

Chapter 1

Overview

While working to bring this book together, I was inspired and overwhelmed by the breadth and vibrancy of vision science. Vision scientists solve problems across the fields of biology, psychology, and engineering. Our field takes on problems ranging from the nature of consciousness to the hurry-up-and-ship-it applications needed to keep a company afloat.

I have written this book for the student who wishes to know *how* to study vision. The pages are filled with measurements and facts; but, my goal in writing this book is to explain to the student how we learned these facts, not the facts themselves. Portions of this book are written with the expectation that the reader has had some experience with linear algebra and calculus. In most sections of the book, however, I have tried to provide the reader with the basic ideas without using mathematical symbols or formal arguments.

To organize the material presented, I have divided the book into three sections. My division reflects three of the basic problems of vision: encoding, representation, and interpretation.

1.1 Encoding

The first section of this book describes how the retinal image is encoded by the visual pathways. The material in this section is particularly important for three reasons.

First, how the visual system encodes light has implications for everything else the visual pathways do. Distortions that are introduced into the signal by poor optics, sparse and uneven spatial sampling of the image, or meager wavelength encoding become part of the signal that must be represented and interpreted by the central visual pathways. We can't understand the central nervous system without understanding the quality of information encoded within the eye.

Second, the properties of the visual encoding have implications for the design of instruments that display visual information. The quality of the representation of pattern and color in display media must be structured to satisfy, but not exceed, the limits of the human visual system. For example, the industry of color imaging, including visual displays, film, and color printing, relies on the fact that human color vision uses three types of cone photopigments to encode light. As a result of this sparse representation of wavelength, color reproductions need not represent the wavelength composition of the original in order to provide a satisfactory appearance match. This is but one example of many in which the initial encoding of the image in the human eye defines practical limits whose properties determine the character of imaging devices.

Third, the methods and standards of proof that are used to understand image encoding set an important example concerning the standards of explanation we aim to achieve at all levels of vision science. The questions of methods and standards of proof are very important in an interdisciplinary field like vision science, which draws on expertise from many different areas. The first section of this book contains several examples that combine physical calculations, biological experiments, and behavioral studies. By examining how these fields come together when we measure the quality of the retinal image formed by the optics of the eye, and again when we establish that human color vision is trichromatic, we see how these diverse fields can forge strong links that define important aspects of visual function. We can learn from these examples as we move on to other problems in vision science.

1.2 Representation

The second section of this volume reviews how the encoded image is represented by the neural response within the peripheral visual pathways. Our understanding of the neural representation is based on work in several different disciplines. This section begins with a review of the anatomical and electrophysiological measurements of the image representation within the retina and primary visual cortex. These measurements characterize the neural hardware of the visual representation. These studies reveal several important points:

- There are many anatomically distinct types of neurons with various distinct functions,
- The different anatomical types of neurons respond to light stimulation in different ways and their signals are communicated to different destinations,
- The microcircuitry of the local neural connections is very precise, and not at all random

The second half of this section reviews psychological and computational studies of image representation. The behavioral studies of image representation involve the simplest performances, such as detection, discrimination and simple recognition. These experiments have led to various proposals about how pattern and color information is represented within the retina and early cortical areas. The computational studies of image representation cover fundamental issues, such as efficient image representation, that are important for image representation.

1.3 Interpretation

Perception is an interpretation of the retinal image, not a description. The third section of this book contains examples of how we interpret the retinal image to assign perceptual properties such as color, motion, and shape to objects.

We perceive object properties from the retinal image by inference, not deduction. The retinal image information is ambiguous and incomplete, and usually the retinal image can be interpreted in many different ways. Because we begin with ambiguous information, we cannot make deductions from the retinal image, only inferences.

When we try to interpret image data in order to infer the color, motion and shape of objects, we confront the same challenges as the visual pathways. To those who have studied tried to develop computational models of these processes, the success of the visual system in interpreting image data represents seem remarkable. The study of these computational problems has made it clear that the visual system succeeds in interpreting images because of statistical regularities present in the visual environment and hence in the retinal image. These regularities permit the visual system to use the fragmentary information present in the retinal image to draw accurate inferences about the physical cause of the image. For example, when we make inferences from the retinal image, the knowledge that we live in a three-dimensional world is essential to the correct interpretation of the image. We are made aware of the existence of these powerful interpretations and their assumptions when they are in error, that is, when we discover a visual illusion.

1.4 Selection of Topics: Methods

Certain theoretical and empirical methods appear repeatedly within vision science. The most important theoretical method, which appears across all areas of vision science, is *linear systems theory*. Whether characterizing optics, neurophysiology, color vision, spatial vision, image compression, or pattern analysis, linear systems play an important role. There is little possibility of understanding the current foundations of vision science without understanding linear systems. I introduce the principles of linear

systems in the first chapter and I refer to them throughout the book.

Linear methods are not a theory of vision; linear systems methods consist of a set of experiments that one should use to analyze a system. If the system's performance satisfies certain experimental properties, such as the principle of superposition, then we can use linear methods to characterize the system completely. Even if the system turns out to be nonlinear, it is useful to begin studying the system using summation experiments to obtain some insights as to the nature of the nonlinearities.

A linear characterization of a system is rarely a satisfactory scientific account of the system. There are usually many theoretical questions that require further explanation before the scientist is done. This will be evident in the first section on image encoding. Optical image formation, photoreceptor sampling and color matching are all fundamentally linear and thus we can characterize the performance of these system components. Even when this work is done, we still must explain the measurements in terms of the purpose of these elements and how their properties serve the goals of visual perception.

In part, the emphasis on linear systems methods is my choice. In part this emphasis is inevitable because of a second choice I made in selecting the material. I have tried to include important problems that vision science has solved, or that I think are close to being solved. At present linear methods are much better understood than nonlinear methods. Consequently, we understand those problems which yield to linear analysis much better than we understand nonlinear problems.

While linear analyses are central, there are some significant examples of successful nonlinear analyses. The first example is the analysis of the relationship between color matching and the cone photocurrent treatment in Chapter ???. This system consists of an initial linear encoding followed by a fixed non-linearity. These types of nonlinear systems can also be treated very thoroughly. The review of pattern sensitivity, in Chapter ??, also includes models that begin with linear encodings followed by a nonlinear stage. In the appendix to Chapter ?? I treat the profoundly nonlinear act of classification. Applications of Bayesian classification to interpret image data is likely to be a very important area in the future.

1.5 Selection of Topics: Problems

As I selected problems to review, I did not distinguish strongly between those that are called basic from those that are called applied. I share Edwin Land's frustration with this distinction. After a theoretical lecture on color appearance, Land, who was both a brilliant inventor and entrepreneur, was asked to explain what applied problem his work would solve and he replied quickly that the work had a wonderful application. He then paused while the audience leaned forward to decide whether to invest in Polaroid

stock. If the theory is right, Land whispered confidentially, we'll finally understand what we are doing.

Vision science finds applications in at least three important areas that I will draw on throughout the book. The first area is medicine. If we are to help the blind, we must understand how the visual portion of the brain functions, including the anatomy and functional properties of nerve cells. Equally important, we must understand how information is represented within the brain results in behavior. The results of behavioral experiments can answer questions about the organization of information within the visual pathways that are inaccessible to the anatomist or the electrophysiologist. Together, these results can guide the development of medical diagnostic tools and prosthetic devices. Tom Cornsweet's beautiful book, *Visual Perception*, was a guide to most of my generation as we first learned about the systematic analysis of the visual pathways, ranging from the visual pathways to behavior. In this book I hope to explain to the new student why so many of us found Cornsweet's presentation exhilarating and to build on Cornsweet's review.

A second area of application is the design of computer algorithms capable of analyzing information in an image. Typical applications range from part inspections in a factory to the identification of a tumor in a medical image. David Marr's book, *Vision*, stimulated the interest of many young scientists in this area. He presented a bold overview that related biological concepts and computer algorithms of visual processing. The contrast between the broad scope of Marr's imagination and the elegant, meticulous discussions by Cornsweet captures something of the creative tension that can arise when different disciplines contribute to a broad scientific endeavor.

The third area of application is the design of visual display devices to communicate information to the human visual system. When two electronic components communicate, the components must be designed to accommodate a set of communication protocols. In the case of communication between an electronic display medium, (e.g. a television display) and the human visual system, the designer can only re-design one of the two components. To communicate information efficiently between the electronic system and the human visual system, we must build displays that are matched to human capabilities. A remarkable harmony between vision science and applications technology has been achieved in some areas, such as color science. I hope that this book will contribute to the further coordination of our basic understanding of vision and the design of useful and efficient visual displays.

1.6 The Principles of Vision.

Defining an overview of how vision works can be a lot of fun. This is particularly so when the author takes a chance, as David Marr did, and suggests an entire framework for thinking about the visual pathways. Marr's speculations about zero-crossings, the

primal sketch, the 21/2D sketch, and so forth, gives the reader a framework to organize the many facets of vision science.

Much of vision science is predicated on the principle that the components of the visual system that limit or govern performance in various tasks can be quite different. In some experiments performance is limited by the lens, while in other experiments performance is limited by a computation performed in visual cortex. Different visual tasks may be limited by completely distinct components of the visual pathways. Hence, a static diagram of the visual pathways, in which zero-crossings are inexorably followed by a primal sketch, and so forth, with all the components play the same role across tasks, does not capture the flexibility and adaptability of the visual pathways.

While a static diagram of visual processing does not suit my view of vision, there are several general principles that I found useful as I wrote and organized this volume. Some of these principles are embedded in the organization of the book, repeated in the introductions to the three sections, and repeated within the chapters themselves. This is the time, however, to introduce you to the principles, briefly, in one place.

The Inescapable Components of Image Encoding. The properties of image encoding, such as the blurring by the lens, receptor sampling, and trichromacy, shape the information available to the rest of the nervous system. The first third of the book is devoted to describing these aspects of vision. The properties of image formation set the stage for what the rest of the nervous system must confront.

The limits of image encoding set limits on the image information available to the visual pathways. As we shall see, the image encoding is a very partial description of the light incident at the eye: There is only a narrow region of high visual acuity in the fovea; the dynamic range of the sensors is very small; the representation of wavelength is very coarse. You would never buy a camera with such poor optics and coarse spatial sampling. Yet, the visual algorithms can interpret the properties of objects from this poor encoding.

Whether you wish to study the eye, or study algorithms embedded in the central nervous system, you will not go wrong by studying image encoding and thinking further about its implications for vision.

Adaptation and flexibility. The visual pathways compensate for the poor quality of the image encoding by their flexibility. Nearly all of the peripheral elements of the visual pathways adapt in response to the viewing conditions. The lens accommodates, the strength of the retinal signal varies as the mean illumination level varies, the eye moves to bring the high visual acuity portion of the retina into a favorable viewing position. The flexible responses of the visual system overcome the mediocre image encoding.

The visual system's adjustments, or *adaptations*, to the environment are fundamental to its design. We see adaptation throughout the visual representation, not just in the peripheral components. Because adaptation is so widespread, it is impossible to characterize the visual system as a static device. The ability to adapt in response to a changing environment is a fundamental design principle of the visual pathways, beginning at the earliest stages. Such adaptation is also an important property of central brain representations.

Image representation: Visual streams As we review the visual representation of the image, we will find that the neural pathways are organized into several distinct pathways. These pathways, sometimes called *visual streams*, can be identified based on anatomical studies. Some cells have different shapes from others; some cells send their outputs this way and others send their outputs that way.

Many of the most important discoveries about vision concern the the identification of visual streams. Many of the important contemporary challenges in vision concern explanations of the functional significance of these streams. Segregation of visual information into these visual streams begins with the photoreceptors (rods and cones). Clarifications concerning the visual streams within the optic nerve have revolutionized our understanding of the visual representation. Understanding the organization of visual information with respect to these visual streams is one of most hotly debated topics in modern visual neuroscience. Identifying new visual streams and understanding their function is an important challenge to vision scientists.

Image interpretation: statistical inferences To me, vision science is about how we see things. The interpretations of the image, or as Helmholtz called them, the *unconscious inferences*, are the purpose of vision. I study vision in order to understand the methods of interpreting images to objects and their properties.

Since the retinal image is often ambiguous, the visual system's success in interpreting images must be because it makes good assumptions about the likely properties of objects in the world. Not all configurations of objects are equally likely; we exist in a three-dimensional world. Not all surface reflectance functions are equally likely; there are regularities in the wavelength properties of surfaces and illuminants. Not all types of motion are equally likely; hard objects cannot pass through one another. The unequal probabilities of different interpretations make it possible to make informed guesses about the color, motion, position and shape of objects. The probabilities of different events are sufficiently skewed so that the visual system succeeds at interpreting the image data. Understanding these regularities, and understanding how to use them to interpret the retinal image, is central to vision science.

My devotion to image encoding and representation, the first two parts of this volume, flows from my conviction that we will not understand visual interpretations of the

image without understanding encoding and representation. The encoding and representation define the environment in which image interpretation takes place. The encoding and representation must be structured to permit image interpretation to succeed. As you look through each section of this volume, you will find ideas about image interpretation. The material in this book will seem unified to you if you continue to ask how image encoding and the image representation serve the ultimate goal of image interpretation.