorrelated representation is to perform principal components analysis (PCA) on the input data. This method produces an orthogonal vector space whose axes are aligned with the directions along which Gaussian data have the highest variance.

2.1.2. Overcomplete Coding and Response Equalization

The vectors generated by PCA will be uncorrelated but their response variance will not be equal. The principal component vectors can be normalized or “sphered” as shown in Figure 2.1B such that the variance along each of the basis vectors is normalized. This normalization—also called response equalization—allows all vectors (or neurons) to respond with the same average magnitude to the family of inputs. This combination of PCA and response equalization is sometimes referred to as “sphering” or “whitening.” However, both terms can be misleading. Sphering is not part of PCA, but for the example shown in Figure 2.1B, sphering produces a representation whose variance is normalized with respect to the basis vectors (that is, response equalized), thus creating a univariate Gaussian distribution. It should be noted that this process creates a sphere only when one is given Gaussian data. If the data are not Gaussian (as shown in Figure 2.1D), the sphering will result in both decorrelation and response equalization, but there will remain higher-order statistical dependencies. Figure 2.1C demonstrates another way to achieve sphering. By choosing the right set of non-orthogonal axes, one can achieve both decorrelation and response equalization for these data. In the sphered space shown on the right of Figs. 2.1B and C, the two transforms are simply rotations of one another. In Figure 2.1C, the gain of the two

neurons is the same but the response is effectively sphered. It is therefore theoretically possible to sphere data without a gain change. However, ganglion cells with different size receptive fields will not see the same stimulus strength. Because of the $1/f^2$ power spectrum, the neurons tuned to higher spatial frequencies (smaller receptive fields) will see less signal strength. As will be argued, in order to achieve equalization in response to natural scenes the neurons with smaller receptive fields must increase their gain. The response equalization hypothesis suggests that the relative gain of neurons tuned to different frequencies is designed to equalize the response of neurons of different sizes. The hypothesis makes no assumptions about the amount of decorrelation achieved by retinal processing. Consider a case where the causes of the data are non-orthogonal and let us assume that the number of causes is over-complete (i.e., there are more causes of the data than there are dimensions in the representation), as in Figure 2.1D. There exists no linear transform of these data that will result in independent responses. One might choose to align the vectors with the causes of the data as shown in Figure 2.1D. If the causes are not orthogonal, the vector outputs will be correlated. But regardless of the correlations, it may be desirable to perform response equalization: for these data, the different causes have unequal variance so vectors of the same length aligned with these causes will therefore have unequal outputs. However, if the vector lengths are adjusted to counter this difference in variance, response equalization is possible (even though the correlations will remain) as shown in the figure. One is therefore left with the possibility of achieving decorrelation with or without response equalization—and of achieving response equalization with or without decorrelation.

2.1.3. Response Equalization and Vector Length

The response equalization model does not make strong arguments regarding the particular shape of any individual tuning curves and it does not depend on the relative spacing of neurons. This model argues that neurons have overall sensitivity set in such a way that different neurons have the same average response to a natural scene. For neurons tuned to different spatial frequency bands, those tuned to higher spatial frequencies must increase their overall sensitivity to counteract the $1/f$ falloff in amplitude. In applying this argument to ganglion cells, it is proposed that the “integrated sensitivity” of neurons of different sizes across the retina is set in such a way that each neuron will respond approximately equally to natural scenes—despite the large differences in receptive field sizes and regions of pooling. This chapter has focused on spatial properties but a full account would consider temporal tuning as well.

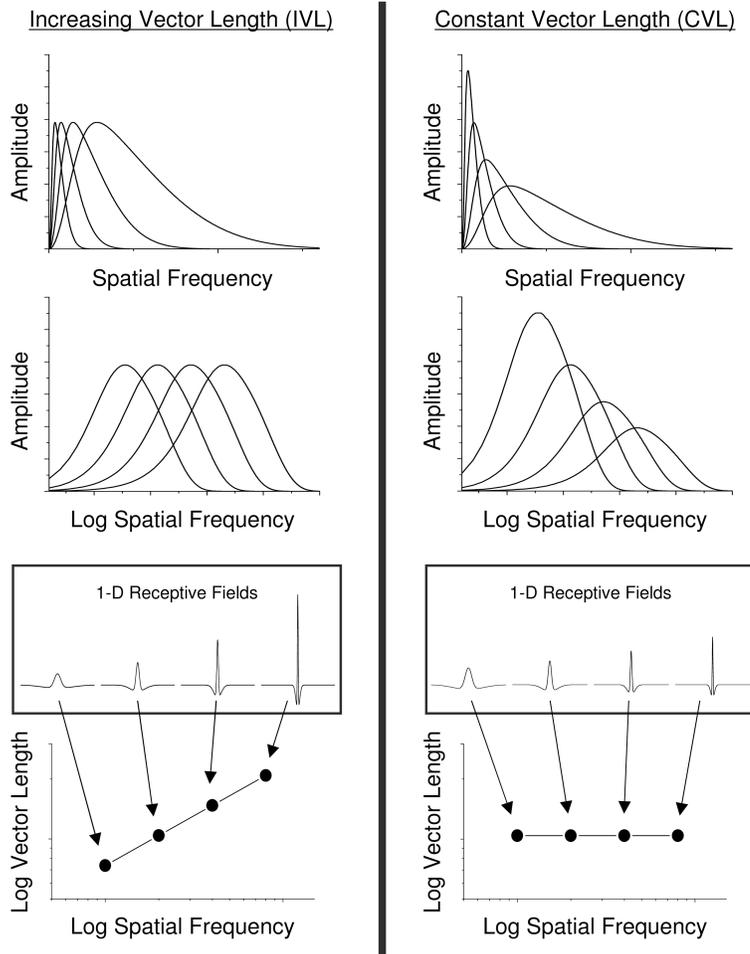


Figure 2.2. Comparison of an increasing vector length (IVL) model of contrast sensitivity, whose vector length increases with spatial frequency (left column), and a constant vector length model (right column), which shows equal vector length across frequencies. Tuning curves of four hypothetical retinal neurons are shown (from top to bottom) as functions of spatial frequency and log spatial frequency; as 1-D receptive field profiles; and in terms of the log of their 1-D vector length. In the IVL model, the peak of each spatial frequency tuning curve is nearly the same, and thus neurons sensitive to high spatial frequencies integrate over proportionally more of the frequency spectrum. This model will show increasing vector length with increasing spatial frequency. The constant vector length model (CVL), on the other hand, predicts decreasing peaks in amplitude for increasing spatial frequencies and approximately equal vector length sensitivities across frequency. A full linear model of the contrast sensitivity function using the vector length metric is being developed. For clarity in this figure, the vector lengths correspond to the 1-D receptive field profiles shown. The 2-D model discussed in the text is a straightforward extension of the 1-D model shown in this figure. In both 1-D and 2-D, the vector length is given by the L_2 norm. The 2-D model used here predicts that vector length increases proportional to spatial frequency (volume under the power spectrum increases in proportion to frequency); in the 1-D model, vector length increases as the square root of spatial frequency.

As an integrated measure of the sensitivity of a neuron, “vector length”(Brady and Field, 1995; Field and Brady, 1997), i.e., the L_2 norm of the neuron’s sensitivity profile, is used.² Figure 1 shows two models for how neural sensitivity profiles, assumed to have constant shape, might scale with the peak spatial frequency of a neuron. On the left, peak sensitivity is independent of peak spatial frequency, so the L_2 norm must increase with peak spatial frequency (“IVL” model). On the right, the L_2 norm is constant with respect to peak spatial frequency (“CVL” model), so the peak sensitivity must decline with peak spatial frequency. In a world with a $1/f^2$ power spectrum, the vector length must increase with frequency in order to achieve response equalization (Brady and Field, 1995).

2.2 Study 1: Calculation of Vector Length Sensitivity of Ganglion Cells

The first study investigates the hypothesis that at the level of the retina, ganglion cells already show evidence of response equalization. The study makes use of the published data of Croner and Kaplan (1995). That previous study is one of the few that provides a measure of the absolute sensitivity of ganglion cell sensitivity for a relatively wide range of receptive field sizes, as well as sufficient information for us to calculate the vector length. This first analysis is therefore a simple reanalysis of their data.

2.2.1 Methods

The Croner and Kaplan (1995) study measured responses of ganglion cells across the retina in anesthetized, paralyzed macaques when presented with gratings of

² The L_2 norm of a vector \mathbf{R} is given by $\sqrt{(R_k)^2}$, where R_k denotes the k th component of \mathbf{R} . Note that the vector length of \mathbf{R} will be the same in any orthonormal basis and it is therefore useful as a unitless, relative measure of sensitivity.

different frequencies. In that study, the tuning functions were Fourier transformed assuming a center-surround phase spectrum, fit to a Difference of Gaussian (DoG) model and the median parameters of those center-surround neurons were published for various cell types and positions. Since linear transforms do not alter the vector length, one can use these data to calculate the vector length as a function of cell size. Absolute sensitivity data was collected across a number of animals and for both M-and P-cells. In the study of the Croner and Kaplan (1995) data, the experimentally determined parameters for the DoG function describing the cells' receptive fields were used. The vector length (L_2 -norm) of the DoG functions for P-cells in the study was calculated. P-cells are the dominant class found in the primate retina, and they have high spatial acuity compared to M-cells.

2.2.2 Results

Vector length sensitivity values are plotted as a function of the log (weighted) mean spatial frequency of each cell (i.e., the weighted mean value of the spatial frequency tuning curve of each cell), shown in Figure 2.3A.³ Parameters for the 84 total P-cells of different sizes represent the median value within bins corresponding roughly to cells of the same eccentricity on the retina. There are five such median values for the parameters that describe the receptive field function. The results of the analysis of the Croner and Kaplan data suggest that vector length is indeed increasing as a function of frequency. What does this vector length sensitivity curve tell us about ganglion cell responses to natural scenes? With a power spectrum that falls as $1/f^2$ (amplitude falls as $1/f$), the response function should be approximately flat, indicating that the response from cells of different sizes to natural scenes will be approximately uniform. Figure 2.3B shows the response of each neuron to a natural scene, computed

³ Qualitatively similar results were obtained using the peak frequency value.

by first multiplying each neuron's tuning curve with a $1/f$ amplitude spectrum and then taking the L_2 norm of that product. This linear model suggests that P-cells perform a significant degree of response equalization.

Although temporal aspects of the neural response are ignored and there is an assumption that the system is linear, these results do imply that P cells have sensitivity that is well matched to the power spectra of natural scenes. If these results hold for all P cells, then it can be predicted that neurons of different sizes distributed across the retina will provide a roughly equal response.⁴ Uniform responses across frequency implies that the cells are maximizing the use of the range of firing rates over which the cell responds, given the regular statistics of the environment.

Note that the data show no clear high-frequency cutoff but at this time, it is not possible to say whether the sensitivity continues to increase out to the highest spatial frequencies to which the neurons respond. It should also be noted that the vector length sensitivity makes a direct prediction regarding how the neurons will respond to noise. Without knowledge of how sensitivity was affected by mean luminance in the Croner and Kaplan (1995) study, one cannot say any more about the noise-reduction properties of primate ganglion cells in this study.

⁴ M-cell data in the Croner and Kaplan study were insufficient for us to draw conclusions about patterns of vector length sensitivity at different spatial frequencies.

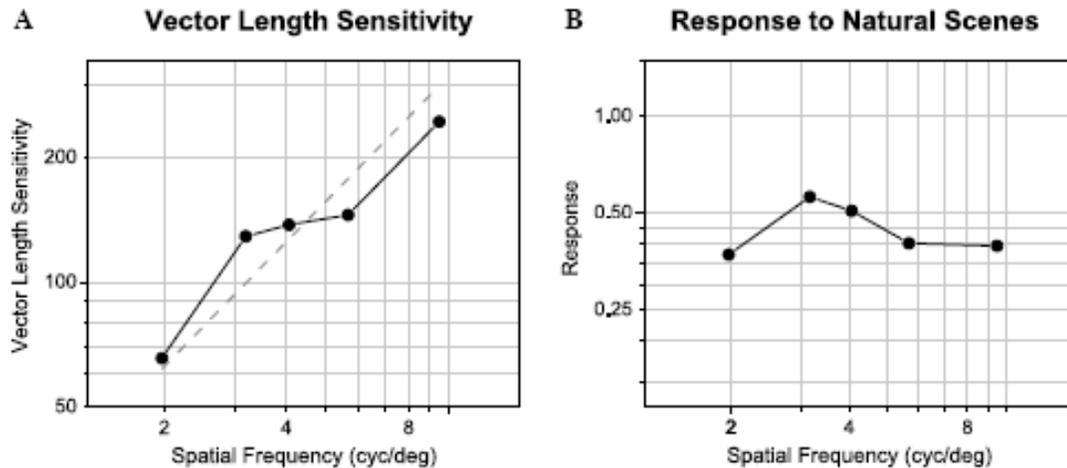


Figure 2.3. (A) Plot of the vector length sensitivity of cells from data by Croner and Kaplan (1995). Each data point in the vector length plot (dots) represents a cell whose receptive field is modeled with the DoG parameters reported by Croner and Kaplan (1995). Vector length sensitivity (unitless) is monotonically increasing proportional to frequency. The x-coordinate of the vector length plot is in units of log mean spatial frequency (cyc/deg) for each cell (see text for definition). (B) Plot of the response magnitude for ganglion cells to a distribution with a spatial frequency power spectrum that falls as $1/f^2$. This $1/f^2$ "input" represents a typical natural scene. Sensitivity is given by vector length (A). Responses show a generally flat shape across spatial frequency. Dotted line in (A) represents a slope of 1 on the log-log plot.

2.3 Study 2: Decorrelation and Sparseness in Model Ganglion Cells

It should be stressed that the response equalization hypothesis provides an account of the overall sensitivity of different neurons but has no implications regarding the spatial correlations between neighboring neurons. As was noted, there is significant evidence that there exist significant correlations between neighboring neurons suggesting that the convolved spectra and/or the spacing is insufficient for producing decorrelated responses.

This study presents a further point. Both the decorrelation hypothesis and the response equalization hypothesis are dependent on the spectra of the neurons' tuning curves and they are not directly dependent on the phase spectra of the neurons. It is

argued that center-surround organization, which depends on the phase spectra, is not directly addressed by either approach.

The focus here is on the question of what function is provided by localized center-surround receptive fields like those of retinal ganglion cells. To explain the oriented receptive fields of neurons in primary visual cortex, it has been argued that the visual system produces a sparse solution that reduces dependencies beyond the second-order correlations (Field, 1987, 1994). Neural networks that attempt to minimize these dependencies among the population of neurons have been found to produce localized, bandpass, oriented receptive fields much like those of simple cells found in V1 (e.g., Bell and Sejnowski, 1997; Olshausen and Field, 1996). Therefore, if the goal of early coding were to produce an efficient or independent solution, one might expect to see a wavelet-like transform similar to V1 in the retina. A wavelet-like transform does not require more neurons than a center-surround system does to achieve a complete representation, so the argument cannot be that more neurons are needed. Here, the relations among center-surround organization, decorrelation and sparseness in model retinal neurons are investigated. The simulation in the following section has two goals: first, it will be used to demonstrate that the decorrelation hypothesis is insufficient to predict center-surround receptive field design. Second, the simulation demonstrates that the ganglion cell produces a more sparse response than other solutions that decorrelate to the same extent.

2.3.1. Methods

For this study, images from van Hateren's database (van Hateren and van der Schaaf, 1998) were randomly selected. Images were then discarded if they did not conform to two criteria: They were required to be devoid of human-created forms and of significant blur. The restriction on blur is the more crucial one: if the camera moved

when the shutter was open, the resulting images were blurry, which introduces uncertainties into the data. The selection process produced a set of 137 stimuli that shows a range of scenes at different scales (images used are listed at <http://tinyurl.com/68mbb>). The mean power spectrum of the images was fit by the function $y = 1/f^k$ where $k = 2.6$. This value of n reflects the fact that the images used are a biased data set within the van Hateren and van der Schaaf (1998) database. However, the relatively uniform ganglion cell response function described in the calculation above is qualitatively the same for $k = 2.0$ and for $k = 2.6$, and the calculations below are not dependent on this fact. Most images in the dataset show grass or forest scenes, some have bodies of water and none has any large vistas. Calibrated images such as these are photometric maps of scenes wherein pixel values correspond linearly to luminance.

A difference-of-Gaussians (DoG) model of retinal ganglion cell receptive fields was used as the basis of the filter kernels (see Figure 2.4). The DoG model is a simplified model that ignores many aspects of ganglion cell function. The radially symmetric DoG function $R(x, y)$ is described by

$$R(x, y) = C_1 \frac{1}{\pi\sigma_1^2} e^{-(x^2+y^2)/\sigma_1^2} - C_2 \frac{1}{\pi\sigma_2^2} e^{-(x^2+y^2)/\sigma_2^2}$$

where C_1 and C_2 are constants that determine the height of the center and surround Gaussians, respectively, and r_1 and r_2 are the variances of the center and surround, respectively (Rodieck, 1965). In this study, $r_1/r_2 = 6.0$ and $C = 20$ and filter kernels were created in a frame of 64 x 64 pixels then centered and zero-padded to make them 1024 x 1024 pixels (the size of the stimuli). Convolutions were performed with phase-aligned DoG filters and with DoG filters whose Fourier frequency components had

been phase-randomized (that is, they were given a norm-preserving, random rotation in the imaginary plane of phase space) before zero-padding. The term “phase-aligned” is used throughout to indicate that the frequency components of the filter are aligned at zero phase before zero-padding. The term “phase-randomized” is used throughout to indicate that Fourier frequency phases were randomized before zero-padding.

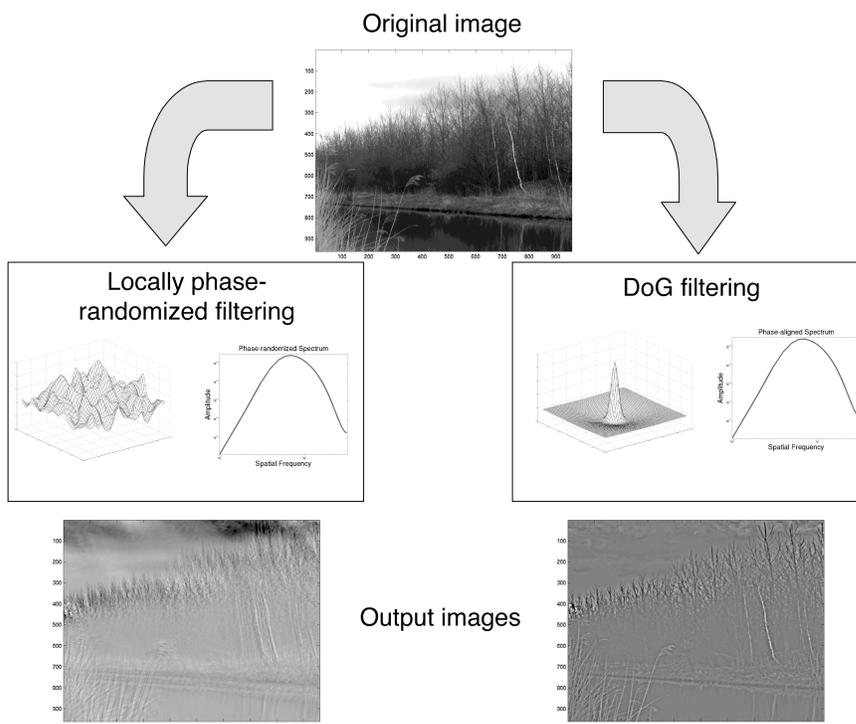


Figure 2.4. Simulation of receptive field filtering of natural scene image. The original linear image (image number 6) was filtered with a phase-aligned difference of Gaussians (DoG) filter and with a locally phase-randomized filter (see text for definitions). The spectra of the two filters (shown on log-log coordinates) are identical. Note that in the output images some image structure is retained in the phase-randomized-filtered image as a result of the localized nature of the phase-randomized filters. One would not expect this to be the case if the filters were the same size as the image. (Filters are 64 x 64 pixels before zero-padding and phase-randomization is done before zero-padding; images are 1024 x 1024 pixels, though the images shown above are 1024 x 1536). The phase-spectrum of the natural scene image was not manipulated before filtering.

By necessity, the power spectra of the phase-aligned and the phase-randomized filters are identical. Six phase-randomized filters were each convolved with the image set, and the mean of these trials was taken as the phase-randomized power spectrum. The 64 pixels at the edges of the images were cropped before spectral analysis in order to remove edge effects (this was necessary for both types of filters because periodic boundary conditions were not used). The convolved power spectra are also identical when the image and the filter kernel are the same size, as found in a separate trial. But in this experiment, because the images are larger than the filters, zero padding is necessary, which leads to differences in the convolved power spectra at low frequencies. However, the mean power spectrum for the set of images convolved with phase-randomized filters falls within one standard deviation of the mean power spectrum of images convolved with phase-aligned filters.

2.3.2. Results

The first result should really be considered a mathematical necessity rather than an experimental finding. By randomizing the phase spectrum of the filter, one may change the phase spectrum of the convolved image, but such randomization can have no effect on the resulting amplitude spectrum. The amplitude spectrum of the convolved image is simply the product of the amplitude spectrum of the image and the amplitude spectrum of the filter. The phase spectrum plays no role.

Since the autocorrelation function is the Fourier transform of the power spectrum, the phase spectrum is also assumed to play no role in determining the correlations (when, that is, we assume stationarity in the phases). Phase-randomized filters achieve the same flattening in the 0.3–3 cycles/degree range and the same high-frequency noise attenuation as do center-surround filters. But the phase-randomized filters do not resemble ganglion cell receptive fields (see Figure 2.4). For filters with a given power

spectrum, each alignment of phases preserves the same amount of information in any convolution with an image.

It is concluded that the center-surround structure does not follow from the constraint that the system simply decorrelates. As with the response equalization hypothesis, there are a wide variety of solutions that achieve equivalent decorrelation, and the center-surround organization is just one example. One must therefore look to other constraints.

2.3.3 Sparseness

The sparseness of the convolved images was calculated using kurtosis as the metric. The ratio of the sparseness of the center-surround-filtered images compared to that of images convolved with phase-randomized filters was on average 3.5 ± 0.40 (mean \pm 95% confidence limits). That is, the mean sparseness of center-surround filtered images is greater than the mean sparseness of phase-randomized-filtered images by a factor of 3.5 ± 0.40 (see Figure 2.5). This value refers to the mean difference in population sparseness, defined as the sparseness across the population of neurons for a given static image. To gauge lifetime sparseness—that is, the sparseness of a neuron’s response through its lifetime as opposed to the sparseness across the population—a total histogram was compiled for all images after filtering with phase-aligned and with phase-randomized filters. In this case, the center-surround filtered images had a sparseness that was 1.4 times greater than that of images convolved with phase-randomized filters. As a control, the same convolutions with the two sets of filters were performed on Gaussian white noise and on white noise whose power spectrum was given by $1/f^2$, both of which gave a kurtosis of 0 for all convolutions.

2.3.4 Compressive Nonlinearities in the Retina

In keeping with the proposal of Srinivasan *et al.* (1982), the same sets of filters were convolved with log-transformed images. The rationale for taking a log of the image before filtering is based on physiological studies of frogs by Norman and Werblin (1974), who showed that photoreceptor sensitivity, when adaptation over time is taken into account, goes roughly as the log of intensity (see also Naka and Rushton, 1966; Baylor, Nunn, and Schnapf, 1987). Moreover, as Field (1987) pointed out, a log transform would recast intensity differences as ratios, a property that could be advantageous for the cell since intensity ratios express contrast.

A log nonlinearity was applied to the images from the previous study then convolved each with the same sets of filters. Sparseness for the phase-aligned DoG filters was higher than the sparseness for the phase-randomized filters in this case by a factor of 1.9 ± 0.17 . The log case is reported in order to show that sparseness is higher for the DoG filters than for the phase-randomized filters when a model of the cone nonlinearity is included. Lifetime response sparseness for log-transformed images was found to be 1.5 times greater for the center-surround filtered images than for those filtered with phase-randomized filters.

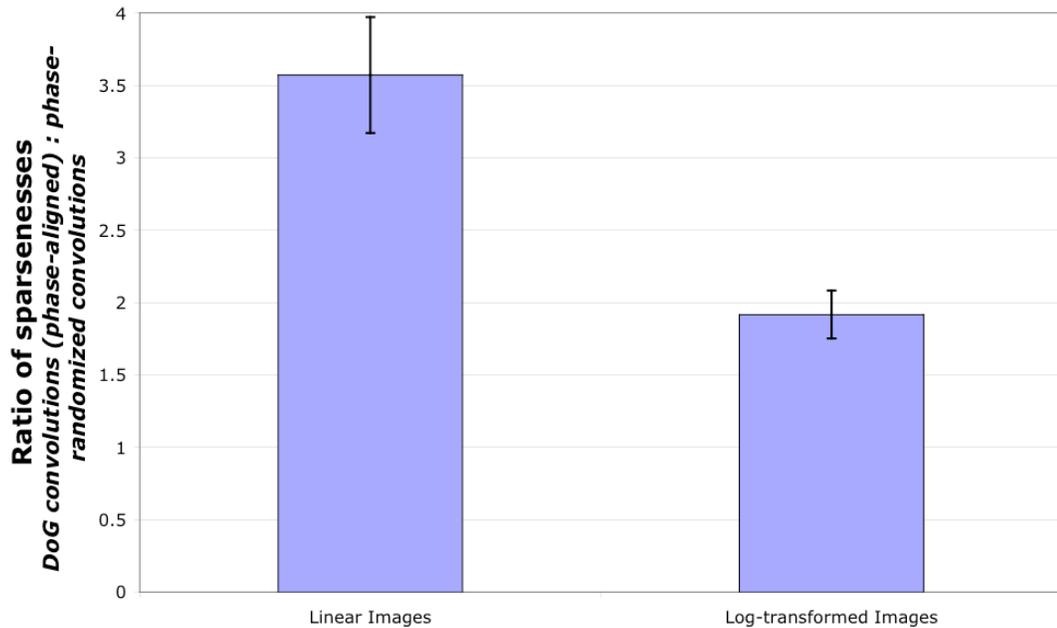


Figure 2.5. Mean ratio across images of population sparseness for DoG convolutions compared to phase-randomized filter convolutions, for linear images (left bar) and for log-transformed images (right bar). For both types of images, DoG convolutions on average show greater sparseness per image. Error bars indicate 95% confidence limits. The mean sparseness ratio is the mean of the ratio of the sparseness for each pair of DoG and phase-randomized image convolutions, in both the linear and log-transformed cases.

It should be noted that the DoG filter and the phase-randomized filter are indistinguishable based solely on the mean response (the first statistical moment) because both filters were designed to have a mean of zero. Nor could they be distinguished based on variance (second moment): Because the filters have the same power spectrum, they will have the same variance. Differences in the skew (third moment) of the filtered images showed no clear pattern, whereas differences in kurtosis (fourth moment) did, as described above.

2.4 Discussion

This chapter investigates several hypotheses regarding why the retina processes information as it does. One prejudice in the past has been to assume that because the retina is one of the earliest major processing units, it “is not expected to have knowledge beyond the simplest aspects of natural scenes” (Atick and Redlich, 1992). This has led many to assume that we have a relatively complete understanding of why the early visual system uses a center-surround receptive field. In this chapter, claims that the early visual system simply “whitens” or decorrelates the input are argued to be insufficient. To account for the center-surround organization, a set of at least three constraints is likely required. Although evidence for full decorrelation among retinal neurons is lacking, decorrelation represents one of the constraints on the shape of the spatial frequency tuning curves and the relative spacing of cells. However, a decorrelation constraint does not account for the relative gain of different receptive field sizes, nor does it account for their center-surround organization. Moreover, the decorrelation hypothesis requires appropriate retinal sampling in order for its predictions to be valid.

Atick and Redlich (1992) make an important point about noise. Because the noise spectrum (e.g., photon noise) is thought to be flat and not declining in the same way as the signal (Pelli, 1981), decorrelation of a $1/f$ spectrum in the high-frequency regime would serve to amplify unwanted noise. This argument is consistent with findings that ganglion cells lose their inhibitory surrounds at low luminance and become low-pass filters. As Atick and Redlich (1992) point out, low-pass filtering increases the signal-to-noise ratio because signal power becomes small at high frequency whereas noise power is constant across frequency. This proposition is important but it is also possible to extend the argument. A full understanding of the underlying noise and vector length sensitivity could account for spatial sensitivity at threshold.

Using the vector length sensitivity measure, sensitivity was found to increase through at least 10 cycles/deg for P-cells in macaques. This result (as shown in Figure 2.3) suggests that neurons with different sizes of receptive fields, from the fovea to the periphery, will respond about equally to a natural scene and maximize the use of the dynamic range available. In a previous study on psychophysical contrast matching it was proposed that this vector length sensitivity increases out to as much as 20 cycles/deg in humans (Brady and Field, 1995).

The vector-length approach to sensitivity may seem to conflict with the standard contrast sensitivity function (CSF) which implies that sensitivity peaks around 4 cycles/deg. However, it is argued here that there is no conflict. A brief comment on this issue should be made here.

2.4.1 Contrast Sensitivity and Vector Length

The contrast sensitivity function measures the psychophysical threshold at which humans (or any species) are able to detect contrast at a given spatial frequency. The CSF is fundamentally a signal-to-noise measure. If one presumes that the noise that limits visual sensitivity is flat, then the vector length sensitivity is a direct measure of the noise magnitude in the system (Field and Brady, 1997). Furthermore, if one assumes that the peak response to gratings is flat out to a range of 20 cycles/deg, and that the linear bandwidth increases with frequency (as shown in the left column of Figure 1), then each neuron will have a constant response magnitude for its optimal grating but it will have a response to noise that increases with increasing frequency. The result is that the system will show a signal to noise level (i.e., the CSF) that decreases with increasing frequency, even though the vector length is increasing with increasing frequency. Here it is proposed that the 4 cycles/deg peak of the psychophysical contrast sensitivity function is the point at which the signal-to-noise

ratio is maximized. The positive slope observed at low spatial frequencies could correspond to a regime that is coded by the lowest spatial frequency channel used by the visual system.

The results shown in Figure 2.3 for P-cells are consistent with this general model. This simple linear model predicts the following:

1. Overall response sensitivity of neurons increases with increasing frequency (out to some limit—in the range of 20 cycles/deg in humans).
2. Response equalization. The response to natural scenes ($1/f^2$ power spectrum) is roughly flat.
3. The contrast sensitivity function will fall at frequencies above and below the peak of the lowest channel.

As noted by Field and Brady (1997), this model also provides an account of why white noise appears to be dominated by high frequencies rather than structure at 4 cycles/deg. Unlike natural scenes or gratings, the model predicts that the response to noise peaks at the highest frequency channel (the neurons with the greatest vector length). If the CSF represented an accurate account of suprathreshold sensitivity, then one would expect that white noise would appear dominated by structure at 4 cycles/deg. A full account of how neural sensitivity relates to the CSF would also need to incorporate the role of the optics and the role of early nonlinearities (Field and Chandler, in preparation). However, the CSF is not incompatible with this simple linear model where integrated sensitivity (vector length) is increasing.

Given that the vector length notion offers a novel explanation of contrast sensitivity, how might one design an experiment to measure vector length sensitivity in neurons directly? If one assumes that the neurons are strictly linear, this is a

relatively simple proposition. However, since neurons saturate, and since they cannot fire negatively, the problem is not straightforward. With nonlinear neurons, there is no guarantee that a neuron's vector length sensitivity with respect to one set of normalized, orthogonal basis set of stimuli will be the same as its sensitivity to another such set. In practice, we must have appropriate controls. Vector length sensitivity could be approximated for neurons that are assumed to be linear by testing a normalized set of Fourier basis gratings and a set of the same gratings where each component has the opposite phase. With respect to the stimulus phase polarity that elicits the greater response, one would then compute the L_2 -norm (i.e., the square root of the sum of the squares) of the resulting receptive field reconstruction in order to measure the vector length sensitivity.

2.4.2 Further Constraints

These results suggest that both the decorrelation constraint and the response equalization constraint remain insufficient to predict the center-surround receptive field organization of ganglion cells. In response to natural scenes, evidence has been presented that the center-surround organization of DoG filters produces a sparse response compared to phase-randomized filters. Since these two classes of filter achieve equal degrees of decorrelation and response equalization, one must consider that an additional constraint is required to account for the center-surround shape. The simulation presented in this chapter suggests that sparseness may be a factor. However, two points must be noted. First, if the only goal to represent the input with maximal sparseness, then the center-surround solution is not the optimal solution. Second, the sparseness may be partly a function of the highly localized nature of the center-surround profiles. As shown by Olshausen and Field (1996), a neural network that optimizes for sparseness and losslessness will settle on a set of oriented, bandpass

filters similar to cortical simple cell receptive fields. If the only additional constraint were sparseness, one would expect to see an oriented wavelet code, which is not the case in the primate retina.

Recent physiological recordings in the primate retina (Berry, Warland, and Meister, 1997) and LGN (Reinagel and Reid, 2000) suggest that random flickering and white noise stimuli can both produce sparse responses in these cells' outputs. This implies some degree of nonlinearity. One would expect sparseness to be lower for white noise than for natural scenes. But at present, no group has directly compared the sparseness of responses to natural scenes with the sparseness of responses to white noise.

The findings in this chapter apply to cells with receptive fields that are well-described by a DoG model (center-surround). Studies of ganglion cell sensitivity that use spike-triggered averages produced for white noise movies (white noise analysis) show that there could be additional microstructure in center-surround receptive fields that is not described by the DoG model (Brown, He, and Masland, 2000).

The fact that significant nonlinearities exist in retinal processing (see Benardete and Kaplan, 1997a; Benardete and Kaplan, 1997b; Kaplan and Benardete, 2001; Shapley and Victor, 1979, 1981; Victor, 1987) could imply that these early nonlinearities are good building blocks for the types of nonlinearities found in the cortical code (end-stopping, cross-orientation inhibition, etc.). Retinal nonlinearities could produce a code that is useful for the kinds of calculations performed in cortex. The focus here has been constrained to a linear model of ganglion cell spatial properties. However, nonlinearities in the early visual system, including temporal and adaptive properties (Hosoya, Baccus, and Meister, 2005), may play an important role in ganglion cell tuning.

The answer to the question of why center-surround organization is highly conserved for receptive fields across species likely requires a broader theory that incorporates noise reduction, nonlinearities, adaptation, and other temporal properties. Dong, among others, (Dong and Atick, 1995a, 1995b; Dong, 2001) has emphasized spatiotemporal decorrelation as an important goal of ganglion cells and of LGN. Linsker (1989) proposed an unsupervised learning algorithm that was optimized with respect to mutual information (equivalently, decorrelation)—this system could produce topographic maps, lateral interactions and Hebb-like modification, though not center-surround receptive fields.

2.4.3 Localization

It is possible that a center-surround arrangement requires a minimum of dendritic wiring given its task (see *Metabolic Efficiency* in Chapter 1). Such arguments have been considered in the context of cortex (e.g., Mitchison, 1991), though such arguments do not address receptive field organization specifically. In the simulation, one effect of phase-randomization was to increase the radial spread of the receptive fields in space, which results in a less sparse response (see Figure 2.4).

Vincent and Baddeley (2003), using a set of simulations, argue that the center-surround operator serves to optimize synaptic efficiency. Presumably, minimizing the dendritic spread will serve to both minimize the total wiring needed in the retina and the number of synapses required to represent the input. From this line of argument, center-surround organization acts to optimize the localization of the receptive field for a given frequency tuning curve. That is, given some constraints on the tuning curve—for example, that it is required to be unimodal—center-surround organization achieves optimal localization.

One possible test of the minimum wiring hypothesis in ganglion cells would involve a neural network architecture that is designed to search for receptive fields that decorrelate and/or sparsify with a minimum of connectivity. Both sparseness/independence and size/efficiency constraints may be inter-related. Both serve useful goals. Although the size constraint may be the more important factor, the additional sparseness/independence should not be ignored.

2.5 Conclusion

The fact that retinal structure and organization are remarkably well-conserved across mammalian species could imply that this organization is a very efficient first step in coding the natural world given the constraints of retinal neurophysiology. Moreover, to the extent that retinal processing is an optimally efficient first step for coding natural scenes, artificial visual systems may benefit from adopting a retina-like strategy as a first step as well. Such a strategy may prove useful as an initial stage in the extraction of features from many classes of natural images. Furthermore, the same types of constraints that contribute to center-surround organization in retinal ganglion cells (sparseness, response equalization, localization, minimal wiring, and other factors) may well explain the center-surround receptive fields in other sensory modalities (such as the tactile system) and the lateral inhibition found in the auditory system.

A minimum of three constraints must be considered to account for the known linear properties—decorrelation, response equalization, and size/sparseness. Although decorrelation plays a role, the evidence does not support the hypothesis that the retina successfully decorrelates. The calculation for P-cells from the primate study of Croner and Kaplan (1995) suggests the sensitivity across neurons of different sizes serves to produce equalized responses in the presence of natural scenes. A full account certainly

must consider the temporal aspects of tuning, and other classes of retinal ganglion cell. However, the results argue that constraints on dendritic wiring and sparseness must also be considered. An account of the retina's linear functional goals would consider all of these factors and perhaps others. This approach has not considered the range of nonlinearities found across different classes of retinal neurons. How many more constraints will be required to provide a full account of retinal processing remains to be seen.

2.5.1 Implications for the Study of Art

Given that retinal ganglion cell sensitivity appears to be efficiently matched to the spatial frequency spectrum of natural scenes (at least in macaque), how might this affect the types of images created by artists for human viewing? That is, could it be that artists in general produce images that match the frequency spectrum of scenes because those images can be efficiently processed by the visual system of the viewer? In order to begin to answer this question, it is first necessary to test whether art images a whole conform to the statistical regularities of natural scenes. Chapter 3 is dedicated to performing this test. Chapter 4 will address whether art from different regions of the globe shows statistical regularities and whether different types of art (landscapes, portraits, abstract works) show characteristic regularities. Chapter 5 offers a consolidated theory of “natural art” based on coding efficiency principles and regularities in art statistics presented in this dissertation.

CHAPTER 3

STATISTICAL REGULARITIES OF ART IMAGES AND NATURAL SCENES: SPECTRA, SPARSENESS, AND NONLINEARITIES.

Ideas from efficient retinal processing are extended in this chapter to the special class of images that humans themselves have created for viewing by other humans, namely artworks. Several outstanding questions in this area motivate this discussion: Does art conform to the statistical regularities of scenes? If so, might its regularity be explained by efficient visual coding arguments? Could art that does not share such regularities be rare because it doesn't produce equalized responses in the retina? Is there a sense in which artist's are generally efficient with respect to the visual system in their methods of representation? In order to begin answering these questions, this chapter explores statistical regularities in paintings from a major university art collection and demonstrates some links between artworks and efficient coding in the visual system.

3.1 Background

From the standpoint of natural vision, painted artworks are interesting because they are willful, hand-made productions of a visual environment, and this is true regardless of whether they seek to faithfully represent a part of the natural environment. As described in earlier chapters, there is now significant evidence that the early visual system takes advantage of the basic redundant statistical properties of natural scenes (e.g., Field, 1987; Field, 1994; Atick and Redlich, 1990; Geisler *et al.*, 2001; Hyvarinen and Hoyer, 2000; see also previous Chapter). Therefore, it may be no surprise that artworks created to stimulate the visual system will also take advantage

of these statistical regularities. For example, natural scenes show a spatial frequency amplitude spectrum that is well-characterized by the function $1/f^k$ where f is spatial frequency and k is approximately 1 (Field, 1987; Burton and Moorhead, 1987; Tolhurst *et al.* 1992; Field, 1993; Ruderman and Bialek, 1994).⁵ The scale invariance of spatial structure of images from the terrestrial world gives rise to this characteristic spectrum. Artworks that deviate strongly from this regularity may lack a key element of natural scene structure.

Artists create images that are intended to stimulate the human visual system. Since the visual system is matched to the statistics of the world, one might expect the images that artists create to match these statistics. Despite the freedom granted artists to put marks on canvas as they see fit, it is argued that artists do not explore the full range of possible images. Instead, they choose to replicate the basic statistics of the world in their paintings, both for representational and nonrepresentational works. Even paintings such as those that belong to the abstract expressionist movement will be shown to share statistical regularities with natural scenes, despite these images' apparent "randomness." Indeed, the work of few artists could be described as white noise which, from a statistics viewpoint, would be a truly random image (see Discussion).

A number of measures of image spatial structure besides the amplitude spectrum are in common use including the fractal dimension. The fractal dimension is related to the amplitude spectrum and has been shown to be regular for many natural forms (e.g., Mandelbrot, 1977) and for natural scene outlines (Pentland, 1984; Kube and Pentland, 1988; Keller *et al.*, 1987). Fractal dimension has been related to perceptions of roughness and complexity (Pentland, 1984; Cutting and Garvin, 1987), discriminability (Knill *et al.*, 1990; Geake and Landini, 1997), and aesthetic

⁵ Other studies have used spectra to classify natural scenes, e.g. Torralba and Oliva (2003).

preference (Aks and Sprott, 1996; Spehar *et al.*, 2003; Hagerhall *et al.*, 2004). For a 2-D luminance surface, fractal dimension D_f is linearly related to the slope k of the spatial frequency amplitude spectrum as plotted on log-log coordinates:

$$(3.1) \quad k = 4 - D_f$$

(Knill *et al.*, 1990). This relation holds for all images whose amplitude spectra are well-described by the function $1/f^k$. The results of the present study are therefore relevant to studies of the regularity of fractal dimension in natural scenes and a prediction for a typical D_f to be found in paintings is given.

Along with these regularities, this chapter considers some ways that painted art varies compared to natural scenes, notably in terms of intensity distributions and modeled neural responses. There are likely to be differences in these statistics as a function of luminance because the possible range of luminances is much smaller for paintings than it is for natural scenes. Anecdotally, the typical intensity range for natural scenes may be 1000:1 or more; Jones and Condit (1948) proposed that it is 760:1, and this figure is viewed as a rough benchmark by computer graphics engineers who model real-world scenes. But whereas natural scene luminance values are a function of illumination and reflectance, paintings typically have only one illumination and therefore the reflected luminances are a function of their reflectances only. Since reflectances are rarely less than 3% or greater than 90%, paintings rarely produce an intensity ratio greater than 30:1 (see Gilchrist, 1979; Dror *et al.*, 2004). Typical illuminations are also much lower for art than for scenes since the former are generally displayed inside museums, though this has no effect on the *ratio* between the most intense and the least intense luminance.

It has been argued (e.g., Brady and Field, 2000; Mante *et al.*, 2005) that the compressive nonlinearity of the early visual system serves to produce an efficient representation of luminances and contrasts in natural scenes. It has also been noted that an early logarithmic transform serves to transform a differencing operation into a contrast measure (Field, 1994). In this chapter the role that such a compressive nonlinearity plays for both art and natural scenes, and how this nonlinearity solves what is termed the “luminance problem” are considered. Because of the large luminance range found in natural scenes, these early nonlinearities will have significant effects on both the skewness (the third statistical moment of a distribution) and the kurtosis (the fourth statistical moment) of luminant intensities of the image. However, because artists use a similar compressive nonlinearity in the production of their work (i.e., a kind of “artist's gamma,”), the nonlinearities have less impact on the statistics of the artworks.

This study compares the distributions of intensities and contrasts of images before and after the application of a compressive nonlinearity. A model of the responses of visual neurons to these images is also presented.

Examples from two databases were analyzed—one of calibrated natural scenes, one of paintings. Measurements were made of pixel statistics, Fourier spatial frequency amplitude spectra, and modeled center-surround and wavelet filter responses. The findings presented here are based on two large and diverse databases but it should be noted that these are biased image sets and therefore there are likely to be some variations in these statistics for different image sets.⁶

⁶ Note that this study was carried out using greyscale images and thus it does not attempt to account for regularities in color statistics. Color plays a crucial role in the creation of art and a full theory of how regularities in art are related to the visual system would undoubtedly include an understanding of color.

3.2 Materials

The images in this study were comprised of 124 uncompressed tiff images of works from the Herbert F. Johnson Museum of Art, Cornell University, (Ithaca, NY) and 137 calibrated natural scene images (van Hateren and van der Schaaf, 1998). The art images were chosen randomly from the 1139 images classified by the museum as "paintings": images were chosen randomly then included in the test set if they were acquired by the museum photographer between 1999 and 2000. Images acquired by this photographer during this time period all used the same lights (5500K fluorescent; Videssence, El Monte, Calif.), camera (4x5; Sinar AG, Feuerthalen, Switzerland), scan back (Phase One Power Phase; Phase One, Inc., Melville, NY) and copy stand (ZBE, Inc., Carpinteria, Calif.). The exposure of each image was set so that whites had pixel values around 250. The images were also white balanced and color corrected to visually match the painting. No gamma correction was applied. These linear images were converted from RGB space to greyscale using the YIQ transform where the luminance $Y = 0.299 * R + 0.587 * G + 0.114 * B$. The art images were rectangular with dimensions between 818 and 3072 pixels. The image size scaled roughly with the physical dimensions of the artwork. The paintings represent a range of time periods (approximately 12th century CE to the contemporary era), provenances (e.g., Europe, India, US, China—42% were from Europe and America, 58% were from the Middle East and Asia), subject matter (e.g., still-life, abstract, landscape, portrait, scene painting), and artistic movements (e.g., Rajput miniatures, abstract expressionism, surrealism, Rococo). Some of the artists are well-known (e.g. Dubuffet, Bouguereau, O'Keefe), others are not. Natural scene images were randomly selected from van Hateren's database (van Hateren and van der Schaaf, 1998) and scenes with significant blur or buildings were excluded. See Chapter 2 for further description of the 137 natural scene images used.

3.3 Study 1: Statistical Regularities in Painted Art

3.3.1 Methods

A patch of each image chosen at random from within the boundaries of the image was extracted. The patch size was 818 x 818 pixels (818 pixels is the smallest dimension in the set of paintings). A 2-D amplitude spectra was generated for each image and performed a rotational average to get the 1-D spectra, then averaged these spectra and measured the slope of a least-squares fit to this spectrum on log-log coordinates. Each individual spectrum was also fit and an average of these slopes was calculated.⁷

3.3.2 Results

The slope k of the best-fit line to the mean amplitude spectrum plotted on log-log coordinates was -1.23 ($R^2 = .97$) for the art and for the natural scenes it was -1.37 ($R^2 = .98$) where R is the correlation coefficient for the fit. See Figure 3.1. The mean of the slopes on log-log axes calculated for each individual image was -1.21 ± 0.017 (standard error used throughout) for the paintings and -1.40 ± 0.017 for the natural scenes. This value for natural scenes again reflects the fact that both databases used are biased samples. The extrema of slopes for the art were -0.70 and -1.56 for *Spring Festival on the River* by Zhang and *Birth of the Virgin* by Giaquinto, respectively (see Figure 3.8).

⁷ Note that the spatial frequency amplitude spectrum is simply the square root of the power spectrum and therefore the slope of the amplitude spectrum is half that of the power spectrum.

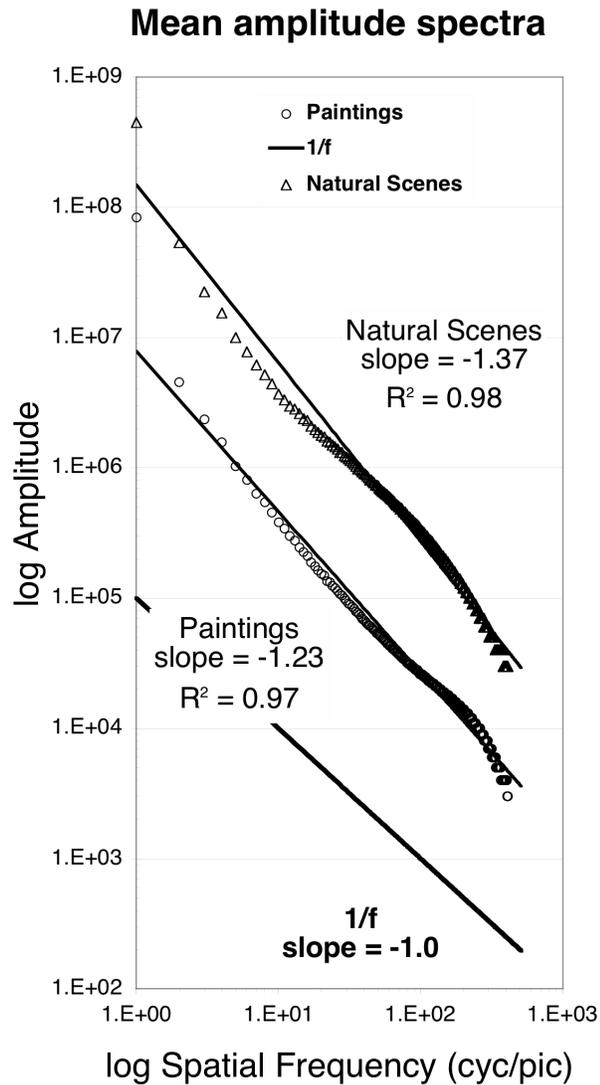


Figure 3.1. Log-log plot of spatial frequency (cycles/picture) versus amplitude (unitless) for the paintings and natural scenes, taken as an average across 1-D spectra. A line with a slope of -1.0 in log-log coordinates is also shown.

A histogram of the distribution of amplitude spectrum slopes is shown in Figure 3.2. It shows that the slopes for the paintings fall in a narrow range around the slopes of the natural scenes but that the two image classes show significantly different means ($p < 0.05$).

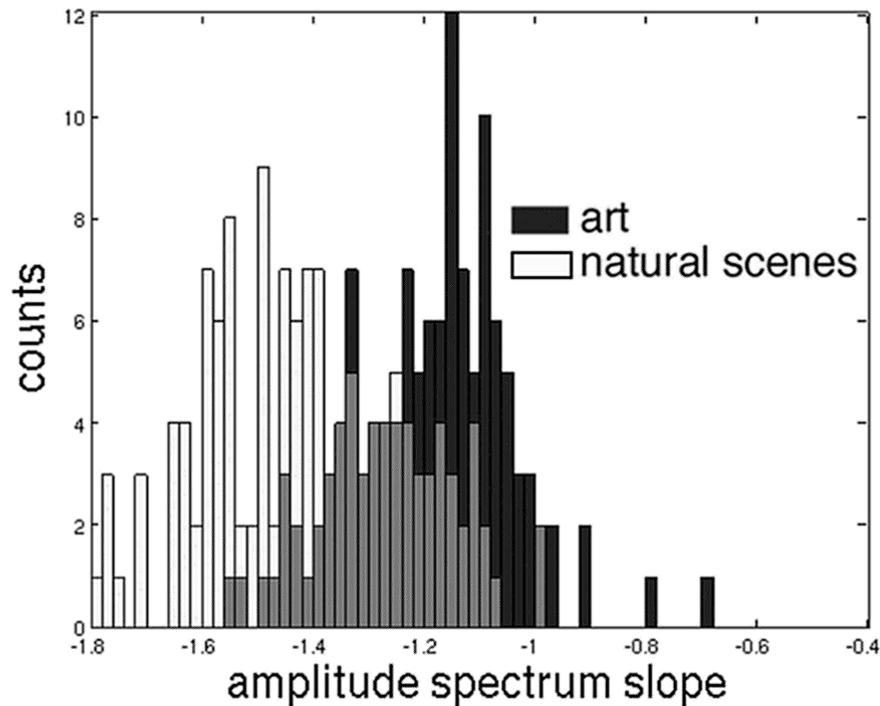


Figure 3.2. Histogram of the best-fit slopes for all art and natural scene images. Grey shaded areas indicate overlap between populations. Note that these slopes were calculated from fits before averaging the 1-D spectra together, whereas the fit in Figure 1 is for the mean spectrum.

In general, paintings with abstract representations had similar slopes as the natural scenes, including *Bleeding Rain* by Bluhm, which had a slope of -1.14 (see Figure 3.3).

3.3.3 Discussion

Judging by this large but biased sample of paintings, the results suggest that a diverse set of painted art selected at random from the collection of a major university museum follows the amplitude spectrum statistical regularities of natural scenes. However, the paintings show a significantly different mean amplitude spectrum slope compared to the natural scenes.

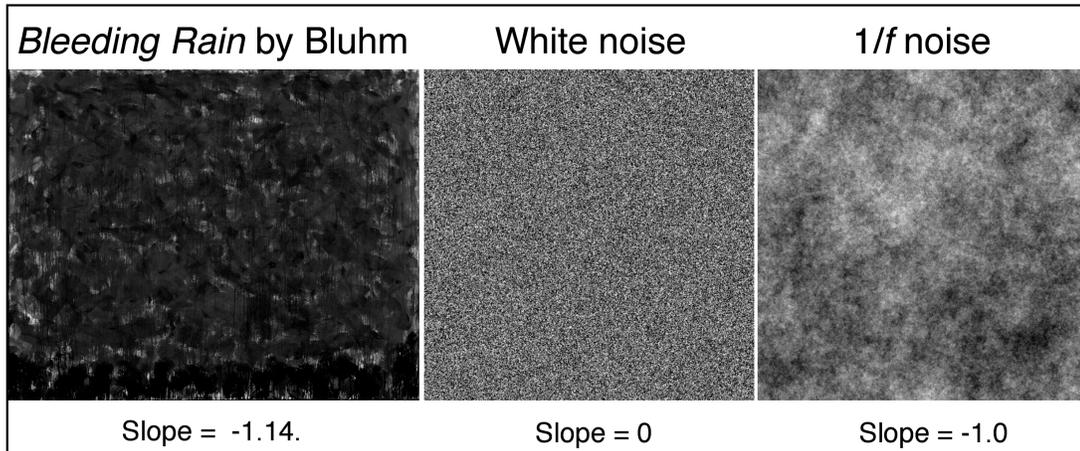


Figure 3.3. *Bleeding Rain* by Norman Bluhm (left) has an amplitude spectrum that is fit by a line whose slope is -1.14 in log-log coordinates; white noise (center) and noise whose amplitude is distributed as $1/f$ where f is spatial frequency (right) have slopes of 0 and 1, respectively. Norman Bluhm, *Bleeding Rain*, 1956, Gift of Katharine Komaroff Goodman. Courtesy of the Herbert F. Johnson Museum of Art, Cornell University.

Slopes for the natural scenes are consistent with other studies, though the values found are different: Tolhurst *et al.* (1992) put the value of k for natural scenes at -1.2 , while both Field (1993) and Redies *et al.* (2007) give a value of -1.0 . The difference between these values and the value presented here is in part due to the fact that each study examined a different biased data. Also, given that blur is a problem with many natural scene databases, it is possible that these different values would be closer to one another if they all used the same frequency cut-offs to calculate slope. That is, because we calculated the slope based on the full range of spatial frequencies, our slope value may be somewhat steeper than the other studies, which did not all include in their calculations the highest frequencies, where blur is most apparent.

Previous work by Taylor *et al.*, (1999; 2002) has demonstrated that Jackson Pollock's drip paintings resemble the basic statistics of natural scenes: the fractal dimension of Pollock's paintings was found to have increased over a small range in a relatively orderly way during Pollock's experimentation with drip-painting techniques.

Pollock paintings' fractal dimension value range spans a similar range as do 1-D natural scene outlines (Spehar *et al.*, 2002). This result implies that paintings as a group will display a fractal dimension D_f of approximately 2.8 (See Equation 3.1). This value is higher than the corresponding value for natural scenes (~ 2.6).

3.4 Study 2: Sparseness and Nonlinearities

3.4.1 Methods

A convolution of a difference-of-Gaussians (DoG) filter with the images described above was used as a simulation of the ganglion cell response. Both the images and a transform of the images where each pixel value was scaled according to its logarithm were tested. The log transform is a rough model of vertebrate cone photoreceptor luminance response (Naka and Rushton, 1966, and Baylor *et al.*, 1987, propose similar compressive functions) and the log recasts differences of intensities as ratios of same, a potentially useful property computationally (Field, 1987). See Rodieck (1965) for specifications for the DoG model. In the present study, $\sigma_1 = 1.7$, $\sigma_2 = 10.2$ and frame size = 129 pixels (see Equation 2.1). 83 images of noise whose amplitude spectrum is distributed proportional to $1/f$, where f is spatial frequency, were tested as a control. Convolutions of the images with log Gabor filters at 3 scales and 4 orientations were performed to model cortical responses (see Field, 1993). Filters uniformly span ~ 1.5 octaves in frequency. The three scales correspond to filters of center wavelength 3 pixels, 6.3 pixels and 13.23 pixels, respectively. Measurements were made of the skewness and sparseness (kurtosis) of the responses for both types of filters to the image, using the linear (untransformed) images and the log transformed images.

3.4.2 Results

In the linear case (i.e., without the nonlinearity), the set of natural scenes gave a

more sparse response than the paintings for the DoG filters, and for wavelet filters at all orientations and scales (see Figure 3.4). When a $\log(x)$ input nonlinearity was applied, the sparseness of the DoG and wavelet filters' responses to the log-transformed paintings was greater than or equal to the response sparseness of the log-transformed natural scenes (Figure 3.5). The same pattern held for the skew (Figure 3.5). Horizontal- and vertical-orientation filters generally showed a more sparse response to the natural scenes than diagonal orientations. Mean sparseness of the noise was approximately zero as predicted (-0.008 ± 0.006 before filtering and -0.003 ± 0.002 after filtering in the linear case).

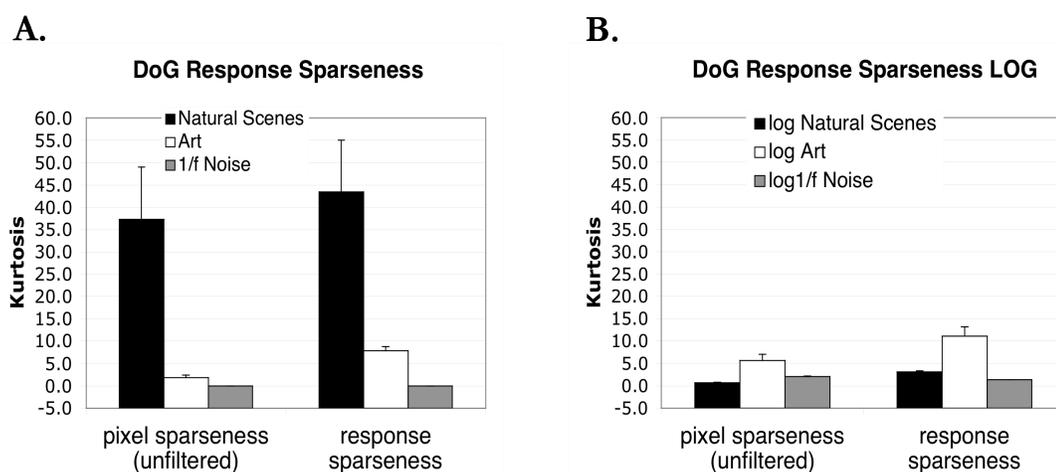


Figure 3.4. Sparseness of the pixels and of the responses of a difference-of-Gaussians (DoG) filter to natural scene images, art images and 1/f-distributed noise, for the linear images (A) and for the log-transformed images (B). Pixel sparseness is the pixel histogram kurtosis of the unfiltered image, response kurtosis is calculated on the convolved image, which is cropped to remove edge effects. Error bars show standard error. Natural scenes show greater sparseness compared to the set of paintings in terms of modeled retinal response (DoG convolution) in the linear case but the paintings gave more sparse responses in the nonlinear case.

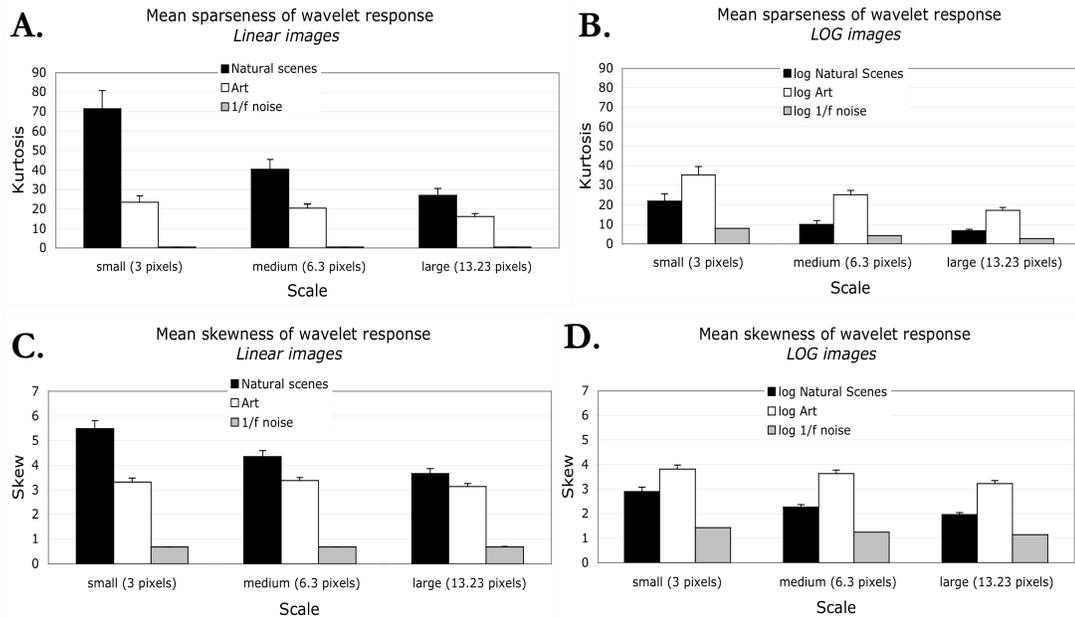


Figure 3.5. Plots showing response sparseness to wavelet filters across 3 scales (high, medium, and low spatial frequency bands in the top, middle and bottom plots, respectively) averaged across orientation. (A) shows results for linear images, (B) for log-transformed images. Note that whereas responses to the natural scenes were more sparse than the paintings in the linear case across all frequency bands and orientations, the paintings were more sparse when the log nonlinearity was applied before filtering. Plots showing response skewness to wavelet filters across 3 scales (high, medium, and low spatial frequency bands in the left, middle and right plots, respectively) averaged across orientation. (C) shows results for linear images, (D) for log-transformed images. Responses to the natural scenes were more skewed than the paintings in the linear case across all frequency bands and orientations, but the paintings were more skewed when the log nonlinearity was applied before filtering. High, medium and low frequency scales correspond to filters of center wavelength 3 pixels, 6.3 pixels and 13.23 pixels, respectively.

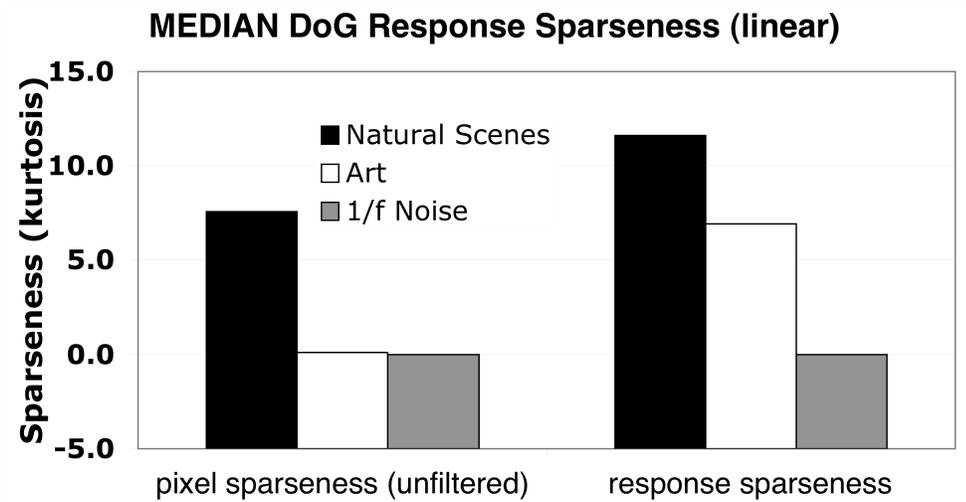


Figure 3.6. Plot of median sparseness (kurtosis) values for natural scenes, paintings, and noise before filtering (pixel sparseness, left plot) and after DoG convolution (response sparseness, right plot). Data are for all 137 linear natural scenes, 124 linear paintings, and 83 linear 1/f noise images. This plot suggests that the high mean kurtosis for natural scenes shown in Figure 3 is the result of a handful of images (~12) with very high skewness and kurtosis. These images generally show sunlight filtered through a dark canopy of trees.

Note that the high mean kurtosis of the linear natural scene images reflects the fact that a small number of these images (12) had extremely large kurtosis values (>70). These images were generally of sunlight filtered through dark foliage and their pixel intensity distributions showed very heavy tails. Figure 3.6 plots the median kurtosis values for the linear images before and after DoG filtering for comparison.

3.4.3 Discussion

As would be expected, a modeled luminance nonlinearity has a greater compressive effect on the high-dynamic range natural scenes than it does on the set of paintings in terms of sparseness. Adding a luminance nonlinearity gives a more sparse response in modeled retinal and cortical cells for the paintings compared to the natural scenes; without the nonlinearity the natural scenes were more sparse. Cases of very high sparseness in the natural scenes' luminance distributions typically result from a

few regions with relatively intense luminance. This produces both high kurtosis (due to the heavy tail) and positive skew, since the tail is typically one-sided toward positive luminances. The compressive nonlinearity reduces this tail in the high luminances, reducing both skew and kurtosis, and it makes the roughly log-normal natural scene distributions more Gaussian. However, this nonlinearity has little effect on the paintings, whose luminance distributions already have a low skewness and a small dynamic range before compression.

Greater sparseness does not necessarily correlate with greater interpretability or aesthetic appeal. But these results suggest that artists apply a type of nonlinear luminance control that is manifest in their paintings. Indeed, if paintings simply took a linear scaling of the luminances in natural scenes, the resulting images would be dominated by low intensities and would appear to be very dark. Figure 3.7 shows an example of a natural scene from the sample scaled linearly with respect to luminance, and scaled using a compressive nonlinearity (a log function).

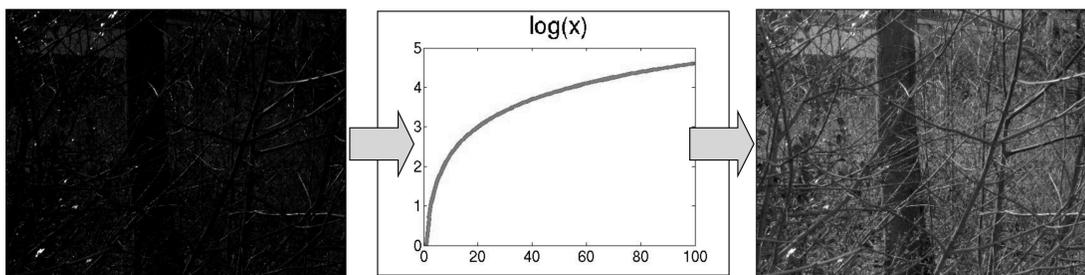


Figure 3.7. A natural scene from van Hateren's collection (image no. 619, van Hateren and van der Schaaf, 1998) displayed with linear scaling in luminance (left) and after application of a $\log(x)$ luminance nonlinearity. Note that the latter image appears less dominated by very dark and very bright regions. See also Fig. 5.2.

3.5 General Discussion

The results presented in this chapter suggest that a diverse set of paintings shares many of the basic spatial statistics of natural scenes but that the paintings show

characteristic differences in terms of pixel intensity distribution statistics and modeled neural responses. These differences reflect artist's solutions to what is referred to here as the luminance problem.

The class of art images labeled “paintings” was used in this study primarily because in general, paintings are constructed by applying pigment over the entire surface of the canvas. They thus represent material reflectances throughout the scene, while drawings, for example, are less involved with representing material reflectances. However, drawings as a class are likely to contain some of the same statistical regularities as paintings (drawings, to a lesser or greater degree, contain detail at many scales, which gives drawings roughly $1/f$ amplitude spectra). Efforts are underway to analyze a number of classes of art, including drawings. In addition, the medium of painting is typically concerned with replicating or interpreting a real or an abstract scene, whereas drawing—and especially caricature—is more concerned with conveying the identity of objects or people (see Gombrich, 1961).

Why does the visual world have a $1/f$ structure, which is reflected in paintings? Field (1994) and Field and Brady (1997) have shown that a self-similar sum of functions produces global $1/f$ structure in synthetic images. Ruderman (1997) has shown a similar distribution of occluding objects will also produce such a spectrum. These results suggest that $1/f$ structure can be generated by simple sums of objects in the proper proportion of sizes and that a world of objects with fractal edges is not required to produce global 2-D fractal structure.

A statistically random painting is simply white noise—i.e., an image in which the luminance at every point is randomly assigned from a Gaussian distribution, and the spectrum is flat. Results from this sample suggest that even abstract non-representational works do not approach this level of randomness. Instead, the sample showed similar amplitude spectrum characteristics across artworks.

These results suggest that artists are applying a form of compressive nonlinearity in the process of utilizing the small range of reflectances available in paint. Ways of measuring this “artist’s gamma”(or, more accurately, the “artist’s look-up table”) are discussed in Chapter 5, which also addresses the question of whether different artists use different compression strategies.

As with the reproduction of geometric form in painting, the above results may imply that this is an example of the “El Greco fallacy” (see Anstis, 2002). This fallacy is the argument that El Greco painted elongated figures because he suffered from astigmatism and saw the world elongated. However, the artist viewing his own canvas would still have the ability to see that his images did not match the world. In a similar way, if the artist was viewing the world through a compressive nonlinearity (as implied by both physiology and psychophysics) then one would not expect the artist to simply produce this nonlinearity on canvas since the image on the canvas must also pass through that nonlinearity. However, in the case of luminance, unlike geometry, the limitations of paint make it impossible to faithfully represent the world on the canvas. Therefore, one can ask the question: What is the most perceptually realistic transformation of luminances under the assumption that the luminance range must be significantly reduced? A linear transform might seem obvious. However, for images with high positive skew, a simple linear compression results in very dark images with a few bright regions (Figure 3.7) since such images have long tails that extend far into the high intensities. The artist's strategy to nonlinearly transform luminances from the world into paint may be a general solution to what is termed the luminance problem. Photography faces a similar problem of compressing the range of intensities onto the small range of intensities available in a print. Although photographic film also employs nonlinear luminance compression, the printing process allows and often employs local adjustments to exposure (burning and dodging). This can certainly

improve the perceptual quality of the image but the method cannot be modeled as a simple function relating input and output. Indeed, artists likely also do something analogous to such local adjustments, but this is beyond the scope of the current study.

The study presented here is not meant to prescribe a formula for acceptable art. It is intended primarily as a means of understanding the visual system through the analysis of artworks. The regularities observed here are of a low statistical order, meaning that they suggest only the most minimal constraints on art, given the infinite variety of possible images. Furthermore, as some have said, artists often view pronouncements about the typical characteristics of art as a challenge to create works that defy those characteristics and are yet undeniably art (Bloom, 2000). The regularities described here are likely to be similarly fallible.

Moreover, it is probable that many painters have already defied these regularities and still managed to produce works deemed art. For example, consider minimalist painter Agnes Martin's grid paintings from the 1960s, which are a group of six-foot square white canvases with perfect rectangular grids drawn in pencil. These images likely have a great deal of energy at the spatial frequency defined by the grid lines, and little in other frequency regimes, and thus their spectra could deviate strongly from a $1/f$ relationship. Also, any technique that results in a large degree of blur would serve to steepen the amplitude spectrum (i.e., make its slope more negative). Highly detailed works that show little low-frequency variation would show a shallower spectrum (slope nearer zero). Examples of the images from the sample that show extreme slope values give some idea of these effects (Figure 3.8). Although the Giaquinto painting is not blurry, the details are smooth enough relative to the strong large-scale (low frequency) luminance changes to cause a steep tilt in the spectrum.

Artists have also challenged the characteristic dynamic range limitations of painting by using light sources as a pigment. Consider for example the work of

American artists James Turrell or Dan Flavin (though one could argue these artists are not painters in the strictest sense).



Figure 3.8. (A) Corrado Giaquinto, The Birth of the Virgin, 1751-1755. Herbert F. Johnson "Class of 1922" Acquisition Fund. Courtesy of the Herbert F. Johnson Museum of Art, Cornell University. This is the painting with the steepest (most negative) amplitude spectrum slope in the sample, -1.56 . (B) Zhang Ziduan, Spring Festival on the River, Late Qing (19th c.). Gift of Drs. Lee and Connie Koppelman. Courtesy of the Herbert F. Johnson Museum of Art, Cornell University. This painting had the shallowest slope (nearest zero), -0.70 .

One difficulty in determining the extent to which art as a class of images follows the statistical regularities of natural scenes is the establishment of a representative corpus. Is the art in museums or art books sufficiently representative? Some interesting work in this area attempts to define a canon of impressionist paintings by enumerating scholarly references to those images. Cutting (2006) has shown that even

within impressionist art, the ‘canon’ is a relatively selective grouping and it is a function of the vagaries of early collectors.

Both the natural scene images and the art images used here represent biased samples of their respective image classes. For example, the natural scenes contain no large vistas, and the majority of the art images in the collection are of Asian provenance. One would expect some differences in these statistical relationships for other collections of images though the deviations would likely be small. Statistical differences among different classes of art using an expanded database are considered in the next chapter. Results suggest there are few low-level statistical differences among classes (unpublished data).

Whether because of taste, history, materials, the structure of the visual system, or the interaction of these factors, human art may tend to adhere to the basic statistical regularities of natural scenes even when the painting is highly abstract or non-representational. These results suggest that nearly all painted art—including abstract painting—shares essential statistical regularities with natural scenes but that art is characteristically limited in terms of dynamic range in ways natural scenes are not.

Moreover, the regularity of spatial statistics in art could implicate efficient neural coding mechanisms as the ultimate cause of the regularity. Evidence presented in the next chapter supports the notion that art from different parts of the world and different types of art (abstract art, portraits, etc.) show similar statistical properties, suggesting that the prevalence of such regularities could be the result of artists attempting to make images that can be perceived by the human visual system. In particular, the proposed goal of response equalization in retinal ganglion cells described in Chapter 2 may in general constrain artists’ creations to the subspace of images which allow this form of efficient retinal coding. If ease of processing or aesthetic judgment are shaped by the ability of the early visual system to efficiently process an image, then the fact that art

that does not match the statistical regularities of the natural world is rare could be due to the inefficiency of processing such images in early vision.

As will be shown in the next chapter, even non-representational art shows a similar fall-off in spatial frequency amplitude as that for natural scenes and other forms of art. After presenting these data, a proposal for why art generally shows statistical regularities will be developed in Chapter 5.

CHAPTER 4

VARIATIONS IN INTENSITY STATISTICS FOR REPRESENTATIONAL AND ABSTRACT ART, AND FOR ART FROM THE EASTERN AND WESTERN HEMISPHERES

4.1 Introduction

The previous chapter suggested that a large and diverse sample of roughly luminance-calibrated paintings shows strong statistical regularities that are quite similar to statistical regularities of natural scenes. This study found that 124 digitally scanned paintings from the Herbert F. Johnson Museum of Art at Cornell University, representing a roughly equal share of works from the Western and Eastern hemisphere, showed the same spatial correlation structure (i.e., Fourier spatial frequency amplitude spectrum) as that seen in natural scenes, though the fall-off of this relationship was significantly different for the paintings and the natural scenes. Redies *et al.* (2007) obtained similar results: this study found that spectra of a large and diverse sample of graphic art are more similar to spectra of natural scenes than they are to spectra of scientific illustrations, or to spectra of close-up views of common objects. The amplitude spectrum measures the relative contribution of different Fourier components to the image as a whole. It is related by a Fourier transform to the autocorrelation function, which measures correlation in neighboring pixels' intensities.⁸ Colloquially, the amplitude spectrum describes the proportion of large-scale structure and fine detail in an image. The fall-off in amplitude plotted on log-log axes goes as $1/f^k$, where f is spatial frequency and k is about 1.4 for natural

⁸ Many researchers express the Fourier decomposition as a power spectrum. The amplitude spectrum is the square root of the power spectrum, and the slope of the amplitude spectrum is simply half the slope of the power spectrum.

scenes (Chapter 3; see also Tolhurst *et al.*, 1992; Field, 1993) and about 1.2 for paintings (see Chapter 3). The paintings, in other words, have a slightly steeper slope, i.e., they have relatively more low-frequency (large scale) amplitude compared to natural scenes. A related study by Taylor *et al.* (1999) showed that for Jackson Pollock's drip paintings of the 1950s, the paint layer's outlines display fractal-like scale invariance and that Pollock's work showed characteristic fractal dimension statistics at different points in his career. But whereas the tests by Taylor *et al.* (1999) relate to the fractal shape of paint mark outlines obtained after thresholding, the results described in Chapter 3 and Redies *et al.* (2007) relate to fractal-like multi-scale structure in the variation of tones over space. In particular, the latter studies suggested that nearly all paintings show $1/f$ scaling of contrasts across spatial frequency in terms of their surface intensity maps. This is true regardless of the pigment application technique used in the works. That is, fractal-like scaling is a property not unique to paint boundary edges in Pollock's paintings—it is rather a characteristic of the great majority of paintings and, indeed, of natural scenes.

Fractal or scale-invariant contrast statistics in images may require a slope of 1.0 though there is active discussion regarding how such $1/f$ structure can arise. Field (1994) argued that a self-similar array of 2-D functions, which increase exponentially in number with decreasing scale, can produce $1/f$ structure and Ruderman (1997) argued that a similar distribution of functions that allow occlusions can also produce such structure. Balboa *et al.* (2001) have argued that exponential scaling of the number of objects that comprise a scene is not necessary for producing a $1/f$ amplitude spectrum. Indeed, Field and Brady (1997) noted that a fractal edge with the right scaling (i.e., an edge with a length that increases in proportion to decreasing scale) can also produce $1/f$ structure in the 2-D amplitude spectrum. See Figure 4.1 for an example. A simple straight edge is insufficient (resulting in a rotationally averaged 2-

D amplitude spectrum of $1/f^2$). However, it should be emphasized that objects with fractal edges—e.g., Mandelbrot or Julia sets—are not required to produce scale-invariant contrast statistics. Nor does $1/f$ scaling in 2-D amplitude spectra guarantee fractal edges or complete self-similarity.

Given that the visual system is well matched to the statistics typical of the natural world, it is perhaps no surprise that painters should reproduce those statistics in their work. Since painters create their art for human viewing it makes sense that they should on average conform to these regularities. Redies (2007) and others have argued that the similarity between natural scene statistics and art statistics is shaped by a universal aesthetic preference for natural statistics. While this is an interesting possibility, there is a need for greater consideration of the importance of perceptibility and manufacture in accounting for this similarity (see Chapter 5).

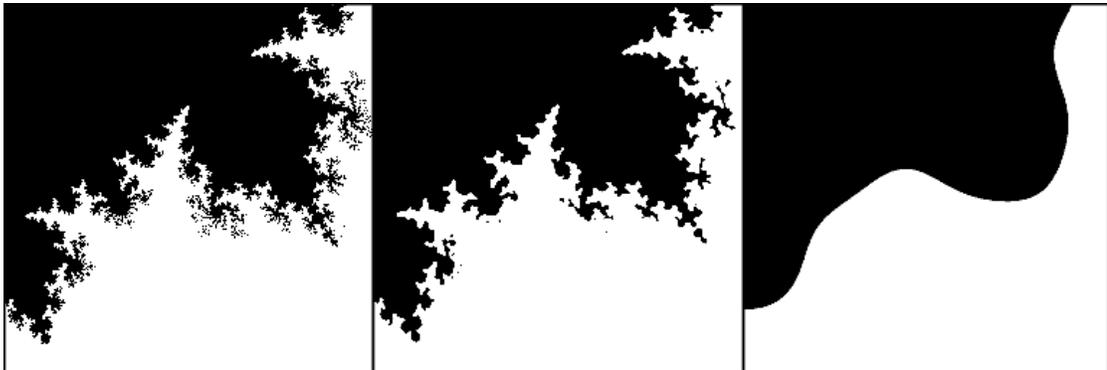


Figure 4.1. The monotone boundaries shown here have amplitude spectrum slopes of -1.04 , -1.22 , and -1.43 (from left to right), showing that a fractal edge by itself can generate a $1/f$ amplitude spectrum, and also that smoothing this single edge is a way to steepen the slope of the spectrum.

It should be stressed that it is not possible to reproduce many of the luminance properties of natural scenes on canvas. As will be noted in Chapter 3, paintings are limited in the ratio of the most luminant point to the least luminant point (i.e., dynamic

range), compared to natural scenes: paintings have a dynamic range of roughly 30:1 while scenes often exceed 1000:1. This difference is related to the number of illuminants for natural scenes (many) and for art (typically one). It has been argued that artists use global nonlinear luminance compression—in a manner analogous to that employed by photoreceptors—as a way to compensate for these dynamic range limitations.

This study tested statistical properties related to provenance. A simple forced-choice classification scheme is also used for the paintings in order to compare typical statistics across content classes and to test whether abstract subject matter is more or less likely to resemble natural scenes in terms of statistics, compared to other classes of paintings. In particular, for the first study, the images were classified by provenance to investigate cultural influences on art statistics. The second study is concerned with how a rough approximation of the range of depths in representational images can affect basic statistics, and how abstract content affects these statistics. The statistical measures are used to identify regularities in grey-scale pixel intensity distributions and in spatial structure. This study and future work with higher order statistics could offer new methods of image search and discrimination (see, e.g., Motoyoshi *et al.*, 2007, for an approach to discriminating surface properties using intensity statistics; see also Torralba and Oliva, 2003, for studies of natural image classification using low-level statistics).

4.2 Study 1: Painting Provenance

4.2.1 Methods

The images used are from the same source as those described in Chapter 3, the Herbert F. Johnson Museum of Art, Cornell University, though the size of the database has been expanded. Briefly, the 140 paintings used were all scanned by the

same museum photographer using the same equipment and procedure. They are a fairly random sample representing about 10% of the museum's diverse collection of paintings. Paintings of Eastern provenance (Asia) represent 51% of the works, Western works (Europe and the Americas) form 49% of the total, and together the works span approximately 900 years of art history. Provenance classification is based on metadata provided for each image by the museum (see Figure 4.2). The first four statistical moments of the pixel intensity histograms (mean, variance, skewness, kurtosis) were calculated as well as the modal intensity for all images using intensity distributions for the whole image. In addition, the amplitude spectrum slope of the images was calculated based on the central 818 x 818 pixel patch of each image (in order to compare across images, an 818 x 818 patch was used, which is the largest square that can be extracted from all images in the collection). Tests showed that choosing a randomly selected square patch produced the same statistical regularities as did choosing the central patch. These calculations relate to basic luminance distribution statistics and Fourier components for the art; there are hosts of other statistics one could measure that may be relevant, e.g., wavelet statistics.

4.2.2 Results

Mean statistics for the two groups are shown in Table 4.1. While the mean intensity was significantly lower for Western works ($p < 0.0001$), this group had a roughly 1.9 times higher variance than the Eastern works ($p < 0.0001$). Both the mean and variance are a function of the scaling applied to the images so both groups' intensities were normalized to their respective means.

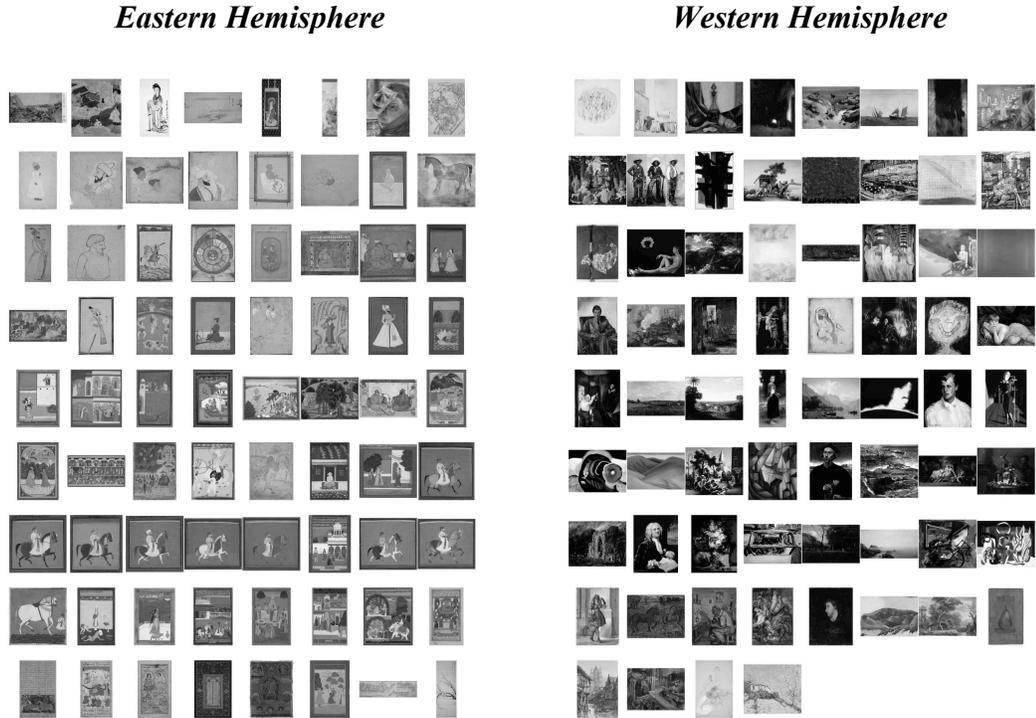


Figure 4.2. Images as classified according to provenance (place of origin). All images are courtesy of the Herbert F. Johnson Museum of Art, Cornell University.

In this case, the variance was 3.2 times higher ($p < 0.0001$) for the Western works (all other statistics are unchanged under normalization). This is equivalent to saying that the ratio of the standard deviation to the mean for the two classes of paintings is 1.8 times greater for the Western works compared to Eastern works (see Equation 4.1).

$$(4.1) \quad 1.8 \times \frac{SD(East)}{Mean(East)} \approx \frac{SD(West)}{Mean(West)}$$

Though the sparseness (kurtosis) of the pixel intensities was not significantly different for images of Eastern and Western provenance, there were significant differences in skewness and amplitude spectrum slope for the two groups, with p values of $p < 0.0001$ and $p < 0.002$, respectively.

Despite the fact that the skewness was different for the two groups, the sparseness of the pixel intensities was not significantly different. Skewness describes how far to the left or right of the mean of a distribution the majority of the samples lie; negative skew is found when more samples are to the left of the mean, positive skew is found when more samples are to the right of the mean.

Table 4.1. Pixel intensity statistics for art of the Eastern and Western hemispheres, with standard error shown in parentheses. Asterisks denote significant differences ($p < 0.05$) between mean statistics for Eastern and Western works.

	No. of Images	Mode (std. error)	Variance (std. error)	Skewness (std. error)	Pixel Sparseness (std. error)	Amp. Spectrum Slope (std. error)
<i>Eastern Hemisphere</i>	72	137.5* (5.67)	1721.6* (99.8)	-0.314* (0.119)	1.15 (0.373)	-1.19* (0.0143)
<i>Western Hemisphere</i>	68	90.53 (10.70)	3251.1 (260)	0.428 (0.137)	0.950 (0.347)	-1.27 (0.0237)

In general, natural scenes' high sparseness is largely due to high positive skew in their roughly log-normal intensity distributions. Paintings have a far smaller dynamic range and skewness than natural scenes (see Chapter 3). Here it is hypothesized that the difference in skew between images of the Eastern and Western hemisphere is related to the background color of the images, since the background appears to account for the majority of the images' pixels. In particular, it is observed that the

Western works generally had a light foreground and dark background while the Eastern works generally had a dark foreground and light background. Such differences should be reflected in the mode of the intensities. As expected, the modal intensity (from among the 256 available in 8-bit greyscale images) for the Eastern works was significantly higher than that for the Western works, with $p < 0.0002$.

4.2.3 Discussion

Results from this sample indicate that paintings from the Eastern and Western hemispheres differ significantly in terms of the normalized variance and the skewness of their pixel intensity distributions, and in terms of their amplitude spectrum slope. This difference in normalized variance suggests Western painters may tend to maximize the variance of their intensities for a given mean, which could be related to the strategies Western artists use to represent a range of strongly varying illuminations. Differences in skewness are attributed to the typical choice of background color for images in this database from the two hemispheres, as shown by the significant difference in the mean modal intensity for the two groups.

Amplitude spectra for Eastern works in this sample were significantly shallower than those of Western works. This may be a result of the media typically used in the production of the work. Eastern painters in the sample—especially the large proportion of 19th century Indian paintings—tend to employ water-based pigments and fine brushes, whereas the Western works tend to employ oil paints, larger brushes and techniques like feathering. It is also possible that these differences are rooted in cultural notions of pictorial representation. A more detailed study of techniques and cultural influences could potentially disentangle these effects.

4.3. Study 2: Painting Content

4.3.1. Methods

Six judges (1F) were asked to classify the 140 images (presented in greyscale) as either "landscape," "portrait/still-life," or "abstract" using a forced-choice paradigm. The landscape and portrait/still-life discrimination was intended to reveal statistical regularities related to the represented range of depths. The abstract category was meant to identify statistical regularities related to non-representational content. Though there are limitless ways to categorize art by content, and though not all paintings would necessarily fit comfortably in these three categories, this method allows an approximate discrimination among paintings with a large implied range of depths, a small implied range of depths, and no implied range of depths.

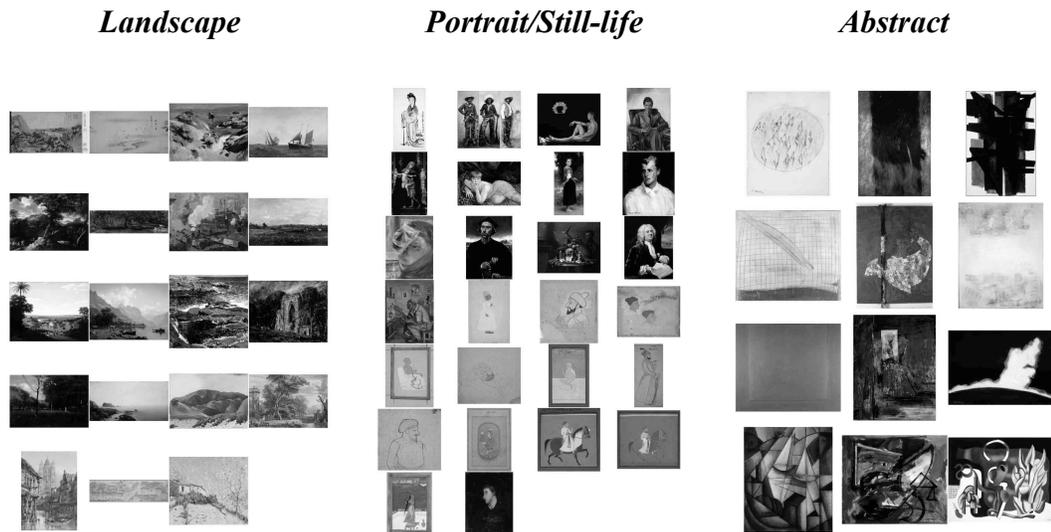


Figure 4.3. Image categories (from right to left: landscape, portrait/still-life, abstract) as determined by the agreement of 6 judges using a forced-choice paradigm. All images are courtesy of the Herbert F. Johnson Museum of Art, Cornell University.

Statistical analysis was carried out for images whose classification was agreed by all judges. The first four statistical moments were calculated for intensity distributions

across the whole greyscale image; spatial frequency amplitude spectrum slope was again calculated for the 818 x 818 pixel square at the center of each image.

4.3.2 Results

Judges agreed on the classifications of 57 of the 140 images (see Figure 4.3). There were no significant differences in distributions of any of the first four statistical moments (mean, variance, skewness, kurtosis) among the classes, nor between concrete works (landscape plus portrait/still-life) and abstract ones. For the amplitude spectrum slope, a significant difference was found between abstract and concrete categories ($p < 0.03$), though both the landscape and portrait/still-life categories separately approached significant differences with the abstract works ($p = 0.058$ and $p = 0.054$, respectively). Mean values are given in Table 4.2 (see also Figure 4.4).

Table 4.2. Table shows how amplitude spectrum slope varies with image content as determined by a forced-choice test. Standard error is shown in parentheses.

Classification	Number of Images	Amplitude Spectrum Slope (standard error)
<i>Landscape</i>	19	-1.26 (0.0387)
<i>Portrait/Still-Life</i>	26	-1.25 (0.0300)
<i>Abstract</i>	12	-1.13 (0.0614)*

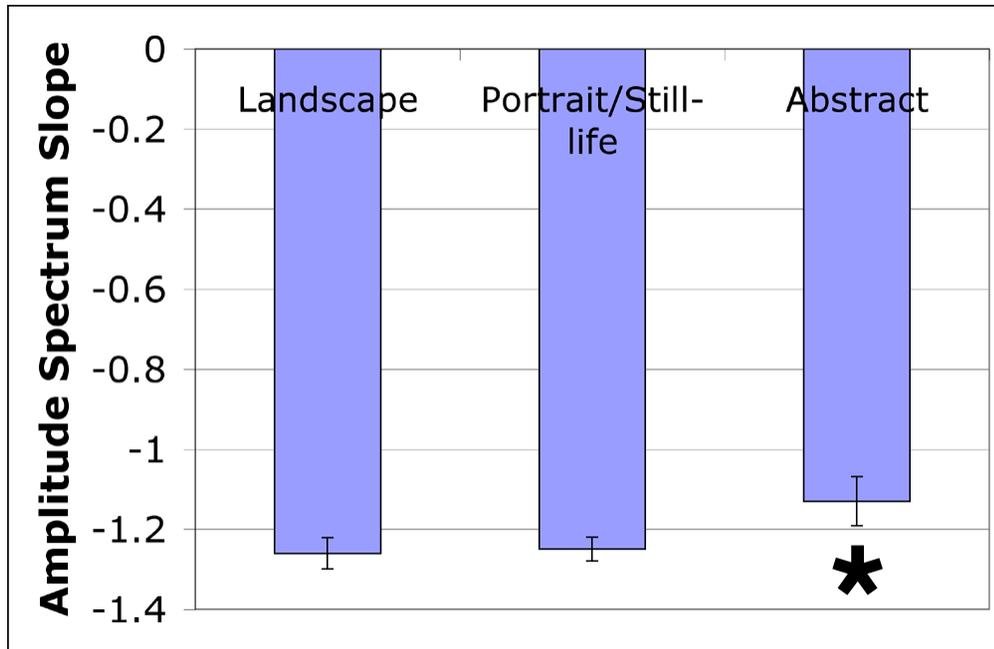


Figure 4.4. Plot of mean amplitude spectrum slope by image category. Abstract works were significantly different from the other two classes taken together ($p < 0.03$). Abstract works also approached significance with both the Landscape and the Portrait/Still-life categories individually, with $p = 0.058$ and $p = 0.054$, respectively. Error bars show standard error.

4.3.3 Discussion

In this diverse though biased sample, only one significant difference between these content-based classes (concrete vs. abstract) was found. This suggests that regardless of subject matter—and even when the subject matter does not correspond to real-world objects or scenes—artists generally reproduce the regular statistics of the visual world in their paintings. Though art that is not strictly representational (i.e., works characterized as abstract here) can be produced with some reliance on random processes (e.g., such artistic movements as automatism, Dada and abstract expressionism), works in this class are hardly random in a statistical sense, at least for this sample. That is, despite their disconnection from concrete imagery and despite the possibility of elements of randomness in their production, abstract works show similar

statistical regularities as representational works. Moreover, basic spatial statistics for art are very similar to those of natural scenes (see Chapter 3), though the portraits, landscapes and abstract works each had significantly different mean slopes compared to the mean slope of the natural scene collection (1.40), with $p < 0.0002$, $p < 0.003$, and $p < 0.0001$, respectively.

Such regularity suggests that in order to produce a work fit for display in a museum, artists on average recreate the statistical regularities of the natural world. As noted in Chapter 3, this does not mean that all paintings must conform to these regularities in order to be deemed art. It merely suggests that the ubiquity of the statistical regularities in the natural world is generally sufficient to cause an artist to be unaware of their presence in paintings. In particular, the amplitude spectrum slope of the natural world—which is related to the proportion of large-scale structure and fine detail—is simulated even by artists who do not attempt to portray a scene from the natural world, as in abstract works.

It is possible that abstract art shows spatial statistics similar to natural scenes in order to suggest the presence of real-world objects, even if no specific objects are depicted. As noted in Rogowitz and Voss' (1990) "cloud watching" experiment, when observers view computer-generated fractal "cloud" images, whose edges display fractal dimensions like those typical of natural images, the observers spontaneously report seeing namable objects. This suggests that in abstract art, as with fractal "cloud watching," spatial statistics that resembles those of natural scenes could promote the perception of object-like structure in non-representational images.

It is not clear why the amplitude spectrum slopes of non-representational works in this study were significantly shallower than those of representational works. This difference may relate to shading, which is used to denote shadows along edges in representational works, but is generally absent from non-representational works.

Shading would tend to make edges softer and thereby lead to the steeper amplitude spectrum than that observed for representational works. Abstract works as a group may also be concerned with an array of similarly sized forms, whereas representational works may tend to show one larger form (i.e., the painting's subject) surrounded by many smaller forms, which could also contribute to the difference in amplitude spectrum slope.

4.4 General Discussion

The major findings of this chapter are the following: for a large and diverse sample of paintings from a major university museum collection, there are systematic variations in the proportion of large- and small-scale detail (mean amplitude spectrum slope) for images with different psychophysically-assigned content identifiers, and for images produced in the two hemispheres. When classified by provenance, paintings also showed significant variations in mean pixel intensity skewness, which is likely related to the background tone typically used in the two hemispheres.

The finding of a systematic difference in the modal intensity, and the related difference in skewness by provenance in this sample stands in contrast to the finding of no significant differences in skewness related to image content. Of course, this could be a result of the smaller sample size used in the image content experiments. But the fact that the relatively small sample used in the image content studies could reveal differences in spectral content suggests that the sample was large enough to reveal first-order statistical differences among the image types.

Generalizing beyond this sample would require more data but based on the statistical analysis given, it is suggested that Western artists of the past millennium typically choose to paint dark backgrounds and light foregrounds while Eastern

painters typically choose light backgrounds and dark foregrounds.⁹ Artists of the Western hemisphere may also have introduced differences in their spatial statistics relative to Eastern artists by dint of their preferred medium (i.e., oil-based pigments).¹⁰

4.5 Conclusion

It seems sensible that paintings—which are created in order to be viewed by the human visual system—reproduce the statistical regularities of the world (see Chapter 3; also Redies *et al.*, 2007). Moreover, as has been shown, paintings as a group show spatial statistics that are quite similar to those of natural scenes despite the need for painters to impose massive nonlinear compression with respect to luminance. Various types of human art work—including abstract works—show essentially the same basic amplitude spectrum shape. However, different classes of art also vary systematically across a number of statistical measures that are relevant to the efficient coding of statistically regular natural scenes. In particular, works from the Eastern hemisphere showed significantly more negatively skewed intensity distributions than works from the Western hemisphere, a difference attributed to systematic differences in background tone. Eastern works also showed a higher proportion of high frequency amplitude compared to Western works in the sample. In addition, abstract works showed a higher proportion of fine detail compared to representational works in the sample, though abstract works followed the general regularities of representational works and natural scenes.

⁹ There are of course exceptions to such trends: Japanese lacquer works typically have deep black backgrounds and lighter-toned foregrounds.

¹⁰ Interestingly, Voss and colleagues have found that examples of 1-D amplitude spectra of Western and Eastern music are quite similar (Voss and Clarke, 1978), and that even stock price data, filtered with 1/f noise and presented aurally, possess enough of the statistical regularity characteristic of Eastern hemisphere music to convince Chinese listeners that they are listening to music from Asia (Voss, 1985).

The next chapter will return to the notions developed in this chapter concerning the implications of the preceding studies of efficient coding of art images to the study of art itself. Similarities between basic statistics for art of all kinds and for natural scenes are viewed by what is termed the *perceptibility hypothesis* as a necessary byproduct of the need to create an image that is readily perceived by the eye and of the nature of manual pigment application techniques. It will also expand on this notion to suggest that art making requires a certain degree of efficiency with respect to the visual system, and that statistical properties of art offer evidence to support this idea.

CHAPTER 5
TOWARDS A THEORY OF NATURAL ART:
THE PERCEPTIBILITY HYPOTHESIS AND
THE ARTIST’S LOOK-UP TABLE

Art, one can suppose, is created in order to be seen. But despite the recent accumulation of evidence from visual neuroscience and other fields, many fundamental questions about the relationship between art and the visual brain remain unresolved. What is it that a painting can convey to the human eye? Are we compelled by some innate sense of aesthetics to look at some images and not at others? Or does the perception of art present us with an inscrutable interaction between the viewer's experience and the artist's choice of colors and forms? Could paintings follow certain subtle rules that facilitate or perhaps permit communication between canvas and cortex? These questions have been approached from a variety of perspectives and by a host of academic disciplines but few have considered recent neuroscientific findings on efficient neural coding of vision. In this chapter, recent studies of efficient coding of statistical predictability in natural scenes and in art described earlier are brought together to construct a new framework for the study of art and art-making. This framework represents a novel mode of learning about the role of the visual system in the trajectory of visual art.

Here the arguments and experiments described earlier in this dissertation are brought together to sketch a theory of “natural art” based on efficiency. The chapter begins by summarizing the literature and the previously described finding of basic similarities and differences in statistics between paintings and the natural world. With regard to the similarities, proponents of the “affect hypothesis” argue that these statistical regularities reflect humans’ aesthetic foundation. Although there may be

some merit to this argument, this position is found to be currently lacking in supporting evidence. Instead, it is argued that statistical regularities are a prerequisite for hand-made images to be perceptible. This fact helps explain the ubiquity of these regularities. Manual paint application techniques strongly favor the production of the same regularities. Aesthetics, then, would seem to play little or no role in human preference for “natural” statistics in art. The perceptibility hypothesis is thus a hypothesis of “natural art.” It is also “natural” in the sense that it is related to art that could be recognized by our earliest fully-human ancestors. In other words, it is specifically related to the problem of depicting the natural world using paint and it ignores complications introduced by artificial lighting, abstract symbolism, etc.

It is suggested here that artists can exploit efficiencies in the visual system and paintings are themselves efficient solutions to what is referred to as the “luminance problem.” The notion of the “efficient artist” can be seen as analogous to the proposed notion of efficient vocal communication (e.g., Levy and Jaeger, 2007). Indeed, art is, if nothing, else a variety of communication.¹¹ As described in previous chapters, there

¹¹ Indeed, this dissertation has implications for the study of many modes of human communication. Since they are generally intended for viewing by conspecifics, artworks certainly constitute a form of communication. And like written language, they are a uniquely human creation. But whereas many species across a number of taxa show vocal, chemical, and physical forms of communication that roughly approximate and occasionally exceed human capacity, no creature matches humans’ sophistication and elaboration in manufactured visual displays. There is much debate in the language literature about whether non-humans show or can learn “recursive” grammars. While it is beyond the scope of this work to comment on this question, there is much evidence that non-human vocal communication shows at least some of the elaboration of human spoken language, e.g., regional dialects in birds, and song in whales. On the other hand, our ape relatives (and indeed all other species) do not possess even the rudiments of the ability to make pictures. Left to their own devices and lacking training, non-human primates create stereotypical fan-shaped scribbles which lack any suggestion of “content” (see Wynn, 1996). In the literature, there is but one anecdotal report of image-making in non-humans: In 1942, Julian Huxley published a report in *Nature* of a gorilla at a zoo tracing his shadow on the wall with his finger. Surely such fleeting, non-permanent evidence is not enough to change my view that while many species possess sophisticated features of human language, none shows an aptitude for picture making. Moreover, unlike vocal communication, artworks leave behind permanent evidence as to their structure. This evidence is far older than that for written language, which some have argued is simply a modification of pictorial communication (see Diamond, 1996). Therefore, to the extent that shared properties of artworks across our history as modern humans reflect evolutionary modifications to our brain, artworks are perhaps more germane to the debate over “human uniqueness” than is language.

exists a basic level of spatial statistical redundancy (predictability) in nearly all art (see also Redies *et al.*, 2007) and such redundancy greatly narrows the space of possible images that can, in the words of Ramachandran and Hirstein (1999), “excite” the visual brain. Moreover, efficient retinal coding of images is largely predicated on the idea that images show predictable spatial frequency spectra. In short, I argue the fact that art in general reproduces natural scene statistics suggests that art is *maximally informative* with respect to visual system coding. This notion is paired with data from a study of preference for artists’ nonlinear luminance compression and a demonstration of how the artist’s gamma (or artist’s look-up table) can be approximated in order to construct a unified hypothesis about human art-making.

By examining and modeling the most basic statistical aspects of art-making, it is possible to put this defining human activity in the context of efficient coding. Making this connection allows us to illuminate the close relationship between human art-making and the fundamental properties of the visual system.

5.1 Background

There are many proposals regarding how and why pictures “work.” Zajonc (1968) wrote the seminal essay in this area, which proposes that humans prefer to look at images that are familiar. Berlyne, on the other hand, found evidence that humans prefer some degree of novelty in their images, which squares with ethological studies showing that some animals seek out novelty even at the risk of injury. Biederman and Vessel (2006) suggest that highly informative scenes and pictures produce opioid compounds in the brain, which contribute to their feeling of pleasure. Winkielman *et al.* (2003) argue that images that are processed most quickly by the visual system are preferred since they are most “fluent.” Some have suggested that pictures that succeed tap into Marr-like visual object models in the brain (Willats, 1997). Others say image-

makers strive to represent the various perceptual constancies (Zeki, 1998). Some have suggested that line drawings, which are seemingly the most simple kind of pictures, are most mysterious of all (Cavanaugh, 1999).

The space of possible images that could be constructed using just 64 trichromatic 8-bit pixels is far greater than the number of particles in the universe. But within this space, images with the statistical properties typical of art occupy a very limited portion of this space. That is, one can exclude the vast majority of possible images from a consideration of art because those images that fall outside this regularity simply aren't produced. It is proposed that this is so because of the efficient coding strategies of the visual system (most especially the proposed goal of response equalization in retinal ganglion cells) and because of the manner in which most art is made. It should be noted at the outset that the observed statistical regularities in paintings do not necessarily limit artistic expression that is relevant to the eye in any significant way. However, the results presented in this dissertation do go some distance in explaining why art through the ages has the general appearance it does.

The overarching goal is to understand art in terms of the most current models of the visual system. In particular, a host of recent studies—which come from neurophysiology, psychophysics, non-Euclidean geometry, information theory, evolutionary biology and other areas—has demonstrated that the visual systems of all sighted creatures are shaped in fundamental ways by the need to operate efficiently when processing the visual environment typical of their habitat (Lythgoe, 1979). The visual system has come to be understood as in many ways a reflection of the predictable statistics of the natural world, both in the system's structure and course of development (see Chapter 1).

Given this information, earlier linear models of visual perception derived from artificial stimuli such as sine-wave gratings are now recognized as being incomplete.

So it is with earlier ideas about art. Many of the ideas about how vision shapes art—e.g., Berlyne (1970), Zajonc (1970), Arnheim (1971), Gombrich (1961)—were conceived of and tested using Gestalt stimuli, which are now recognized as having limited applicability to “natural” images. As will be argued here, models based on efficient coding of predictable natural statistics can now be profitably employed to inform theories of art perception.

5.2 Efficiency of Regular Spatial Statistics: The Perceptibility Hypothesis

Early television researchers and later Field (1987) and Burton and Moorhead (1987) noted that the spatial frequency power spectrum of scenes from the terrestrial earth is predictable. In other words, the proportional contribution of large scale structure and fine detail to image power (in a Fourier sense) can be guessed with reasonable accuracy for any image taken on earth. As described in earlier chapters and in Redies *et al.* (2007), hundreds of works of art sampled from across centuries of history and from dozens of provenances show remarkable consistency in the slope of the Fourier power spectrum, and that this slope is very near to that typical of natural scenes (roughly -2.4).

In a gross statistical sense, the vast majority of artworks on display in museums share spatial statistical redundancies. In particular, the spatial correlation of neighboring points in a painting—or equivalently, the slope of the Fourier spatial frequency amplitude spectrum of the work—can be roughly predicted. In general, amplitude falls as $1/f$ where f is spatial frequency. This is true across cultures, through centuries of art history and for both representational and non-representational works. While such predictability is unlikely to inform higher level perception of art, it could be an example of the deep relationship between the notion of efficient coding in visual system design and the types of patterns humans tend to create with pigment. In this

view, art is something of an uncontrolled experiment in vision, one that offers a translation between the language of the visual world and the language of 2-D human-made images intended to represent some aspect of the visual world. Because there exist examples of visual art from most human cultures, art is therefore quite useful for testing theories of efficient visual coding.

Why should art follow this regularity? One view, which is termed the affect hypothesis, suggests that natural, fractal-like statistics are inherently aesthetically pleasing and therefore artists should in general strive to match such statistics in their paintings (Redies, 2007; Redies *et al.*, 2007; Spehar *et al.*, 2002; Hagerhall *et al.*, 2004). This is an intriguing proposal. Indeed, observers show relatively consistent preference over a range of 1-D fractal scaling exponents for artificial fractal images and for natural scene silhouettes (Aks and Sprott, 1996; Spehar *et al.*, 2002; Sprott, 1993), though it should be noted that other studies have questioned the effect of fractal scaling on preference in urban design (Stamps, 2002).

Redies (2007), in particular, has articulated this hypothesis regarding universal aesthetics of natural statistics with appropriate nuance: “Fractal-like image properties may be necessary, but they are not sufficient to induce [a]esthetic perception.” Redies (2007) has also argued that the bias towards natural scene-like statistics could extend to other affective responses, e.g., disgust or ugliness. That is, images with $1/f$ structure may be judged to be more disgusting or ugly than random images (this notion is currently being tested). Philosophers traditionally see these affective responses as being components of aesthetic response, according to Redies.

Here I propose an alternative view, called the *perceptibility hypothesis*, which states that regularities in art statistics result ostensibly from the need to make an image that is readily visible to the human eye, and one that is amenable to creation by hand. The statistical regularities in art would thus be largely the result of the way pigment is

typically applied to canvas, and of the way those marks are processed by the visual system.

Consider for example a comparison with white noise, which has a flat amplitude spectrum (slope of 0). Almost any collection of hand-made marks introduces spatial correlations, meaning that the resulting image would be likely to deviate from having a flat spectrum. A simple line will produce correlations; a patch of paint will do so to a greater extent. If a single line or paint stroke has the proper fractal structure, it alone can give rise to a $1/f$ spectrum (see Field and Brady, 1997). Thus, art with spatial statistical regularities *unlike* those of natural scenes—e.g., white noise—is rare because of the challenges involved in producing such images by hand, not because such images are necessarily less aesthetically pleasing. The same logic holds for images with a spectrum much steeper than that of natural scenes (slope > 2), as for example in a blurry image (see Figure 5.1).

It is true that artists like Mondrian and Pollock have used application techniques that differ from traditional brush-based techniques and yet produced art whose statistical regularities match those typical of art as a whole.¹² However, any technique that results in markings that resemble an array of locally correlated regions is likely to differ strongly from a technique that chooses tones randomly with respect to space (as in white noise). Despite the fact that random images comprise the great majority of possible images, artists are rarely able create images with such statistics using either their traditional or non-traditional paint application techniques. Therefore, the point is not that artists are forced to reproduce the spatial statistics of natural scenes, but that it is exceedingly hard work to avoid doing so without the assistance of a computer.

¹² Tests on 3 digitized Pollock drip paintings from 1947-1950 [Scholar's Resource, Freeport, ME] found that their amplitude spectra had a mean slope of -1.0 ± 0.05 (standard error); four Mondrian "compositions" [Scholar's Resource, Freeport, ME] had a mean slope of -1.4 ± 0.06 . These data were obtained with the same procedure used elsewhere in this dissertation. The Mondrian images were nearly square, while the Pollock images were cropped to make them square for testing.

Consider Attneave's (1954) classic paper on sensory coding, which testifies to the difficulty involved in creating a random image with a flat amplitude spectrum (i.e., white noise): Attneave employed two junior enlisted airmen to painstakingly fill in a grid of close to 20,000 small squares by hand based on values drawn from a random-number generator. Few human artists have likely gone to the necessary trouble to escape the basic statistical regularities of scenes.

Even if one could easily produce images with "unnatural" spectra by hand, these images would not be perceptible in the way images with natural scene-like spectra are (see Knill *et al.*, 1990), and such images would thus be rarely attempted. Indeed, there is evidence that the human visual system is well-adapted to the range of 2-D fractal dimensions characteristic of natural scenes, which closely matches that for art (Knill, *et al.*, 1990; Field and Brady, 1997; Parraga *et al.*, 2000; Billock, 2000; see also Chapter 2).

The perceptibility hypothesis is also consistent with the results presented in Chapter 1, which show that a response equalization sensitivity model for retinal ganglion cells can explain retinal spatial coding without reliance on decorrelation arguments. Images that do not show natural scene-like amplitude spectra would not produce the relatively flat spectrum predicted by the response equalization model. This coding bias towards images with natural scene-like spectra could contribute to the scarcity of non- $1/f$ art, though the exact mechanism for this process is unknown.

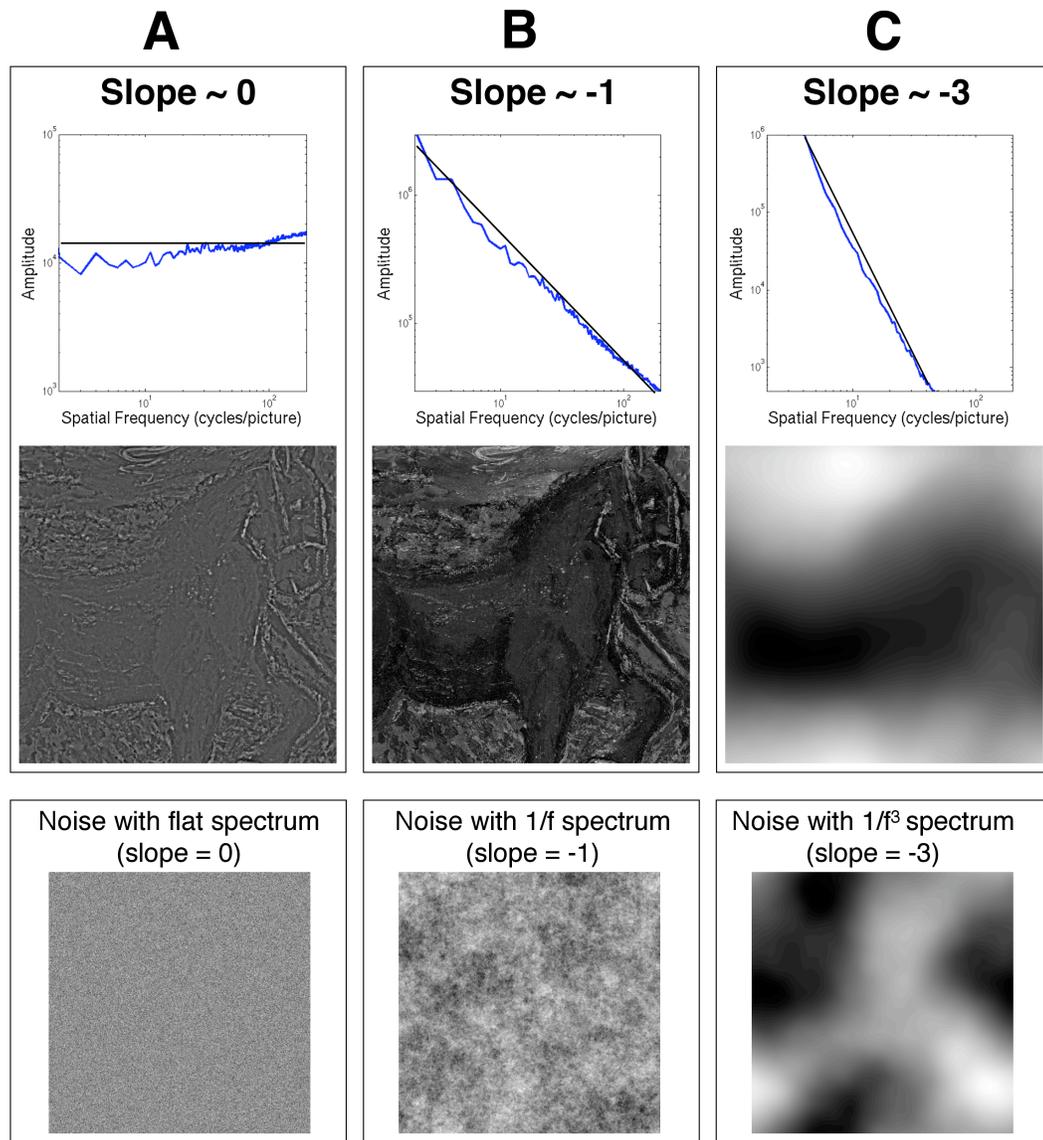


Figure 5.1. Examples of images with spatial frequency amplitude spectrum slope of approximately 0 (A), -1 (B) and 3 (C). The amplitude spectrum for a portion the painting *Untitled (Woman with Red Horse)* is plotted on log-log axes. Below the spectrum is the portion of the image for which the spectrum was calculated. (A) and (C) were generated by adjusting the spectrum of the original image (B) in the frequency domain. Gaussian-distributed noise with amplitude spectra similar to those of the corresponding paintings are also shown. Painting is courtesy of the Herbert F. Johnson Museum of Art, Cornell University: David Burliuk, *Untitled (Woman with Red Horse)*, 1951, Gift of Hans Namuth accession.

Moreover, regularities in amplitude spectrum slope hardly constrain the space of possible images that the affect hypothesis would judge to be aesthetically pleasing, nor does this hypothesis account for artistic styles that do not match the amplitude spectra of natural scenes (e.g., monochromes). In addition, nonlinear luminance compression in paintings fundamentally alters intensity statistics for art compared to the natural world and therefore art is unlikely to "induce a specific resonant state in the visual system...based on the adaptation of the visual system to natural scenes." (Redies, 2007). Nonlinear luminance compression strategies reflect the need to make an image that is perceptible to the eye as depicting the world. Work is under way to explore statistical regularities in these compression strategies across artists.

No studies have yet addressed whether there is a similar range of preference for 2-D fractal dimensions in images. The term "2-D fractal dimension" is used in reference to images whose greyscale intensities vary continuously across two spatial dimensions. The term "1-D fractal dimension" pertains to images comprised entirely of binary boundaries. In other words, the 1-D fractal dimension measures the degree to which a 1-D boundary fills up 2-D space, while the 2-D fractal dimension measures the degree to which a 2-D intensity surface fills up 3-D space. All fractal images tested in Spehar *et al.* (2003) are 1-D by this definition, though it should be noted that the "control" images presented in Figure 7 in that paper were 2-D by this definition. These control images—which are presumed to be derived from white noise—showed no statistical difference in preference.

If preference were found to be limited to a range of 2-D fractal dimensions, this would not by itself invalidate the perceptibility hypothesis, since this hypothesis proposes that art is rarely statistically random because random images are difficult to perceive and to produce. That is, I do not predict a preference for one 2-D fractal dimension over another, except in the sense that the preference will be lower for

images lacking $1/f$ scaling. As has been noted, random images do not show $1/f$ scaling, yet they comprise the great majority of possible images. Visual system coding is matched to natural spatial statistics such that viewers find random images to be equally imperceptible, while images with $1/f$ scaling stand a better chance of being perceived and preferred. I remain agnostic as to the amount of aesthetic pleasure that is to be derived from images possessing a particular fractal dimension—the present study concerns only the high prevalence of art with 2-D amplitude spectra that are well described by $1/f$, not such images' relative aesthetic quality.

Artists have occasionally strayed from the statistical regularities of natural scene amplitude spectra—as well as the limitation on the dynamic range of luminances—and yet produced fine art. Statistical differences in spectra and in pixel intensity distributions for representational and non-representational art, and for Eastern and Western art, are also subject to exceptions.

Regularities in the amplitude spectra are only one of many types of redundancy in the spatial statistics of images. Chandler and Field (2007) have estimated that around 40% of the statistical predictability of natural images is due to regularities in the amplitude spectra. It is unclear at present the extent to which higher order redundancies of natural scenes are reproduced in paintings, nor is it known how these statistics vary across painting type and provenance.

5.3 The Artist's Look-Up Table

Evidence presented in this dissertation suggests that artworks share a number of statistical regularities with natural scenes. As has been argued in previous chapters, the ubiquity of this statistical regularity is most simply explained as a form of efficiency relative to coding strategies employed at early stages of visual system processing, rather than as a feature of innate human aesthetics (e.g., Redies *et al.*, 2007). The

necessity of nonlinear luminance compression in the creation of art was also noted in Chapter 3. In particular, a log nonlinearity was shown to be useful for approximating the way in which artist's solve the "luminance problem," i.e., how they compress the high-dynamic range scenes they depict into the low dynamic range available in paint.

Here I propose that artist's nonlinear luminance compression strategies may themselves constitute an effective solution to the luminance problem. If the goal of an artist were simply to compress scene luminances such that they are reproducible in paint, all artists might do well to apply a log-like nonlinearity. This nonlinearity has been suggested as an efficient solution for compressing scene luminances into neural responses: Cone photoreceptors in vertebrates all show roughly log-like transforms, indicating that this is a common approach to solving the luminance problem faced by most visual systems. As McCann argues (McCann, 2005), the visual system is able to process natural scenes that together span a 10^{10} :1 luminance range though it should be noted that a typical scene (and most scenes depicted in paintings) will show a far smaller dynamic range. McCann estimates that photoreceptors have a 10^8 :1 dynamic range and that ganglion cells can only fire at rates in a roughly 100:1 range. A log transform model of vertebrate cone photoreceptors (when adjusted for adaptation) has been argued to be efficient with respect to regularities in scene luminances and contrasts (Brady and Field, 2000; Mante *et al.*, 2005) and because it turns intensity differences into ratios (Field, 1994). A log nonlinearity is also a good first approximation of the effect of the "artist's gamma" (more accurately called the artist's look-up table) on natural scene skewness. This previous study suggested that a linear scaling is not sufficient to bring about the required compression, but a log function is sufficient (see Figure 5.2). Given that retinal coding strategies are remarkably consistent across vertebrate taxa (Finlay *et al.*, 2004), one might expect artist's to all employ some form of log nonlinearity, or a similar function, in their paintings.

However, one must be careful not to fall for a variety of the El Greco fallacy (see e.g., Anstis, 2002) as described in Chapter 3. That is, an artist must view his own paintings through the cones' compressive nonlinearity, meaning that the image will pass through this nonlinearity twice. Log-like luminance compression in paintings may be useful for many of the same reasons that a log-like nonlinearity is useful at the photoreceptor level but it may not be a result of photoreceptor luminance compression.

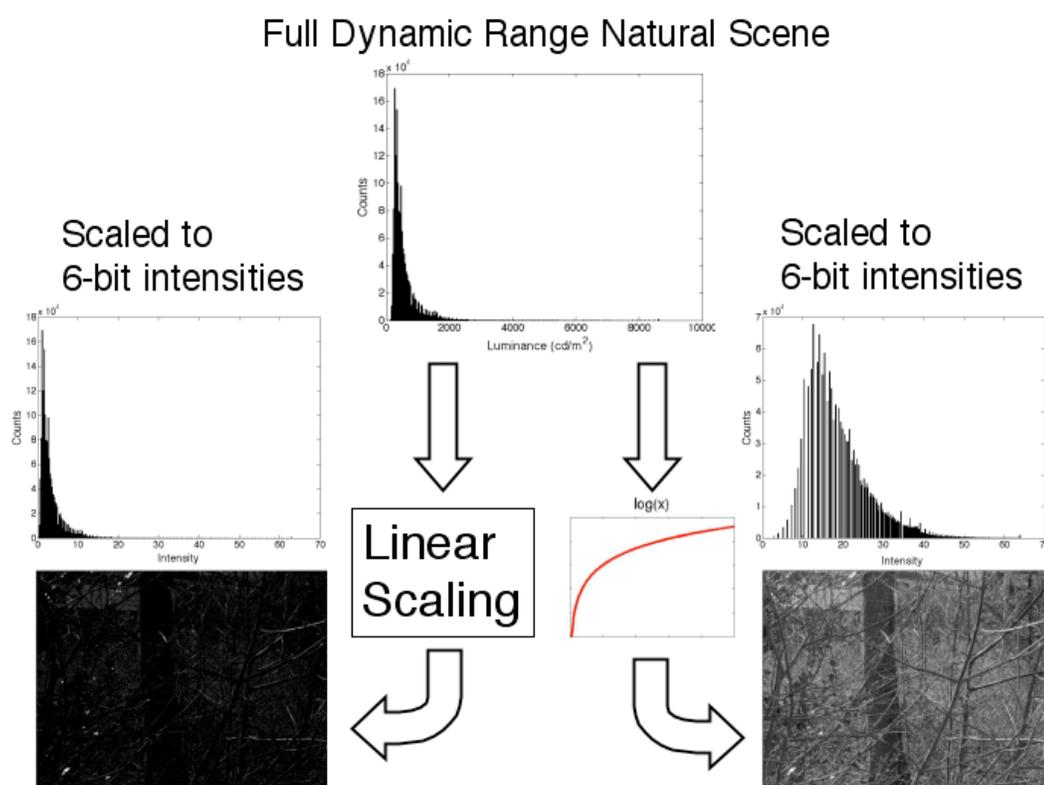


Figure 5.2. This figure illustrates the fact that linear scaling of natural scene luminances produces a very dark scene with a few highlights, but log luminance compression can generate an acceptable scene. The luminance histogram of the original scene (van Hateren #619; van Hateren and van der Schaaf, 1998) is shown at the top. The luminances are linearly scaled to 6 bits of intensity in order to produce the image and histogram on the left, while the luminances are scaled with a log nonlinearity, then stretched to 6-bits to generate the image and histogram on the right.

Clearly the goal of the artist is not simply to compress the dynamic range of luminances. The study described in this Chapter tests whether human artists use a consistent set of strategies to compress scene luminances, and it compares these compressive transforms to a log in terms of observer preference. If a single transform were sufficient to compress images in the way artists do, one would expect these transforms all to be log-like and on average, there should be little or no difference in observer preference for the collection of natural scenes when they are compressed according to these transforms. Though a log nonlinearity would appear to be the most efficient method of compression, I show that it cannot be used to model the variability in the artist's look-up table. I propose that like the regularity of power spectra for art, the simple model of the artist's look-up table describes a minimal constraint on art making, but a fundamental one. Both of these statistical constraints have underlain the artist's task since the inception of art, though most artists were and remain unaware of them. I speculate that the type of scene depicted by the artist or perhaps the manner in which the artist learned to paint may dictate the use of a distinct compression strategy and discuss other factors that may affect the shape of this transform. The fact that photographers have long known that it is necessary to *locally* adjust intensities in an image (see Discussion) suggests that one could even assume that there is no single strategy common to all artists. However, this question has not been directly addressed in the literature using calibrated natural scenes so it warrants the more rigorous attention it is given here.

In the first study, I provide a model of the artist's look-up table for an accomplished landscape painter. In the second study, I present evidence that observers show consistent preferences for compressive transforms of natural scenes images that match scene histograms to the mean histogram of various collections of art, and observers prefer many of these transforms over a log transform.

5.3.1 Study 1: Model of the Artist's Look-up Table

An example of how the artist's look-up table can be modeled is presented. There remain technical limitations of the current approach with respect to very high dynamic range scenes, like landscapes containing significant amounts of sky. This means that measurements of the artist's look-up table are currently most profitable for scenes with dynamic ranges that are significantly smaller than that for typical landscapes (but which are still much larger than the range available in paint). I chose therefore to analyze a painting by a professional local painter with many years of experience, Neil Berger (<http://neilberger.com>). The painting depicts a scene of the bottom of a waterfall (containing no sky), which it was possible to recreate in a photograph.

The task of finding the artist's look-up table involved mapping luminances from the scene that inspired a painting onto the luminances of the resulting painting. This measurement is most straightforward for outdoor scenes. Of course, few paintings are "literal" reproductions of the spatial layout of the scene, and as I have argued, linear reproductions of natural scene luminances are impossible to make using paint. The artist whom I have studied (Neil Berger) generally makes quick sketches of the outdoor scene he aims to capture but does the bulk of the painting in his studio. As can be seen in Figs. 5.3 and 5.4, however, this method leads to an image that is quite similar to the scene that inspired it.

The pixels in images of the scene and the painting were photometrically calibrated. To do this, a Macbeth Color Checker (Gretag-Macbeth LLC, New Windsor, NY USA) was placed in a stationary location within the scene and its intensities were recorded as raw pixel values from the camera's CCD chip, and as luminance values measured with a Minolta LS-100 photometer (Konica Minolta Inc., Tokyo Japan). Light reflected from the color checker's six greyscale reflectance patches was measured for each patch. The relationship between reflected luminances and their corresponding pixel

values can then be established and it can be extrapolated to other values by fitting the relation with a polynomial or exponential function. However, this procedure will not generally allow calibration for typical high dynamic range scenes. Debevec and Malik (1997) offer a discussion of this problem and one solution, which is to take pixel intensity samples using different shutter speeds and the same aperture.

The painting titled “Taughannock Falls” by Neil Berger (2006) was selected for study for reasons described above. It was photographed in the artist’s studio lit by a bank of north-facing windows at 14:00 on a partly cloudy day in summer. It was imaged along with a Macbeth Color Checker using a tripod-mounted Canon PowerShot S60 (Canon Inc., Tokyo Japan) digital camera. The full image was 2592 x 1944 pixels and cropped to 916 x 860 (exposed for 0.1 sec at f/5.8, focal length 8.6 mm) and recorded at 16-bits per channel in raw format. It was converted from RGB coordinates to YIQ coordinates and the resulting map of greyscale intensities (Y) was retained. When fit with a polynomial, the relationship of pixel value (x) and luminance (L) was found to go as $L = 0.0035x - 0.15$ in units of candelas/(meter)², with $R^2 = 0.99$ for the fit. The image is thus converted to luminance values. See Figure 5.3 for the luminance map and a histogram of the luminance values. The painting showed a dynamic range of 20.9:1.

The same procedure was then performed with the scene that the painting depicts, which is located in Taughannock State Park, Trumansburg, NY USA. The scene was photographed using the same apparatus as that used above on an overcast day at 15:00 in autumn, a day much like that depicted. A second image was acquired using the same exposure and focal length (shutter speed: 0.025 sec, aperture: f/6.3, focal length: 20.7 mm) with the Macbeth chart placed 1.5 m from the camera, which was used for calibration. The luminance map and luminance histogram are given in Figure 5.4 below. Using this calibration, the scene is found to have a dynamic range of

luminances of 41.3:1, which is roughly twice that for the painting. These data were fit according to $L = 0.0038x^{1.11}$ where x is pixel value and L is luminance in units of cd/m^2 , with $R^2 = 0.99$ for the fit. Note that the camera saturated on the fifth most reflective chip (reflectance = 0.64), meaning that the pure white chip (reflectance = 0.95) had the same pixel value. However, the scene itself did not have surfaces that reflected light at this intensity. I therefore calibrated the scene luminances based on the first five data points, which gave good agreement with spot photometric measurements taken at the base of the falls and on the surface of the pool.

The method for calculating the look-up table for this image pair involves transforming the histogram of the scene to match that of the painting using histogram matching (also called histogram specification; see Gonzalez *et al.*, 2004). Given the scene luminance map A , one can minimize for a transform T using the equation $|c_1(T(k)) - c_0(k)|$, where c_0 is the cumulative histogram of A , c_1 is the cumulative sum of a specified histogram (i.e., of the corresponding painting) for all intensities k . This minimization is subject to the following constraints: (1.) T must be monotonic and (2) $c_1(T(a))$ cannot overshoot $c_0(a)$ by more than half the distance between the histogram counts at a . This transformation is used to map the gray levels in the scene A to their new values in the image B , with intensities $b = T(a)$. After performing histogram matching on the scene (using the luminance histogram of the painting), the “artified” scene appeared recognizable as a waterfall scene (Figure 5.5). It should be stressed that the images shown here are luminance maps so in order to reproduce them on paper or on a low dynamic range display, linear intensity scaling was applied to each image. The point of this demonstration was not to make the “best” image or tone mapping, as is the case for most computer graphics applications. Rather it was to show that a greatly reduced model of the painter’s strategy can by itself achieve the

nonlinear luminance compression, and that this transform is unlike a simple log function. The transform for the Taughannock Falls scene is shown in Figure 5.6.

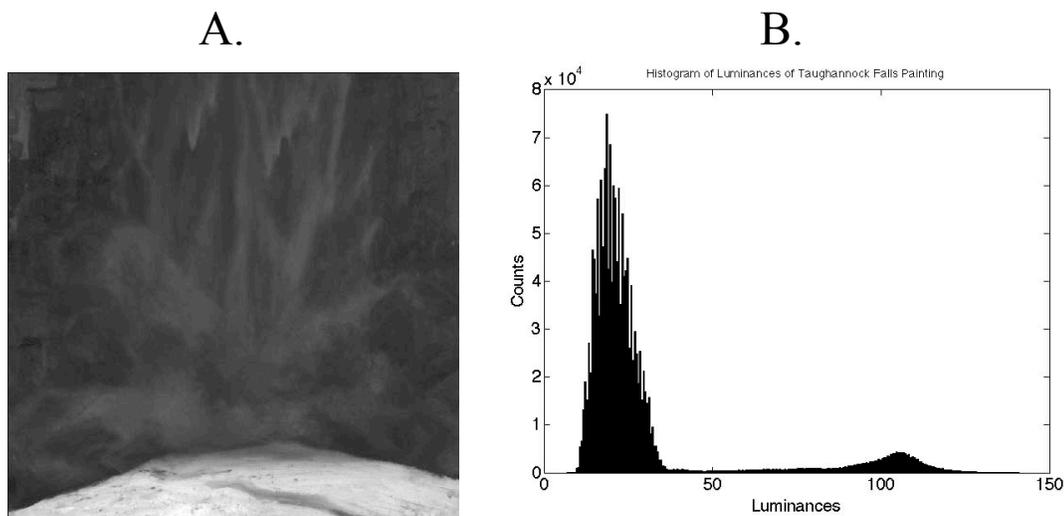


Figure 5.3. Luminance calibrated image of painting “Taughannock Falls” by Neil Berger (A.; linearly scaled for display) and luminance histogram of painting (B.). The painting shows a dynamic range of 20.94:1.

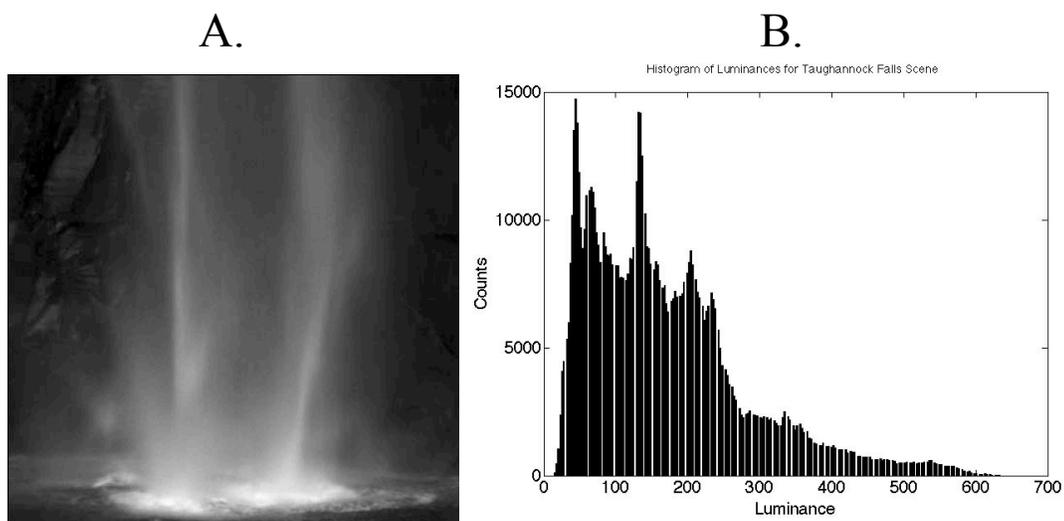


Figure 5.4. Luminance calibrated scene (A.; scaled linearly for display) and luminance histogram of scene (B.). The scene shows a dynamic range of 41.25:1.

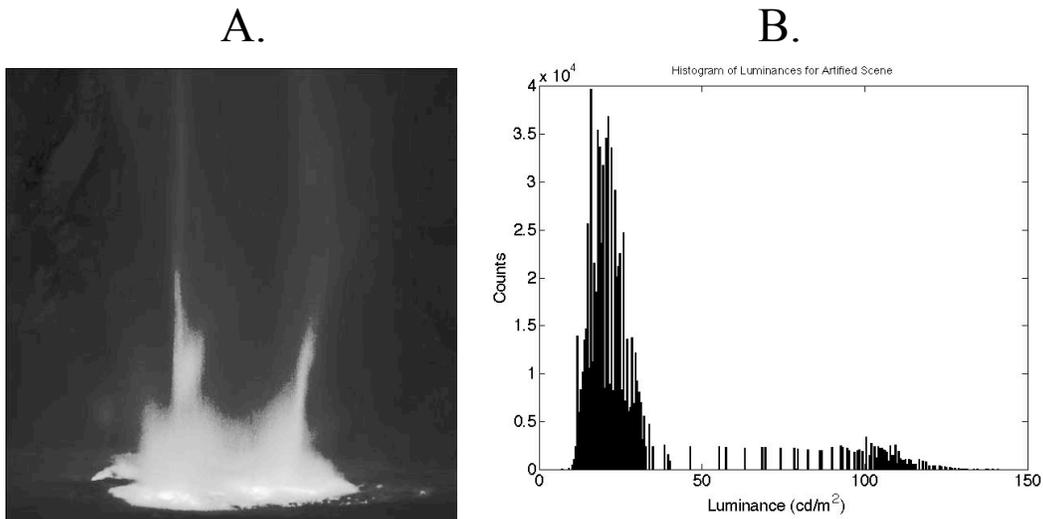


Figure 5.5. “Artified” scene (A.; scaled linearly for display) and luminance histogram of scene (B.).

I speculate that another artist painting the same scene could produce a distinctly different image by simply changing the number of inflection points in the transform. Note that unlike a log function, whose second derivative does not change sign, the second derivative of the transform for “Taughannock Falls,” T , changes sign three times. Most simple functional compressive nonlinearities have the same limitation as does the log. Clearly, this artist does not employ a single, simple functional look-up table for all his paintings.

Of course, the artist’s look-up table model has many limitations, not the least of which is the obvious quantization in the mid-range tones. As described in the discussion, the transformed scene has been boosted in the mid-range intensities because of the nature of the histogram matching technique.

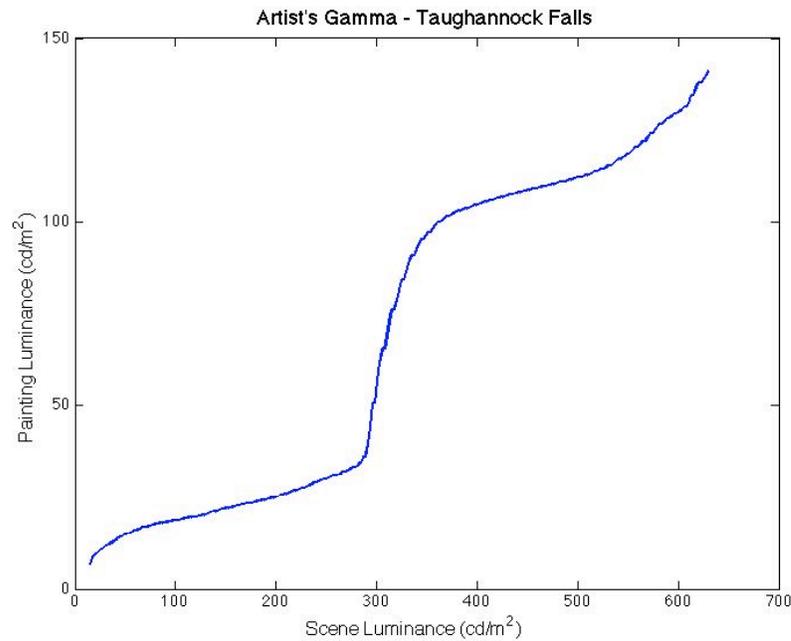


Figure 5.6. Luminance transform T mapping Taughannock Falls scene luminance histogram into “Taughannock Falls” painting luminances in units of candelas/meter².

In the next study, I performed an experiment to compare different mappings of natural scene images such that their histograms matched those of a number of groupings of artworks from a major university collection. I also tested observer preference for these transformed scene images, and I compared these to scenes that had been compressed using a log luminance nonlinearity. The goal of this study was to test the notion that observers prefer natural scene images to have a particular histogram, and to ask whether a simple log transform of the scene is preferred over histogram-specified scenes.

5.3.2 Study 2: Preference for Artist’s Transforms

Since one does not generally have access to the scenes that inspired most paintings held by major art museums, it is difficult at present to directly calculate an artist’s look-up table for most artists. However, the question of what type of transform most

people prefer can be addressed. Here I tested whether observers prefer images with a certain histogram, regardless of the image's subject matter. I calculated compressive nonlinearities that transformed a given natural scene image so that its histogram matched the mean histogram of a variety of collections of art from a major university museum. Then I asked observers to rank each of these scenes, along with a log-transformed version of the scene, in order of preference. The natural scenes that were used in this study were 31 scenes from the van Hateren database (van Hateren and van der Schaaf, 1998). The transform for each image and class of artworks was again calculated using histogram matching as described in Study 1. The sets of artworks whose mean histograms were measured included:

- a collection of Hudson River School painters (10 images)
- a collection of Eastern art from the Herbert F. Johnson Museum of Art, Cornell University (72 images)
- a collection of Western art from the Herbert F. Johnson Museum (68 images)
- the set of images chosen as “abstract” works in a previous study² (12 images)
- the entire collection of Eastern and Western works, not including Hudson River School images (140 images).

Three Western observers (1F), who were naïve to the purpose of the experiment, were asked to rank the six transforms of each image in order of preference. The 31 sets of 6 images were shown in three rows and two columns on an Apple 30-inch Cinema Display (Apple Inc., Cupertino, CA USA) at approximately 0.4 m from the observer. The placement of the images was randomized for each scene. See Figure 5.7 for an example.

For the observers tested, it was found that scene images transformed to match the mean histograms of the entire collection (all art), the set of Hudson River School paintings, and the set of Western painting were preferred over the log-transformed images, and over those transformed to match the mean histogram of the abstract works and of the Eastern works. Table 5.1 shows the total number of top 3 votes and bottom 3 votes for each transform, and their difference, as well as corresponding percentages. Figure 5.8 displays a histogram of the rankings by transform.

First, note that there is no single histogram transform that can be used to predict all viewer preference data measured here. However, there remains a consistent preference for certain transforms. Observers prefer scenes that have been transformed so that their histograms match the mean histogram of a number of the collections of art considered here. In other words, viewers are not indifferent to the range of transforms that allow a given image to be matched to the histogram of a variety of artworks. Moreover, many of these transforms produce images that are preferred over images compressed with the more parsimonious strategy, i.e., a log nonlinearity. The flexibility afforded artists in this respect allows for the production of distinctive images drawn from a diversity of scenes, which could help explain the preferences observed. It is unclear at present why the transforms for Eastern and abstract art were not preferred over the log transform.

Of course, the method for calculating these transforms is not strictly a model of the luminance compression performed by painters (see Discussion). Moreover, definitive tests of the relationship between distinguishability and liking for natural scenes have yet to be performed. However, one can conclude that no single solution can characterize the strategies taken by broad groupings of artists. There may thus be good reason in future to measure consistency among and across artist's look-up tables, since

these transforms may be shaped by a number of factors. In the following section, some of these factors are discussed.

Table 5.1. Preference data for natural scenes transformed so that each image histogram matched the mean histogram for each grouping of art, and they were also transformed with a log function. From left, the table lists the number of images from each class ranked in the top half (number of top 3 votes) and in the bottom half (bottom 3 rankings) in terms of preference, summed over observers. There were a total of 93 rankings (3 observers x 31 images) for each transform. Corresponding percentages are also shown. The rightmost column shows the difference between top half rankings and bottom half rankings.

Transform	Rankings in Top Half		Rankings in Bottom Half		Difference
	<u>Number</u>	<u>Percent of total</u>	<u>Number</u>	<u>Percent of total</u>	<u>Number</u>
All Art	76	82%	17	18%	59
Hudson River School	67	72%	26	28%	41
Western Art	41	44%	52	56%	-11
Log Transform	36	61%	57	39%	-21
Eastern Art	31	33%	62	67%	-31
Abstract Art	28	30%	65	70%	-37

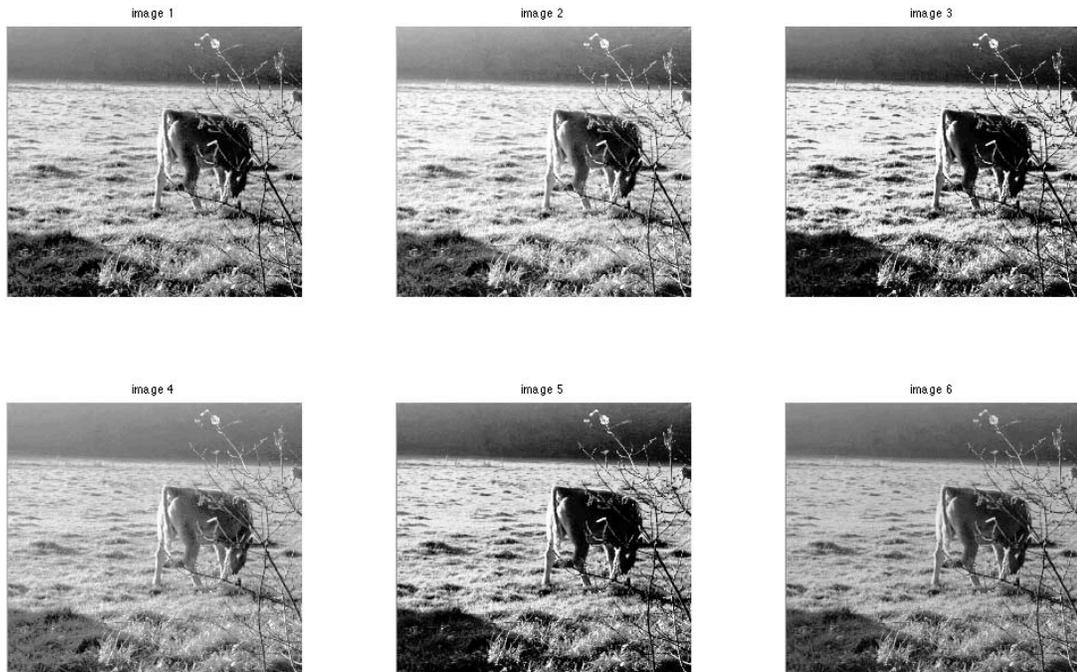


Figure 5.7. Example of display viewed by observers for scene number 9 in the test set transformed according to the following histograms (left to right from top): all art, log, abstract art, Eastern art, Western art, Hudson River School. Placement of transforms was randomized for each scene.

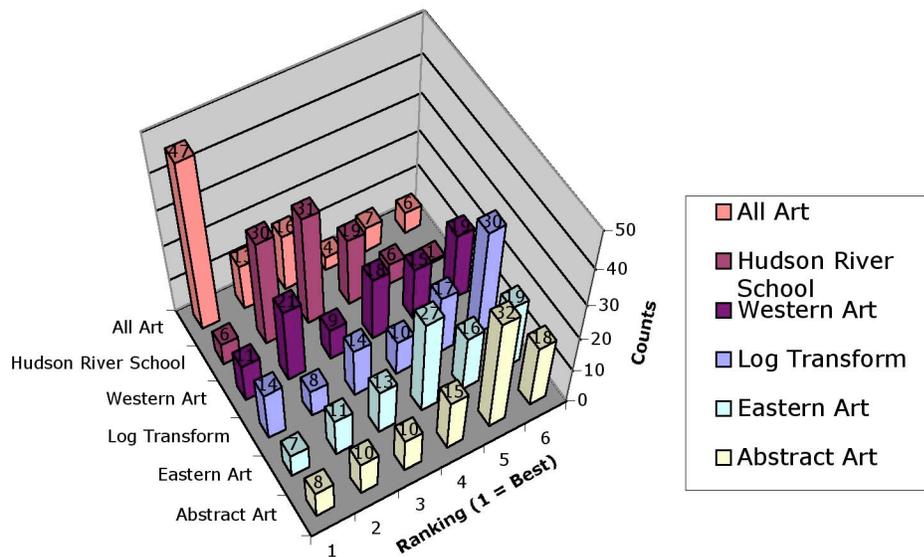


Figure 5.8. Histogram of preference rankings by transform summed over 3 observers. There were 93 total rankings for each transform class (3 observers x 31 images), and each ranking was used a total of 93 times. Note that the transform classes All Art, Hudson River School and Western Art were preferred over the log transform.

5.3.3 Discussion

Like the regularity in power spectra in art shown previously, the artist's look-up table is a minimum necessary constraint for nearly all artists. The demonstration here of a method for approximating the artist's look-up table does not prove that the artist examined here produced the most efficient compression, nor that there do not exist a host of equally compelling images that could be generated for this scene with different transforms. But together with the results regarding observer preference for histogram-specified scenes, this work indicates that no single function or fit parameter is able to account for artist's nonlinear luminance compression strategies. Different groupings of artworks produce different transforms, many of which are preferred by observers over a log transform. The artist's look-up table model therefore captures a fundamental aspect of art making, one that may be characteristic to a given artist, artistic movement, time period or scene type (i.e., landscapes, portraits, etc.). I am currently exploring this range of possible factors.

Why might one expect artists to have discovered effective nonlinear transforms to solve the luminance problem? A number of caveats should be added to this notion. This hypothesis is most easily applied to landscape paintings since it is often possible to find a rough input-output relation in terms of luminance for paintings of specific landscapes. But landscapes as such became common only in recent times. Portraying the natural environment for its inherent beauty using paint is a relatively new innovation in the history of Western art (Gombrich, 1961), though I would argue that this is for cultural, not technological, reasons.¹³ On the other hand, all painters face some form of the luminance problem.

In addition, the notion of luminance compression assumes that the artist has the ability and the intention to faithfully map luminances in the world into reflectances of

¹³ Indeed, Korean painters have been producing such images for over a thousand years (Yi, 2006).

pigment in the same relative spatial arrangement. This is not always the artist's goal. Luminance compression may be facilitated by tricks such as a knowledge of atmospheric perspective and the use of grids and camera obscurae or other optical projections, but it does not require such short-cuts. It may even be possible to judge whether an artist used a specific short cut by calculating a collection of artist's look-up table functions under experimental conditions.

There are other issues with the idea of an artist's look-up table, which relate to the limitations of the technique of histogram matching. Because histogram matching is imperfect by necessity (Gonzalez *et al.*, 2004), it is impossible to fully model luminance compression using this technique. Histogram matching essentially interpolates between the cumulative distribution functions of the input and the output, and the resulting transform (which is applied to the input) is thus required to be monotonically increasing. As can be seen in the example given above, the bimodal painting luminance distribution (Figure 5.3) is approximated as a unimodal one in the "artified" scene (Figure 5.5). Histogram-based methods of this sort (e.g., histogram equalization) all suffer from these limitations when they correspond to images because one cannot choose some fraction of pixels at a given intensity to transform while leaving the rest at that intensity unchanged.

Artists—particularly photographers—have been aware of the limitations of global luminance compression for some time. Ansel Adams' "zone system" of camera exposure is in a sense a practical model for solving the luminance problem using *local* adjustments to exposure. The zone system is designed to make use of the ability of the photographer to locally over- or underexpose regions of an image in the print-making process (Adams, 1948). The general rule of this system is to "expose for the shadows; develop for the highlights," thereby maximizing the effective dynamic range of the output. That is, the response properties of camera film are such that the global effect of

an overexposed negative (which captures a larger dynamic range in the shadows) can be compensated by locally adjusting the exposure of the print (processes referred to as “burning and dodging”).¹⁴ Similar techniques can be applied digitally. The development of the zone system for the purpose of capturing diverse outdoor scenes suggests that a flexible strategy based on the spatial distribution of scene luminances can be effective for diverse collections of scenes.

McCann (2007) has argued that until the Renaissance, all human art was incapable of depicting high-dynamic range scenes with any degree of fidelity. As has been mentioned, Western landscape paintings as we know them today were unknown in the Pre-Renaissance and rare indeed until the 19th century, so evidence for McCann’s argument is typically drawn from candle-lit scenes. Leonardo da Vinci is thought to be the inventor of the technique of *chiaroscuro* (literally, “light-dark”) but Rembrandt, Caravaggio and others perfected it in order to portray high dynamic range indoor scenes. This is another example of a trick that artists can use to nonlinearly compress scene luminances. In the case of candle-lit scenes, *chiaroscuro* is a technique that may succeed in part because of the relative ease with which the artist can model the light produced by a single candle. Compared to a candle-lit scene, a typical outdoor or window-lit scene—with its multiple sources of direct, diffuse and specular illumination—may be more difficult to recreate.

Painters are given more flexibility than photographers in that every area of a painting can be locally “burned and dodged,” so to speak. Indeed, local adjustments to luminance and contrast play a large role in the visual effect of a painting. The artist’s look-up table notion is a significantly simplified model, which does not capture the

¹⁴ This can be aided by changing the amount of time a negative spends in chemical developer, and by using toner chemicals during development.

rich landscape of local adjustments. Nor does it account in its present guise for the use of color, which is a major concern of painters.

5.4 Conclusion

Although it has been shown that viewers have consistent preferences for images whose histograms match the histograms of various collections of art, the results presented here suggest that artists do not have a single, universal strategy for compressing real-world luminances into painting luminances. And despite the limitations of the artist's look-up table model presented here, the evidence gathered suggests that the artist's look-up table is a useful approximation of an effective approach used by artists to solve the luminance problem. It is not a full model of the painting process, nor of luminance compression in art in particular, but it is a simple description of an artist's unique strategy for compressing the dynamic range of luminances. As such it may be a defining characteristic of an artist and it could potentially be used to determine the "stylometry" for a given artist. This can be done if (1) an artist is consistent in her look-up table over time, or has predictable changes over time; and (2) a reasonably representative, calibrated image of the scene depicted in a painting can be generated. Both of these conditions require much further study. Lacking further evidence, I nonetheless speculate that reproducing the depicted scene may not be necessary if the artist's look-up table is found to be of a *fundamental* nature with respect to the artist's visual system. That is, an artist's typical look-up table may be shaped simply by the ambient light in the area where the artist learned to paint or was raised. A more extreme suggestion comes from Charles Falco (2007) who suggests that optical projection devices developed in the Renaissance produced images of a scene that were so much like what we think of as "paintings" that the mere exposure of an artist to that image would have changed all of the artist's subsequent

paintings. In other words, merely witnessing a projection would, in Falco's view, allow painters to more accurately depict many aspects of natural scenes, including their large dynamic range. While I do not endorse this view, it does show that the luminance problem has been attacked in various ways for hundreds of years (McCann, 2007). Moreover, a full model of artists' luminance compression strategies could be a useful test of Falco's hypothesis.¹⁵ It is also possible that the shape of the artist's transform may be related to the number and intensity of illuminants in the depicted scene, or to the transforms typical of a painter's contemporaries.

The statistical measures demonstrated in this dissertation, as well as those from other groups¹⁹, can also be useful for digital art authentication. The refinement of such techniques is ongoing. The ability to establish such statistical signatures based on non-invasive photometric measurements could prove extremely useful to museums, collectors and critics. Such has been done with some success for literature (Mosteller and Wallace, 1964).

¹⁵ However, it should be noted that many projections—when viewed in a dark room—will show the same large dynamic range as the scene itself, which makes it impossible to “copy” the scene using paint. Falco (2007) suggests, on the other hand, that artists would only have used the projection to sketch where the defining shadows are in a scene and, as with landscape painters like Neil Berger, the bulk of the composition would be done later in the studio.

CHAPTER 6

GENERAL DISCUSSION AND FUTURE DIRECTIONS: PERCEPTIBILITY, FAMILIARITY AND PICTURE LIKING

6.1 General Discussion

As has been shown earlier in this dissertation, there exist measurable statistical “signatures” for collections of art: Various samples of human art works have been shown to vary systematically across a number of statistical measures that are relevant to the efficient coding of statistically regular natural scenes. In particular, natural scenes and artworks were found to show similar though significantly different spatial frequency amplitude spectra. The amplitude spectrum describes the relative degree of large scale (low spatial frequency) structure content and fine detail (high frequency). This regularity held for abstract as well as representational art. Similar findings by other researchers (Redies *et al.*, 2007) have led some to propose that this regularity is due to the most part to humans’ innate aesthetic sensibility (Redies, 2007). One major goal of this dissertation is to elaborate an alternative proposal, namely the perceptibility hypothesis. This hypothesis argues that the regularity of basic spatial statistics in art is ultimately a consequence of visual system design. Starting at the retina, for which a model of coding efficiency is presented in Chapter 2, the visual system has been shaped over the course of evolution such that it is highly attuned to images with natural scene-like statistics. Moreover, manual pigment application techniques (painting, dripping, etc.) strongly favor images with these same statistics.

We have also considered basic statistical moments of intensity distributions for large collections of art. It has been shown that works from the Eastern hemisphere show significantly more negatively skewed intensity distributions than works from the

Western hemisphere, a difference attributed to systematic differences in background tone. Eastern works also show a higher proportion of fine detail compared to Western works. Abstract works have been found to show a higher proportion of fine detail compared to representational works in our sample, though abstract works follow the general regularities of representational works and natural scenes.

Figure 6.1 summarizes the effect of changing these statistics for a $1/f$ noise image. While the original noise image has a spatial frequency amplitude spectrum slope of -1.0 , adjusting the slope of this image to be -0.5 and -1.5 results in the images that appear in the top right. The effect of moving the original image's intensity distribution skewness (roughly zero) to the left (resulting in a skewness of roughly -1.0) and to right (skewness of roughly $+1.0$) is shown on the bottom right of the figure. Note that the noise image shows roughly the same amplitude spectrum slope as a typical natural scene, it lacks much of the structure of scenes, especially in the form of continuous edges. However, in terms of these basic statistics, most art images from across art history show amplitude spectrum slopes and intensity distribution skewnesses that fall between the examples shown.

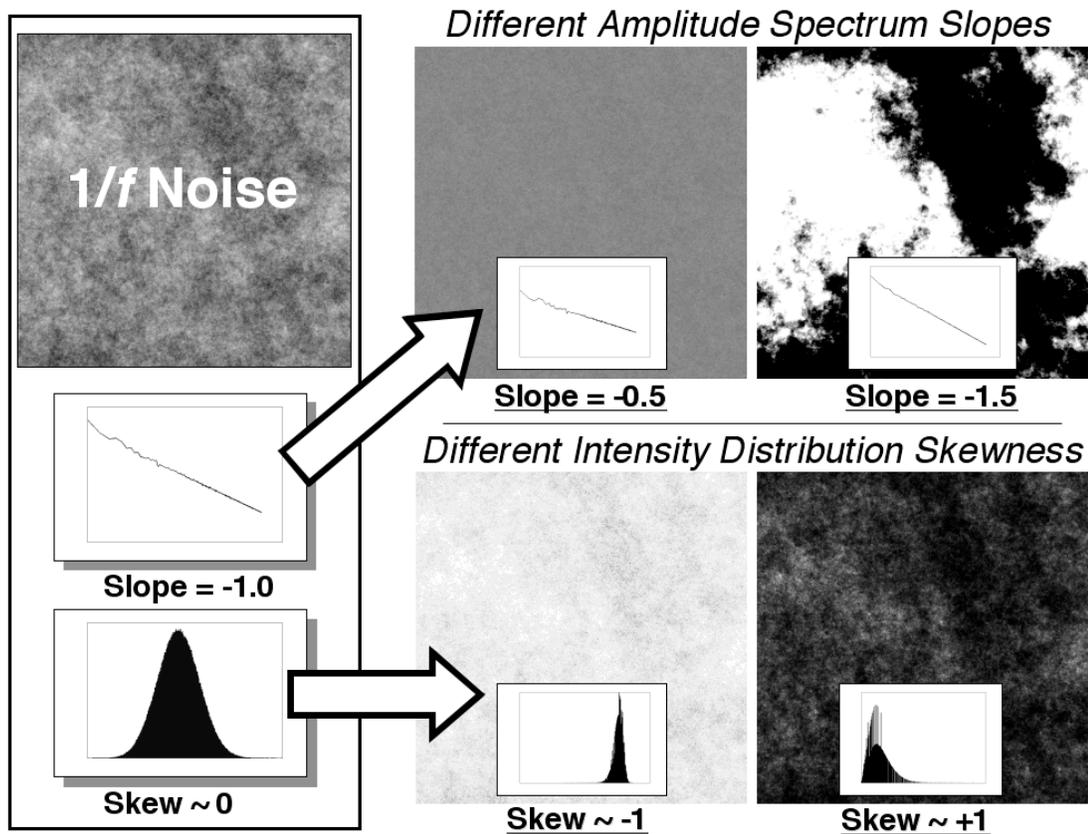


Figure 6.1. The image at left is random noise whose amplitude spectrum falls as $1/f$ (slope of -1) and whose intensity distribution is Gaussian (skew of zero). Both of these statistics can be independently varied, such that the image can have an amplitude spectrum that falls as $1/f^{0.5}$ and $1/f^{1.5}$ respectively (top right) or it can have a skew of roughly -1 and $+1$, respectively (bottom right).

It was also demonstrated that artists must perform nonlinear luminance compression in order to represent the visual world, since the range of light intensities in the world far exceeds what it is possible to create using paints. A model was deployed of the “artist’s look-up table,” based on a photometric analysis of one painter’s image of a waterfall. A study of observer preference showed that no single function can describe the transform used by different classes of artists, and that a simple log compression is less preferred than transforms based on distribution statistics from different groupings of Western art.

The statistical measures presented in this dissertation, as well as those demonstrated by other groups (Lyu *et al.*, 2004) may be useful for digital art authentication, and the refinement of such techniques is ongoing. The ability to establish such statistical signatures—or stylometries—based on non-invasive photometric measurements could prove to have great value for museums, collectors, art historians and critics. This suggests that statistical measures are useful not simply for understanding commonalities in human artistic production and liking, but also for understanding their differences. These measures may be employed for other practical applications like image search as well.

6.2 Future Directions: Perceptibility and Picture Liking

In the remainder of this dissertation, I will consider another avenue of study, which brings together the findings on statistical regularities of artworks with classic studies of viewer preference for pictures. The perceptibility hypothesis is joined with findings about familiarity and novelty in order to construct a new framework for analyzing gross trends in viewer preference for images.

To briefly rehearse the perceptibility hypothesis, consider that though the realm of pictures constitutes an unfathomably large array of possible images, only a small portion of this realm is occupied by images one might consider art. This smaller, though also quite large, space of images that are candidates to be art shares the basic statistical regularities of natural scenes. Images that fall outside of this space still have a chance of being called art but because they are on average perceptually identical to most other such images, the chances of this are slim.¹⁶

¹⁶ There is an intriguing parallel with another framework for understanding how the brain shapes the appearance of art through the ages. Zeki, in one of his most eloquent papers on art (1998), argues that artists seek to capture the invariant visual qualities of the world, in much the same way that the brain seeks to isolate objects and their motions from an ever changing stream of retinal images. These “visual qualities” are similar to Plato’s idea of the ideal object, as Zeki describes. Though both artistic

Once we've narrowed our attention to the space within which art typically lives, we can apply more conventional notions of picture liking derived from studies of familiarity and novelty (Zajonc, 1970 Berlyne, 1970). However, as with statistical regularity, psychophysically-measured viewer preference or response (what we call picture-liking) is a metric that often wrongly invokes aesthetics. Unlike the notion of a universal aesthetic sensibility, we propose that picture liking is predictable given psychological findings related to mere exposure, novelty and distinguishability. But prediction is only possible given a detailed understanding of the observer's visual history. Thus, this framework unites elements of the universal aesthetic argument (though stated in terms of the perceptibility hypothesis), in that perceptual response for works with certain statistical properties is roughly predictable for all observers; and elements of the contextualist standpoint, since only knowing the viewer's visual history allows further predictability. However, such knowledge is in general impractical to obtain.

Since it is difficult to control or document what a person views for long periods of time (see e.g., Hochberg and Brooks, 1962), in this chapter I offer a more general hypothesis that describes basic relationships between how familiar a person is with an image and how likely it is that she will like it. This proposition requires a cleaving of the notion of picture liking from the larger study of aesthetics, and it also requires a provisional delineation between style and content. In a nutshell, I propose that the number of images a person sees in her lifetime affects aspects of the artistic production of contemporary painters, which in turn reinforces the existing preference for the current style and content. In particular, relatively small changes in style allow perpetuation of the accepted subject matter. Examples are given from Western art of

representation of objects and brain representations are not fully understood, the perceptibility hypothesis acts as a sort of bookend to Zeki's theory: whereas Zeki describes constancy in terms of object-level perception, the perceptibility hypothesis concerns low-level statistical constancy.

the middle ages and from landscape painting of the late Qing dynasty in China. One consequence of this paradigm is that some new, more eccentric types of content can open up large tracts of “unexplored space,” which in turn begins the process of stylistic elaboration. That is, there is the chance for large, unforeseen, fruitful innovation in terms of content (often the result of cultural mores being loosened) and a subsequent explosion in modes of representation for that content. However, innovation in low picture-density societies is mainly propelled by relatively small changes in style, and when such innovation is extinguished, entire artistic traditions can die off.

In modern times, on the other hand, a person in an industrialized country is likely to see dozens if not hundreds of images in a day, which I argue shifts the emphasis of image makers to innovations in content, rather than style. Paradigm shifts can still be effected by producing stylistic innovations (and by new representational technologies), but in general, image makers can simply introduce new content in order to keep observers “happy.”

Here I trace the logic for these conclusions and offer a number of outstanding questions to be addressed by this course of research. The theoretical framework developed for understanding statistical regularities in art is relevant to many areas of art study, including aesthetics, art history, archaeology, psychology and economics.

6.2.1 Familiarity and Novelty

Psychologists of the last century made a number of important discoveries related to picture liking, though it should be noted that they often couched their findings in terms of aesthetics. There are two important ideas that grew out of this work, which are largely due to Zajonc (1968, 1970, 1980) and Berlyne (1970), respectively. Zajonc, over decades of research, demonstrated what Cutting (2003) calls “a nonconscious acquisition of information about, and attitudes toward, objects and events through their

repeated presence in our lives.” This so-called familiarity effect can be seen as a form of predictability in the response of a viewer to an image. In particular, viewers are consistently more likely to say they prefer an image they have seen earlier in the course of psychophysical testing, or even a prototype that resembles the stimulus under consideration. This effect is observed even in very brief experimental presentations, which is referred to as the mere exposure effect.

The second course of research that is relevant here is related to variety and complexity in repeated visual presentations. Berlyne showed that for many classes of visual stimuli (e.g., simple shapes, drawings of animals, patterns of lines, textures and random distributions), viewers predictably prefer irregular arrangements, irregular shapes, changes in the number of objects or material, heterogeneous elements and incongruous juxtapositions over symmetrical, regular or predictable ones. This finding is related in Berlyne’s framework to the preference of many animals for new and different smells, routes, visual investigations and other stimulation over monotony, even at the risk to themselves of injury or death. At the level of society, it may operate on long time scales as well and serve as a force guiding the evolution of art historical canons.

It is obvious that these two effects can counteract one another: if you’ve seen something before, you are likely to say you like it, but if you’ve seen it or something very much like it too many times before, you will stop liking it. In any practical experiment with fine art, this trade-off is difficult to predict.

I propose that this paradigm can be extended to model viewer response when the purpose is to gauge preference and when the relevant viewing history of the observer is known. Moreover, the ubiquity of fast computers and the development of high resolution computer monitors and high dynamic range displays mean that psychologists can gather a great deal more data in this vein. This is especially

important when held in comparison to the tachystoscope experimental set-ups used in the original work of Zajonc, Berlyne and others. In sum, this neo-contextualist argument requires a detailed knowledge of the viewer's experience prior to sampling her degree of liking. And by definition, it is separated from the wider notion of aesthetics. But how does this argument square with our earlier observation that practically all art shares statistical regularities? Could there be some quality of art perception that is universal, that can be attributed to a universal aesthetic sentiment, perhaps one that arose through evolution to appreciate this very regularity, regardless of prior exposure or conditioning? Or might the novelty hypothesis of Berlyne suggest that humans would tire of these predictable statistics? I argue that there is no contradiction here: the fact that paintings share such statistical predictability is in part a consequence of the need to replicate the predictable statistics of the world for the purpose of perceptibility and in part an accident of handmade manufacture. In other words, there are certain essential qualities that make art images comprehensible to the brain, which art has been found to possess, and the absence of these qualities would spawn only a short-lived vogue for "unnatural art."

But what of those images that lack such regularities (knowingly or not)? Are they all excluded from the realm of art? Such images could well be art but the impact of the work is likely to be short-lived. This is so because a typical white noise image, for example, which lacks the regularity of scenes, is likely to be perceptually identical to any other typical white noise image (See Figure 1.1). Colloquially, this argument can be summed up thus: "If you can perceive it, it might be art; if you can't perceive it, it could still be art but no one will buy it for very long."

As Paul Bloom (2000) has noted, "Art is a self-conscious endeavor, and so any attempt to define art is likely to inspire a clever artist to react with a counterexample." So the fact that artists have rarely produced statistically random images could motivate

them to start a movement consisting of white noise images. This movement could expand with the addition of various nuances, or changes in other statistics. But because the visual system is adept at processing only images that approximate the characteristic statistics of natural scenes, artists will not produce white noise art for very long. For once the viewer has seen one white noise image, she has seen them all, at least in a perceptual sense. The viewer cannot tell one from another, so even a large change in statistics could go unnoticed by the viewer. This is so because, without altering the mean spatial correlation, one can produce large statistical differences among perceptually-identical white noise images. Indeed, one can produce changes far larger than are possible for natural scenes, without changing the underlying statistics. The perceptibility hypothesis views each such image as equally imperceptible.

6.2.2 Picture Liking

It should be said that those in the art world are right to be skeptical of the proposition that human artistic production can be understood or predicted in relation to an aesthetic sensibility shared by all humans throughout our history as a species. Indeed, this is not our claim. Our argument requires a clear delineation between aesthetics and picture liking, and a detailed understanding of what the observer has seen before. We define aesthetics as the viewer's longstanding sentiment *vis à vis* art. Aesthetics as such is characterized by an emotional mien that may or may not be expressible in words, and that may change with presentation of new or varied works. Picture liking, on the other hand, is a simple psychometric tool used to gauge pleasantness of appearance. The degree to which a person likes an image is certainly not the only factor in an assessment of art, nor in what would impel an artist to paint. It is instead a practical tool for understanding which images persist in culture, which become valuable in their own time, and which the prevailing economic systems have

tended to favor. Picture liking is more akin to economic metrics in this sense and as such it should not be construed as a measure of artistic quality, emotional, or historical resonance, or what can loosely be called aesthetic value. It is a rough measure of short-term affect.

There is an interesting question regarding the degree to which subject matter and representational style depend on familiarity and on novelty. Consider for example religious art of Europe in the pre-Renaissance era in terms of our framework. Art of pre-Renaissance Europe shows a remarkable consistency in subject matter across centuries, with relatively small changes in style and formal composition. Aside from the social pressure exerted on images by the hyper-religious culture of the time, the mere exposure effect could have been crucial to this progression. As David Freedberg argues (1989) religious art was omnipresent in the homes of those who could afford it but the range of subject matter was strictly prescribed. Freedberg quotes a 15th century primer on family care which mentions a few of the acceptable compositions for the home: “A good figure would be Jesus suckling, Jesus sleeping on his mother’s lap, Jesus standing politely before her, Jesus making a hem and the mother sewing that hem...little girls should be brought up in the sight of the eleven thousand virgins, discussing, fighting and praying. I would like them to see Agnes with the fat lamb, Cecelia crowned with roses, Elizabeth with many roses, Catherine on the wheel...” Compared to naturalistic images of the world’s various landscapes, this is a highly restrictive set of possible subjects. So the force of novelty preference, in this context, can be seen as giving rise to the tremendous formal advances made by Duccio, Giotto and others. Once cultural forces began to permit a wider array of subject matter, the new, though to modern eyes, rather subtle innovations in style brought on by late pre-Renaissance painters were let loose on a vast new territory of representation. That is, small changes in style held viewers attention during hundreds of years when content

was proscribed. As soon as the formula for content was relaxed, existing styles and methods of representation found wider applicability, and subsequently these were elaborated. However, I stress that this framework still lacks adequate measures of the relative changes in style and content, though it is possible that basic statistics could act as proxies for these quantities.

This brings up an important factor in understanding preference for novelty and familiarity in style and content: The effect of picture density (i.e., the number of pictures a typical observer will see in a lifetime). It is not simply because images were prescribed by religious, governmental and cultural authorities in the pre-Renaissance that led to the familiarity effect—this is also due to their extreme rarity. The choice of images for the home was so important in large part because the average art buyer could only afford to keep a handful of images on view over a lifetime. The same goes for images in the church. If the only images one sees are very similar to one another, only small changes in form or style are needed to keep one's interest. In addition, Zajonc's familiarity effect appears to act on content in this case: if artists use new representational techniques (e.g., chiaroscuro, linear perspective, atmospheric perspective), viewers—and more importantly art buyers—demand that content remain essentially constant.

One can see this if one considers each image a viewer sees as a single point in a cloud of points in a high-dimensional vector space. Images with similar content and general appearance cluster in a relatively limited portion of this space. If the degree of variety necessary to generate liking of a new image is given by a movement outside of this cloud (but not too far outside), then the smaller the cloud, the less change is needed to effect liking. And when artists hit upon powerful new or rediscovered techniques, like linear perspective, the size of the cloud of preferred images grows with increased viewer exposure, and there is room for myriad images to fill out that

cloud. Once the cloud reaches a certain density, some of the techniques used to fill it in can be pushed out beyond the edge of the cloud, and the process begins again (see Fig. 6.2).

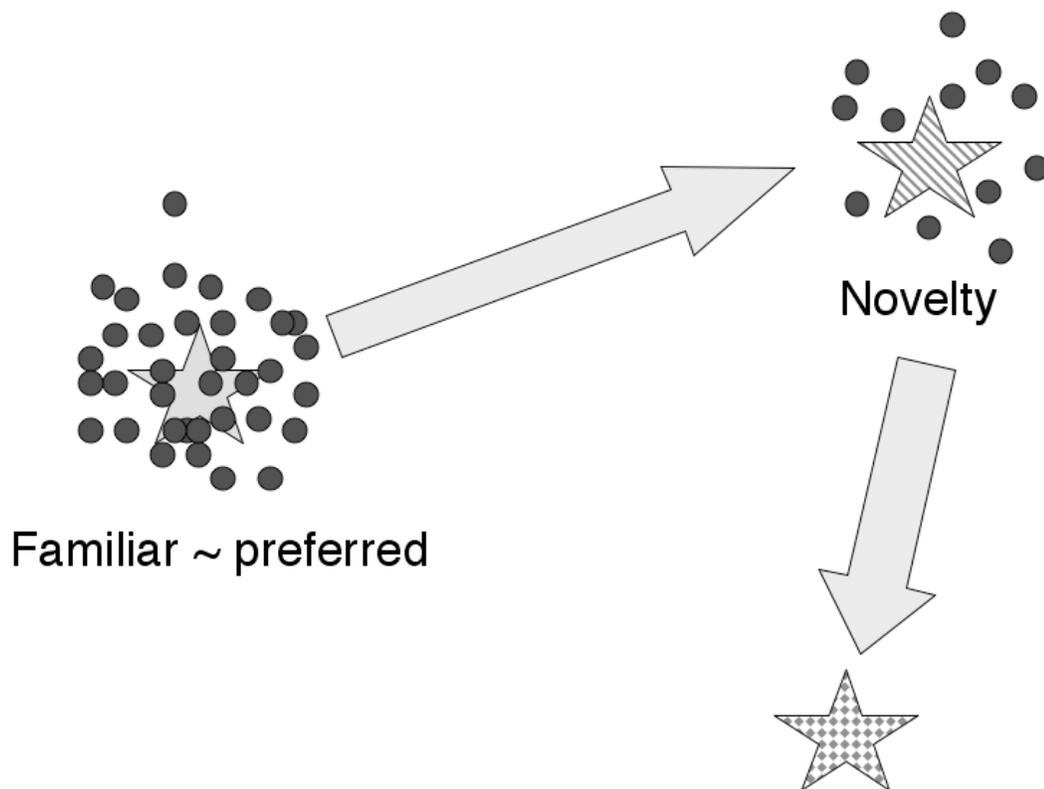


Figure 6.2. This figure shows a schematic picture of how familiarity and novelty produce movements within a state space of images that show natural scene-like statistics. Here, a cluster of familiar images (shown as dots)—say European devotional paintings from the pre-Renaissance—forms in a limited area of the state space. With innovation in style and/or content, the center of a cluster (stars) can shift so that it is outside the original cluster. If this innovation captures viewer attention effectively, the new cluster will fill out as the old one had, once viewers have become familiar with the innovation, and the process can start again. However, if the novelty does not produce a family of distinguishable images, this process will reach a dead end and new directions in this space will be explored. Such is the case for white noise images which, though novel to art galleries, are not perceptually distinguishable. Note that for simplicity, the high-dimensional space in which these images reside has been projected into two dimensions.

There is further evidence that picture density has a strong influence on the output of artists, and especially on innovation. Historians of Asian art (Whitfield, 1979; Lee,

1962; Sullivan, 1970) have noted that after a thousand years of development, landscape painting in China saw a low ebb in the late Qing dynasty (17th c.). As Lee says, “by 1800 landscape and all painting ha[d] run dry in theme, technique and mood.” Chinese landscape paintings, which to that point had been a hugely important and venerated class or art works, became scarce. It is interesting that while Asian art saw a millennium of innovation in images that depict landscape scenes for their inherent beauty, it was at roughly the time that Chinese interest in such images declined that Western painters took an interest in them.¹⁷ Only in modern (post-Imperial) times has there been a renewed interest and vitality in Chinese landscapes. Why did this happen?

In a sense, these older landscape paintings were loved to death by the Emperor Qianlong. He collected nearly all the paintings of any note and hid them away in the palace, where few budding artists or viewers of any kind would have access to them. At the same time, court painters had deconstructed the methods of the old masters and pronounced these as the orthodoxy; any older ideas that did not conform to this orthodoxy were expunged. Sullivan (1970) describes the emperor’s impact: Qianlong was “a voracious art collector, a niggardly and opinionated connoisseur, an unstoppable writer of inscriptions and stamper of seals who was determined, as a function of his imperial role, to leave his indelible mark upon China’s artistic legacy. His seals obliterate some of the finest paintings in the imperial collection....which grew to such enormous size that there were few ancient masterpieces that were not gathered behind the high walls of the Forbidden City, shut away forever from the painters who might still have studied them had they remained in private hands.”

¹⁷ As Watson (1979) has argued, Asian artists’ interest in depicting the natural world may spring in part from the Buddhist tradition, which in the early days of landscape painting was perhaps the only religious practice situated in the out-of-doors; Watson calls it an “open-air cult.”

As the number of images that a typical viewer sees in a lifetime dwindles to zero, artists concurrently lose their points of reference and eventually stop producing innovative art altogether. Note that at this time in China, when paintings depicted trees, rocks and water only, even the limited menu of content in pre-Renaissance art described by Freedberg was unavailable to painters. Of course, such scenes may not have had the same resonance as they would have in Europe but the point is that painters in both cultures had a relative paucity of subject matter with which to work.

This example illustrates how the cementing of formal and stylistic techniques can kill off an otherwise innovative artistic tradition. That is, possible advancements enabled by explorations in the realm of content were precluded by the narrowing of acceptable style. In contrast, pre-Renaissance painters were permitted to develop new stylistic approaches in order to sell paintings, which allowed them to search the space of new images more fully, and eventually to expand the range of acceptable content. Once content undergoes innovation, the process of formal refinement and elaboration begins the process anew. But when style alone is prescribed, advances in both style and content are severely limited.

So far I have only considered the effects of familiarity and novelty preference in cultures with low picture density, namely pre-Renaissance Europe and Qing dynasty China. But where images of all kinds are common in daily life, a different paradigm may hold. In image-rich cultures such as those in 21st century industrialized nations, new, though similar, subject matter is likely to be in demand and most changes in technique count for little. Consider for example the work of painter Thomas Kinkade, which by any reasonable measure of gross preference, is emblematic of the sort of images contemporary Americans prefer. Estimates by the Associated Press in 2006 suggested that 10 million Americans had a Kinkade work in their home, and Kinkade had sold over \$100 million worth of images. Kinkade's work has been dismissed by

art critics as formulaic and lacking inspiration (they also complain that most of his works are made by machines) but I would argue that their popularity among typical viewers is due in large part to the low threshold that these viewers set in terms of content novelty. That is, though the representational style of these works is highly consistent, there is enough variety in the content to keep millions of people from becoming bored. Their craving for novelty is satisfied by new content. So in this case, viewers prefer familiarity in style and novelty in content, whereas cultures with low picture density prefer familiarity in content and novelty in style.

Can these principles be used to map out the future progression of art? It is possible that the challenge of innovating in the domain of style may be facilitated by advances in other technologies. High dynamic range displays offer a novel way for image makers to overcome the “luminance problem” described in the previous chapter. Under ideal conditions, these displays are capable of a far better representation of the large dynamic range of luminances in typical scenes, without the need for nonlinear compression. This technology may soon give artists the tools to radically alter the current mode of representation by obviating the need for massive luminance compression. There is ongoing research in computer graphics that addresses one specific question in this area: how does one represent low dynamic range images in a high dynamic range (HDR) display? This is referred to as the problem of “legacy footage,” and there is likely to be a large range of inverse artist’s look-up tables that are capable of producing a pleasing image. Here, nonlinearities are used to expand rather than compress the range of luminances, in what is called *companding*. Already, artists have been moving in the direction of HDR imaging: photographer Jeff Wall displays his huge prints in light boxes, so that bright areas radiate light rather than reflect it. Once HDR technology becomes widespread and affordable, I predict that one will see a profusion of innovative images that exploit this technology in order to

produce images that were impossible to produce previously. Early adopters of the technology will sketch out the previously unpopulated space of images to which HDR technology allowed them access. Such is the power of new content to dull the search for novel stylistic techniques in cultures with high picture density that it takes a fundamental change in image photometry to allow such stylistic invention.

Of course, a daunting question that is left outstanding in this chapter is, Where does subject matter end and representational style begin? Given predictable regularities in basic statistics exhibited by art from different hemispheres and by art with different content (see Chapter 4), it is possible that these low-level statistics could help measure differences in style and content, provided an adequate definition of these terms can be found. It may be possible to test these notions using $1/f$ noise images, which lack all recognizable content, whose basic statistics can be manipulated in order to model style and content.

6.3 Conclusion

It should be stressed that the ideas presented here relate largely to simple preference and other low-level measures. Indeed, the emotional or spiritual resonance of a work or any other dimension relevant to a full and typical response is likely to be subject to the very effects we are describing. That is, one can only ask so many questions about viewer response to a particular image before the viewer craves to see something else, or perhaps reinforces her judgments in ways that would go against those responses if they were elicited in isolation.

My proposal is somewhat akin to proposals related to facial preference. Many studies have shown that observers recruited for testing at modern universities in many countries prefer faces that have been averaged with other faces over any particular face. That is, we have a preference for 2D frontal images of a face that does not

actually exist but rather is a theoretical composite of the faces we could have seen over a lifetime. The number of faces—or for my purposes, artworks—is an important predictor of changes in picture liking over time.

The central idea presented in this chapter is that the desire for novelty and the competing draw of the familiar, along with the statistical regularities of paintings, have important implications for the study of picture liking. In particular, if we consider only the narrow definition of picture liking, and we accept the proposition that content and style can be meaningfully differentiated, then certain principles can be devised. These principles, which may govern gross trends in the development and invention of pictures, are as follows:

(1) In areas of low picture density, small advances in technique can perpetuate liking for a relatively limited array of content. When technique is dogmatized, all innovation can cease.

(2) In areas of high picture density, new content is the main source of the novelty most viewers crave, while advances in style are often not sufficient to overcome viewers' quest for novelty. However, advances in picture technology (e.g., HDR displays) can offer new opportunities to artists in the stylistic domain.

REFERENCES

- Abeles, M., Vaadia, E. and Bergman, H. 1990. Firing patterns of single units in the prefrontal cortex and neural network models. *Netw.: Comput. Neural Syst.* **1**, 13-25.
- Adams, A. 1948. *The Negative: Exposure and Development*. Boston: New York Graphic Society.
- Aizenberg, J., Tkachenko, A., Weiner, S., Addadi, L., and Hendler, G. 2001. Calcitic microlenses as part of the photoreceptor system in brittlestars. *Nature* **412**, 819-822.
- Aks, D., and Sprott, J. C. 1996. Quantifying aesthetic preference for chaotic patterns. *Empirical Studies of the Arts* **14**, 1-16.
- Albert, M. V., Schnabel, A. and Field, D. J. 2008. Innate visual learning through spontaneous activity patterns. In preparation.
- Anstis, S. 2002. Was El Greco astigmatic? *Leonardo* **35**, 208.
- Arnett, D. and Sparker, T. E. 1981. Cross-correlation analysis of maintained discharge of rabbit retinal ganglion cells. *J. Physiol.* **317**, 29-47.
- Arnett, D. W. 1978. Statistical dependence between neighboring retinal ganglion cells in goldfish. *Exp. Brain Res.* **32**, 49-53.
- Arnheim, R. 1971. *Entropy and Art*. Berkeley: University of California Press.
- Atick, J. J. and Redlich A. N. 1992. What does the retina know about natural scenes? *Neural Comput.* **4**, 196-210.
- Attneave, F. 1954. Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183-193.
- Attewell, D. and Laughlin, S. B. 2001. An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* **21**, 1133-1145.
- Baddeley, R., Abbott, L. F., Booth, M. C., Sengpiel, F., Freeman, T., Wakeman, E. A. and Rolls, E. T. 1997. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. Lond. B* **264**, 1775-1783.

- Balboa, R. M., Tyler, C. W. and Grzywacz, N. M. 2001. Occlusions contribute to scaling in natural images. *Vision Res.* **41**, 955-964.
- Banks, M. S., Geisler, W. S. and Bennett, P. J. 1987. The physical limits of grating visibility. *Vision Res.* **27**, 1915-1924.
- Barlow, H. B. 1961. Possible principles underlying the transformation of sensory messages. In: *Sensory Communication*, W. A. Rosenblith, ed. Cambridge, MA: MIT Press.
- Barlow, H. B. 2001. Redundancy reduction revisited. *Netw. Comput. Neural Syst.* **12**, 241-253.
- Baum, E. B., Moody, J. and Wilczek, F. 1988. Internal representations for associative memory. *Biological Cybernetics* **59**, 217-228.
- Baylor, D. A., Lamb, T. D. and Yau K.W. 1979. Responses of retinal rods to single photons *J. Physiol. (Lond.)* **288**, 613-634.
- Baylor, D. A., Nunn, B. J. and Schnapf, J. L. 1987. Spectral sensitivity of the cones of the monkey *Macaca fascicularis* *J. Physiol.* **390**, 145.
- Bell, A. J. and Sejnowski, T. J. 1997. The 'independent components' of natural scenes are edge filters. *Vision Res.* **37**, 3327-3338.
- Beloozerova, I. N., Sirota, M. G. and Swadlow, H. A. 2003. Activity of different classes of neurons of the motor cortex during locomotion. *J. Neurosci.* **23**, 1087-1097.
- Benardete, E. A. and Kaplan, E. 1997a. The receptive field of the primate P retinal ganglion cell, I: linear dynamics. *Vis. Neurosci.* **14**, 169-186.
- Benardete, E. A. and Kaplan, E. 1997b. The receptive field of the primate P retinal ganglion cell, II: nonlinear dynamics. *Vis. Neurosci.* **14**, 187-205.
- Berkes, P. and Wiskott., L. 2005. Slow feature analysis yields a rich repertoire of complex cell properties. *J. Vision* **5**, 579-602.
- Berlyne, D. E. 1970. Novelty, familiarity and hedonic value. *Perception and Psychophys.* **8**, 279-286.
- Berry, M. J. Warland, D. K. and Meister, M. 1997. The structure and precision of retinal spike trains. *Proc. Natl. Acad. Sci. USA* **94**, 5411-5416.

- Biederman, I. and Vessel, E. A. 2006. Perceptual pleasure and the brain. *Am. Scientist* **94**, 249-255.
- Billock, V. A. 2001. Neural acclimation to 1/f spatial frequency spectra in natural images transduced by the human visual system. *Physica D* **137**, 379-391.
- Blaschke, T., Berkes, P. and Wiskott, L. 2006. What Is the Relation Between Slow Feature Analysis and Independent Component Analysis? *Neural Comput.* **18**, 2495-2508.
- Bloom, P. 2000. *Descartes' Baby*. New York: Basic Books.
- Brady, N. and Field, D. J. 2000. Local contrast in natural images: normalisation and coding efficiency. *Perception* **29**, 1041-1055.
- Brady, N., and Field, D.J. 1995 "What's constant in contrast constancy: the facts of scaling on the perceived contrast of bandpass patterns". *Vision Res.* **35**, 739-756.
- Brecht, M. and Sakmann, B. 2002. Dynamic representation of whisker deflection by synaptic potentials in spiny stellate and pyramidal cells in the barrels and septa of layer 4 rat somatosensory cortex. *J. Physiol.* **543**, 49-70.
- Brown, S. P., He, S. and Masland, R. H. 2000. Receptive field microstructure and dendritic geometry of retinal ganglion cells. *Neuron* **27**, 371-383.
- Buchsbaum, G. and Gottschalk, A. 1983. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proc. R. Soc. Lond. B* **220**, 89-113.
- Burton, G.J. and Moorhead, I. R. 1987. Color and spatial structure in natural scenes. *Appl. Optics* **26**, 157-170.
- Butts, D. A. 2002. Retinal Waves: Implications for synaptic learning rules during development *Neuroscientist* **8**, 243-253.
- Butts, D. A., Feller, M. B., Shatz, C. J. and Rokhsar, D. S. 1999. Retinal waves are governed by collective network properties. *J. Neurosci.* **19**, 3580-3593.
- Campbell, F. W. and Green, D. G. 1965. Optical and retinal factors affecting visual resolution. *J. Physiol.* **181**, 576-593.
- Carrasco, M. M., Razak, K. A. and Pallas, S. L. 2005. Visual experience is necessary for maintenance but not development of refined retinotopic maps in superior colliculus. *J Neurophysiol.* **94**, 1962-1970.

- Cavanaugh, P. 1999. Pictorial art and vision. In: *The MIT Encyclopedia of Cognitive Sciences*, R A. Wilson and F C. Keil, eds. Cambridge: MIT Press.
- Caywood, M. S., Willmore, B. and Tolhurst, D. J. 2004. Independent Components of Color Natural Scenes Resemble V1 Neurons in Their Spatial and Color Tuning. *J. Neurophysiol.* **91**, 2859-2873.
- Chandler, D. M. and Field, D. J. 2007. Estimates of the information content and dimensionality of natural scenes from proximity distributions. *J. Opt. Soc. Am.* **24**, 922-941.
- Changizi, M. A. and Shimojo, S. 2005. Parcellation and area-area connectivity as a function of neocortex size. *Brain, Behavior and Evolution* **66**, 88-98.
- Changizi, M. A., Zhang, Q., Ye, H. and Shimojo, S. 2006. The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. *American Naturalist* **167**, E117-E139.
- Chapman, B. and Godecke, I. 2000. Cortical cell orientation selectivity fails to develop in the absence of on-center retinal ganglion cell activity. *J. Neurosci.* **20**, 1922-1930.
- Chen, Y., Geisler, W. S. and Seidemann, E.. 2006. Optimal decoding of correlated neural population responses in the primate visual cortex. *Nature Neurosci.* **9**, 1412-1420.
- Cherniak C, Mokhtarzada Z, Rodriguez-Esteban R, Changizi K. 2004. Global optimization of cerebral cortex layout. *Proc. Natl. Acad. Sci. USA* **101**, 1081-6.
- Cherniak, C. 1994. Component placement optimization in the brain *J. Neurosci.* **14**, 2418-2427.
- Chklovskii, D. B. and Koulakov, A. A. 2004 Maps in the brain: What can we learn from them? *Ann. Rev. Neurosci.* **27**, 369-92.
- Chklovskii, D. B., Schikorski, T. and Stevens, C. F. 2002. Wiring optimization in cortical circuits. *Neuron* **34**, 341-347.
- Croner, L. J. and Kaplan, E. 1995. Receptive fields of P and M ganglion cells across the primate retina. *Vision Res.* **15**, 7-24.
- Crotty, P., Sangrey, T. and Levy, W. B. 2006. Energy cost of action potential velocity. *J. Neurophysiol.* **96**, 1237-1246.

- Crowley, J. C. and Katz, L. C. 2000. Early development of ocular dominance columns *Science* **290**, 1321 – 1324.
- Cutting, J. E. 2003. Gustave Caillebotte, French Impressionism and mere exposure. *Psychonomic Bull. Rev.* **10**, 319-343.
- Cutting, J. E. 2006. *Impressionism and Its Canon*. University Press of America, Lanham, MD, USA.
- Cutting, J. E. and Garvin, J. J. 1987. Fractal curves and complexity. *Perception and Psychophysics* **42**, 365-370.
- Daugman, J. G. 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am.* **2**, 1160-1169.
- David, S. V., Vinje, W. E. and Gallant, J. L. 2004. Natural stimulus statistics alter the receptive field structure of V1 neurons. *J. Neurosci.* **24**, 6991-7006.
- Debevec, P. E. and Malik, J. 1997. Recovering high dynamic range radiance maps from photographs. *Proc. SIGGRAPH 97*, 369-378.
- De Valois, R. L. and De Valois, K. K. 1988. *Spatial Vision*. New York: Oxford Press.
- De Valois, R. L., Morgan, H., and Snodderly, D. M. 1974. Psychophysical studies of monkey vision-III. spatial luminance contrast sensitivity tests of macaque and human observers. *Vision Res.* **14**, 75-81.
- Denk, W. and Webb, W. W. 1989. Thermal-noise-limited transduction observed in mechanosensory receptors of the inner ear. *Phys. Rev. Lett.* **63**, 207-210.
- DeVries, S. H. 1999. Correlated firing in rabbit retinal ganglion cells. *J. Neurophysiol.* **81**, 901-920.
- DeWeese, M., Wehr, M. and Zador, A. 2003. Binary spiking in auditory cortex. *J. Neurosci.* **23**, 7940-7949.
- Diamond, J. 1996. *Guns, Germs and Steel: The Fates of Human Societies*. New York: W. W. Norton.
- Dong, D. W. 2001. Spatiotemporal inseparability of natural images and visual sensitivities. In: *Motion Vision: Computation, Neural and Ecological Constraints*, J. M. Zanker and J. Zeil eds. Berlin: Springer-Verlag.

- Dong, D. W. and Atick, J. J. 1995a. Statistics of natural time-varying images. *Netw. Comput. Neural Syst.* **6**, 345-358.
- Dong, D. W. and Atick, J. J. 1995b. Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Netw. Comput. Neural Syst.* **6**, 159-178.
- Dror, R. O. Willsky, A. S., and Adelson, E. H. 2004. Statistical characterization of real-world illumination. *J. Vision* **4**, 821-837.
- Durbin, R. and Mitchison G. 1990. A dimension reduction framework for understanding cortical maps. *Nature* **343**, 644-47.
- Falco, C. 2007. *Personal communication*.
- Farah, M. 2004. *Visual Agnosia*. Cambridge, Mass.: MIT Press.
- Field, D. J. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* **4**, 2379-2394.
- Field, D. J. 1993. Scale-invariance and self-similar 'wavelet' transforms: an analysis of natural scenes and mammalian visual systems. In: *Wavelets, Fractals and Fourier Transforms: New Developments and New Applications*, M. Farge, J. C. R. Hunt and J. C. Vassilicos, eds. Oxford: Oxford University Press.
- Field, D. J. 1994. What is the goal of sensory coding? *Neural Comput.* **6**, 559-601
- Field, D. J. and Brady, N. 1997. Wavelets, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Res.* **37**, 3367-3383.
- Finlay, B. L., de Lima Silveira, L. C. and Reichenbach, A. 2004. Comparative aspects of visual system development. To appear in: *The Structure, Function and Evolution of the Primate Visual System*, ed. J. Kremers. New York: John Wiley.
- Frazor, R. A. and Geisler, W. A. 2006. Local luminance and contrast in natural images. *Vision Res.* **46**, 1585-1598.
- Freedberg, D. 1989. *The Power of Images*. Chicago: University of Chicago Press.
- Gabor, D. 1946. Theory of Communication. *J. Inst. Elect. Eng.* **93**, 429-457.
- Geake, J. and Landini, G. 1997. Individual differences in the perception of fractal curves. *Fractals* **5**, 129-143.

- Geisler, W. S., Perry, J. S., Super, B. J. and Gallogly, D. P. 2001. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.* **41**, 711-724.
- Ghose, G.M., Yang T. and Maunsell, J. H. 2002. Physiological correlates of perceptual learning in monkey V1 and V2. *J. Neurophysiol.* **87**, 1867–1888.
- Gilchrist, A. 1979. The perception of surface blacks and whites. *Sci. Am.* **240**, 112-123.
- Gombrich, E. H. 1961. *Art and Illusion*. Princeton University Press, Princeton.
- Gonzalez, R. C., Woods, R. E. and Eddins, S. L. 2004. *Digital Image Processing Using Matlab*. Upper Saddle River, N.J.: Prentice-Hall.
- Graham, D. J. and Field, D. J. 2007a. Efficient neural coding of natural images. In *New Encyclopedia of Neuroscience*, L. R. Squire, ed. London: Elsevier, in press.
- Graham, D. J. and Field, D. J. 2007b. Statistical regularities of art images and natural scenes: spectra, sparseness and nonlinearities. *Spatial Vision* **21**, 149-164.
- Graham, D. J., Chandler, D. M. and Field, D. J. 2006. Can the theory of "whitening" explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Res.* **46**, 2901-2913.
- Graham, D. J., Rockmore, D. N, and Field, D. J. 2008. Global nonlinear compression of natural luminances in painted art. *Proc. SPIE Electronic Imaging*, submitted.
- Graham, N. 1989. *Visual Pattern Analyzers*. New York: Oxford University Press.
- Gross, C. G. 2002. Genealogy of the "Grandmother Cell." *Neuroscientist* **8**, 512-518.
- Hagerhall, C. M., Purcell, T. and Taylor, R. 2004. Fractal dimension of landscape silhouette outlines as a predictor of landscape preference. *J. Environmental Psych.* **24**, 247-255.
- Hahnloser, R. H. R, Kozhevnikov, A. and Fee, M. S. 2002. An ultrasparse code underlies the generation of neural sequences in songbirds. *Nature* **419**, 65-70.
- Hochberg, J. and Brooks, V. 1962. Pictorial recognition as an unlearned ability: a study of one child's performance. *Am. J. Psych.* **75**, 624-628.
- Hofman, M. A. 1983. Energy metabolism, brain size and longevity in mammals. *Quant. Rev. Biol.* **58**, 495.

- Hoyer P. O. and Hyvärinen, A. 2000. Independent component analysis applied to feature extraction from colour and stereo images. *Netw Comput. Neural Syst.* **11**, 191-210.
- Huxley, J. S. 1942. Origins of human graphic art. *Nature* **149**, 637.
- Hyvärinen, A. and Hoyer, P. O. 2000. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput.* **12**, 1705-1720.
- Johnsen J. A. and Levine, M. W. 1983. Correlation of activity in neighbouring goldfish ganglion cells: relationship between latency and lag. *J. Physiol.* **345**, 439-449.
- Jones, J. P. and Palmer, L. A. 1987. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* **58**, 1233-1258.
- Kaplan, E. and Benardete, E. 2001. The dynamics of primate retinal ganglion cells. *Prog. Brain Res.* **134**, 17-34.
- Kayser, C., Körding, K. P. and König, P. 2003. Learning the nonlinearity of neurons from natural visual stimuli. *Neural Comput.* **15**, 1751-1759.
- Keller, J. M., Crownover, R. M. and Chen, R. Y. 1987. Characteristics of natural scenes related to the fractal dimension. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **9**, 621-627.
- Kendrick, K. M. and Baldwin, B. A. 1987. Cells in temporal cortex of conscious sheep can respond preferentially to the sight of faces. *Science* **236**, 448-450.
- Knill, D. C., Field, D. and Kersten, D. 1990. Human discrimination of fractal images. *J. Opt. Soc. Am.* **77**, 1113-1123.
- Knudson, E. I. and Konishi, M. 1978. Center-surround organization of auditory receptive fields in the owl. *Science* **202**, 778-780.
- Koulakov, A. A. and Chklovskii, D. B. 2001. Orientation preference patterns in mammalian visual cortex: A wire length minimization approach. *Neuron* **29**, 519-527.
- Kreutz-Delgado, K., Murray, J. F., Rao, B. D., Engen, K., Lee, T.-W. and Sejnowski, T. J. 2003. Dictionary learning algorithms for sparse representation. *Neural Comput.* **15**, 349-396.

- Kube, P. and Pentland, A. 1988. On the imaging of fractal surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**, 704-707.
- Lamme, V. A. 1995. The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.* **15**, 1605–1615.
- Land, M. 1985. The Eye: Optics. In *Comprehensive Insect Physiology, Biochemistry and Pharmacology*. Eds. G. A. Kerkut and L. I. Gilbert. London: Pergamon.
- Laughlin, S. B. 1981. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.* **36C**, 910-912.
- Laughlin, S. B. and Sejnowski, T. J. 2003. Communication in Neuronal Networks *Science* **301**, 1870-1874.
- Laughlin, S. B., de Ruyter van Steveninck, R. R. and Anderson, J. C. 1998 The metabolic cost of neural information. *Nature Neurosci.* **1**, 36-41.
- Lee, S. 1962. *Chinese Landscape Painting*. Cleveland: Cleveland Museum of Art/Harry Abrams.
- Lee, T. W., Wachtler, T. and Sejnowski, T. J. 2002. Color opponency is an efficient representation of spectral properties in natural scenes. *Vision Res.* **42**, 2095-2103.
- Lennie, P. 1998. Single units and visual cortical organization. *Perception*, **27**, 889-935.
- Lennie, P. 2003. The cost of cortical computation. *Curr. Biol.* **13**, 493-497.
- Levy, W. B. and Baxter, R. A. 1996. Energy efficient neural codes. *Neural Comput.* **8**, 531-543.
- Levy, R. and Jaeger, T. F. 2006. Speakers optimize information density through syntactic reduction. *Proc. Neural Inform. Processing Syst.* **19**, 1-6.
- Linsker, R. 1989. How to generate ordered maps by maximizing the mutual information between input and output. *Neural Comput.* **1**, 402-411.
- Lythgoe, J. N. 1979. *The Ecology of Vision*. Oxford: Oxford University Press.
- Lyu, S., Rockmore, D. N. and Farid, H. 2004. A digital technique for art authentication. *Proc. Nat. Acad. Sci. USA* **101**, 17006-17010.
- Malo, J. and Gutiérrez, J. 2006. V1 non-linear properties emerge from local-to-global non-linear ICA. *Netw. Comput. Neural Syst.* **17**, 85-102.

- Maloney, L. T. and Wandell, B. A. 1986 Color constancy: a method for recovering surface spectral reflectance. *J. Opt. Soc. Am. A* **3**, 29-33.
- Mandelbrot, B. B. 1977. *The Fractal Geometry of Nature*. W. H. Freeman, New York.
- Mante, V., Frazor, R. A., Bonin, V., Geisler, W. S. and Carandini, M. 2005. Independence of luminance and contrast in natural scenes and in the early visual system. *Nature Neurosci.*, **8**, 1690-1697.
- Marcelja, S. 1980. Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am.* **70**, 1297-1300.
- Masland, R. H. 2001. The fundamental plan of the retina. *Nature Neurosci.* **4**, 877-887.
- Mastrorarde, D. N. 1989. Correlated firing of retinal ganglion cells. *Trends Neurosci.* **12**, 75-80.
- McCann, J. 2005. Rendering high-dynamic range images: Algorithms that mimic human vision. *Proc. AMOS Technical Conf.*, 19-28.
- McCann, J. J. 2007. Art, science and appearance in HDR images. *J. Soc. Inform. Display* **15**, 709-719.
- Meister, M. 1996. Multineuronal codes in retinal signaling. *Proc. Nat. Acad. Sci. USA* **93**, 609-614.
- Meister, M. and Berry, M. J. 1999. The neural code of the retina. *Neuron* **22**, 435-450.
- Meister, M., Lagnado, L. and Baylor, D. A. 1995. Concerted signaling by retinal ganglion cells. *Science* **270**, 1207-1210.
- Mitchison, G. 1991. Neuronal branching patterns and the economy of cortical wiring. *Proc. Roy. Soc. B Biol.* **245**, 151-158.
- Mosteller, F. and Wallace, D. L. 1964. *Inference and Disputed Authorship: The Federalist*. Reading, MA: Addison-Wesley.
- Motoyoshi, I., Nishidam S., Sharan, L. and Adelson, E. H. 2007. Image statistics and the perception of surface qualities. *Nature* **447**, 206-209.
- Mumford, D. 1994. Neuronal architectures for pattern-theoretic problems. In *Large scale neuronal theories of the brain*, C. Koch and J. L. Davis, eds. Cambridge, Mass.: MIT Press.

- Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., and Woods, D. L. 2002. Shape perception reduces activity in human primary visual cortex. *Proc. Natl. Acad. Sci. USA* **99**, 15164-15169.
- Naka, K. I. and Rushton, W. A. 1966. S-potentials from colour units in the retina of fish (*Cyprinidae*). *J. Physiol.* **185**, 536-555.
- Nirenberg, S. Carcieri, S. M., Jacobs, A. L. and Latham, P. E. 2001. Retinal ganglion cells act largely as independent encoders. *Nature* **411**, 698-701.
- Norman, R. A. and Werblin, F. S. 1974. Control of retinal sensitivity. I. Light and dark adaptation of vertebrate rods and cones. *J. Gen. Physiol.* **63**, 37-61.
- Olshausen, B. A. 2003b. Learning sparse, overcomplete representations of time-varying natural images. *IEEE International Conference on Image Processing*. Sept. 14-17, 2003. Barcelona, Spain.
- Olshausen, B. A. and Field, D. J. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607-609.
- Olshausen, B. A. and Field, D. J. 2005. How close are we to understanding V1? *Neural Comput.* **17**, 1665-1699.
- Olshausen, B. A. 2003a. Principles of Image Representation in Visual Cortex. In: *The Visual Neurosciences*, L.M. Chalupa, J.S. Werner, eds., MIT Press.
- Olshausen, B. A., Anderson, C. H. and Van Essen, D. C. 1993. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neurosci.* **13**, 4700-4719.
- Osorio, D. and Vorobyev, M. 1996. Colour vision as an adaptation to frugivory in primates *Proc. R. Soc. Lond. B.* **263**, 593-599.
- Parraga, C. A., Troscianko, T. and Tolhurst, D. J. 2000. The human visual system is optimised for processing the spatial information in natural visual images. *Curr. Biol.* **10**, 35-38.
- Pelli, D. G. 1981. *Effects of visual noise*. Ph. D. thesis. Cambridge University, Cambridge, England.
- Pentland, A. 1984. Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 661-674.

- Prenger, R., Wu, M. C.-K., David, S. V. and Gallant, J. L. 2004. Nonlinear V1 responses to natural scenes revealed by neural network analysis. *Neural Netw.* **17**, 663-679.
- Puchalla, J., Schneidman, E., Harris, R. A. and Berry, M. J. 2005. Redundancy in the population code of the retina. *Neuron* **46**, 493-504.
- Rainer, G. and Miller E. K. 2000. Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron* **27**, 179–189.
- Rainer, G., Lee, H. and Logothetis, N. 2004. The effect of learning on the function of monkey extrastriate visual cortex. *PLoS Biol.* **2**, 275-283.
- Ramachandran, V. S. and Hirstein, W. 1999. The science of art: A neurological theory of aesthetic experience. *J. Consciousness Stud.* **6**, 15-51.
- Ranganath, C. and Rainer, G. 2003. Neural mechanisms for detecting and remembering novel events. *Nat. Rev. Neurosci.* **4**, 193–202.
- Rao, R. P. N. and Ballard, D. H. 1997. Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Comput.* **9**, 721-763.
- Redies, C. 2007. A universal model of esthetic perception based on the sparse and efficient coding of natural stimuli. *Spatial Vision* **21**, 97-117.
- Redies, C., Hasenstein, J., and Denzler, J. 2007. Fractal-like image statistics in visual art: similarity to natural scenes. *Spatial Vision* **21**, 137-148.
- Reinagel, P. and Reid, R. C. 2000. Temporal coding of visual information in the thalamus. *J. Neurosci.* **20**, 5392-5400.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R. and Bialek, W. 2001. *Spikes*. Cambridge, Mass.: MIT Press.
- Riesenhuber, M. and Poggio, T. 2000. Models of object recognition. *Nature Neurosci.* **3**, 1199-1204.
- Rodieck, R. W. 1965. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Res.* **5**, 583-601.
- Rogowitz, B. E. and Voss, R. F. 1990. Shape perception and low-dimensional fractal boundaries. *Proc. SPIE* 1249, 387-394.
- Rolls, E. T. and Tovee, M. J. 1995. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol.* **73**, 713-726.

- Ruderman, D. L. 1997. Origins of scaling in natural images. *Vision Res.* **37**, 3385-3398.
- Ruderman, D. L. 1994 Designing receptive fields for highest fidelity. *Netw. Comput. Neural Syst.* **5**, 147 – 155.
- Ruderman, D. L. and Bialek, W. 1994. Statistics of natural images: scaling in the woods. *Phys. Rev. Lett.* **73**, 814-817.
- Schwartz, O. and Simoncelli, E. P. 2001. Natural signal statistics and sensory gain control. *Nature Neurosci.* **4**, 819-825.
- Shapley R. and Victor, J. D. 1979. Nonlinear spatial summation and the contrast gain properties of cat retinal ganglion cells. *J. Physiol.* **290**, 141-161.
- Shapley R. and Victor, J. D. 1981. How the contrast gain control modifies the frequency responses of cat retinal ganglion cells. *J. Physiol.* **318**, 161-179.
- Shoham, S., O'Connor, D. H. and Segev, R. 2006. How silent is the brain: is there a "dark matter" problem in neuroscience? *J. Comp. Physiol. A* **192**, 777-784.
- Simoncelli, E. P. and Olshausen, B. A. 2001. Natural image statistics and neural representation. *Ann. Rev. Neurosci.* **24**, 1193-1215.
- Spehar, B., Clifford, C. W. G., Newell, B. R. and Taylor, R. P. 2002. Universal aesthetics of fractals. *Computers and Graphics* **27**, 813-820.
- Sprott, J. C. 1993. Automatic generation of strange attractors. *Computer & Graphics* **17**, 325–32.
- Squires, T. M. 2004. Optimizing the vertebrate vestibular semicircular canal: Could we balance any better? *Phys. Rev. Lett.* **93**, 198106.
- Srinivasan, M. V., Laughlin, S. B. and Dubs, A. 1982. Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. B Biol. Sci.* **216**, 427-459.
- Stamps, A. E. 2002. Fractals, skylines, nature and beauty. *Landscape and Urban Planning* **60**, 163-184.
- Sullivan, M. 1970. *A Short History of Chinese Art*. Berkeley: University of California Press.
- Sundar, V. C., Yablon, A. D., Grazul, J. L., Ilan, M. and Aizenberg, J. 2003. Fibre-optical features of a glass sponge. *Nature* **424**, 899-900.

- Taylor, D. R., Finkel, L. H. and Buchsbaum, G. 2000. Color-opponent receptive fields derived from independent component analysis of natural images. *Vision Res.* **40**, 2671-2676.
- Taylor, R. P., Micolich, A. P. and Jonas, D. 1999. Fractal Expressionism. *Physics World* **12**, 25-29.
- Taylor, R. P., Micolich, A. P. and Jonas, D. 2002. The construction of Jackson Pollock's fractal drip paintings. *Leonardo* **35**, 203-207.
- Thompson, L. T. and Best, P. J. 1989. Place cells and silent cells in the hippocampus of freely-behaving rats. *J. Neurosci.* **9**, 2382-2390.
- Tolhurst, D. J., Tadmor, Y. and Tang, C. 1992. The amplitude spectra of natural images. *Ophthalmic Physiol. Opt.* **12**, 229-232.
- Torralba, A. and Oliva, A. 2003. Statistics of natural image categories. *Netw.: Comput. Neural Syst.* **14**, 391-412.
- Treves, A., Panzeri, S., Rolls, E. T., Booth, M. and Wakenan, E. A. 1999. Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. *Neural Comput.* **11**, 601-631.
- Van Essen, D. C. 1997. A tension-based theory of morphogenesis and compact wiring in the central nervous system. *Nature* **385**, 313-318.
- van Hateren, J. H. and Ruderman, D. 1998. Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proc. R. Soc. Lond. B* **265**, 2315-2320.
- van Hateren, J. H. and Snippe, H. P. 2001. Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells, *Vision Res.* **41** 1851-1865.
- van Hateren, J.H. and van der Schaaf, A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Roy. Soc. B Biol. Sci.* **265**, 359-366.
- Victor, J. D. 1987. The dynamics of cat retinal X cell centre. *J. Physiol.* **386**, 219-246.
- Vincent, B. T. and Baddeley, R. J. 2003. Synaptic energy efficiency in retinal processing. *Vision Res.* **43**, 1283-1290.

- Vinje, W. E. and Gallant, J. L. 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* **287**, 1273-1276.
- Vinje, W. E. and Gallant, J. L. 2002. Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *J. Neurosci.* **22**, 2904-2915.
- Voss, R. F. 1985. Random fractal forgeries. In *Fundamental Algorithms for Computer Graphics*, R. A. Earnshaw, ed. New York: Springer-Verlag.
- Voss, R. F. and Clarke, J. 1978. 1/f noise in music: Music from 1/f noise. *J. Acoust. Soc. Am.* **63**, 258-263.
- Wachtler, T., Albright, T. D. and Sejnowski, T. 2001. Nonlocal interactions in color perception: nonlinear processing of chromatic signals from remote inducers. *Vision Res.* **41**, 1535-1546.
- Wainwright, M. J. 1999. Visual adaptation as optimal information transmission. *Vision Res.* **39**, 3960-3974.
- Wässle H. and Boycott B. B. 1991. Functional architecture of the mammalian retina. *Physiol. Rev.* **71**, 447-480.
- Watson, W. 1979. Landscape elements in the early Buddhist art of China. In: *Landscape Style in Asia* W. Watson ed. London: School of Oriental and African Studies.
- Watson, A. B. 1983. Detection and recognition of simple spatial forms. In: *Physical and Biological Processing of Images*, O. J. Braddick and A. C. Sleight, eds. Berlin: Springer-Verlag.
- Webster, M. A. and Mollon, J. D. 1997. Adaptation and the color statistics of natural images. *Vision Res.* **37**, 3283-3298.
- Whitfield, R. 1979. Landscape painting in the late Ch'ing period. In: *Landscape Style in Asia* W. Watson ed. London: School of Oriental and African Studies.
- Willats, J. 1997. *Art and Reresentation: New Principles in the Analysis of Pictures*. Princeton, NJ.: Princeton University Press.
- Willmore, B. and Tolhurst, D. J. 2001. Characterizing the sparseness of neural codes. *Netw. Comput. Neural Syst.* **12**, 255-270.
- Winkielman, P., Schwarz, N., Reber, R. and Fandero, T. A. 2003. Cognitive and affective consequences of visual fluency: When seeing is easy on the mind. In:

- Persuasive Imagery: A Consumer Response Perspective*, L. M. Scott and R. Batra, eds. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wiskott, L. 2005. How does our visual system achieve shift and size invariance? In: *23 Problems in Systems Neuroscience*, J. L. van Hemmen and T. J. Sejnowski, eds. Oxford University Press.
- Wiskott, L. and Sejnowski, T. J. 2002. Slow feature analysis: unsupervised learning of invariances. *Neural Comput.* **14**, 715-770.
- Wolff, J. G. 2006. *Unifying computing and cognition: The SP theory and its applications*. ebook.
- Wong, R. O. L. 1999. Retinal waves and visual system development. *Ann. Rev. Neurosci.* **22**, 29-47.
- Wynn, T. G. 1996. The evolution of tools and symbolic behaviour. In: *Handbook of Human Symbolic Evolution*, A. Lock and C. R. Peters, eds. Oxford: Blackwell.
- Yu, Y. and Lee, T. S. 2004. Adaptive contrast gain control and information maximization. *Neurocomputing* **65-66**, 111-116.
- Zajonc, R. B. 1968. Attitudinal effects of mere exposure. *J. Personality & Social Psych.* **9**, 1-27.
- Zajonc, R. 1970. Brainwash: Familiarity breeds comfort. *Psych. Today* **3**, 33-35, 60-62.
- Zajonc, R. B. 1980. Feeling and thinking: preferences need no inferences. *Am. Psychologist* **35**, 151-175.
- Zeki, S. 1998. Art and the brain. *Dædalus* **127**, 71-103.
- Zetsche, C. and Röhrbein, F. 2001. Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Netw. Comput. Neural Syst.* **12**, 331-350.