1 Principles and Consequences of the Initial Visual Encoding

Brian Wandell and David Brainard

1.1 Introduction 2
1.2 Scene to Retinal Image 5
1.2.1 Light Field 5
1.2.2 The Incident Light Field 6
1.2.3 Spectral Irradiance and the Plenoptic Function 6
1.2.4 The Initial Visual Encoding 7
1.3 Mathematical Principles 9
1.3.1 Linear Systems 9
1.3.2 Linearity Example: Cone Excitations and Color Matching 10
1.3.3 Matrix Formulation of Linearity 12
1.3.4 Color-Matching Functions 13
1.3.5 Noise in the Sensory Measurements 14
1.3.6 Image Formation 14
1.3.7 Shift-Invariance and Convolution 16
1.4 Computational Model of the Initial Encoding 17
1.4.1 The Value of Computational Modeling 17
1.4.2 Shift-Varying and Wavelength-Dependent Point Spreads 18
1.4.3 Shift-Varying Sampling 19
1.4.4 Spatial Derivatives of the Cone Excitations Mosaic 22
1.5 Perceptual Inference 23
1.5.1 Ambiguity and Perceptual Processing 23
1.5.2 Mathematical Principles of Inference 23
1.5.3 Thresholds and Ideal Observer Theory 26
1.5.4 Computational Observers 30
1.5.5 Image Reconstruction 31
1.5.6 Optimizing Sensory Measurements 34
1.6 Summary and Conclusions 35
1.7 Related Literature 36
Acknowledgments 36
References 36

Only infrequently is it possible to subject the manifold phenomena of life to simple and strict forms of mathematical treatment without forcing the data and encountering contradiction, probably never without a certain abandonment

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

> of the immense multiplicity of details to which those phenomena owe their aesthetic attractiveness. Nevertheless, however, it has often proved to be possible and useful to establish, for wide fields of biological processes and organic arrangements, comparatively simple mathematical formulas which, though they are probably not applicable with absolute accuracy, nevertheless simulate to a certain approximation a large number of phenomena. Such representations not only offer preliminary orientation in a field that at first seems completely incomprehensible, but they also often direct research into a correct course, in as much as first an insight into those fundamental formulations is sought, and then the deviations from their strict validity, which become apparent here and there, are made the subject of special investigations. Among the fields of physiology which have permitted the establishment of such guiding formulas the theory of visual sensations and of color mixture assumes a particularly distinguished position. (von Kries, 1902)

1.1 Introduction

Vision research has many purposes. Medical investigators aim to diagnose and repair visual disorders ranging from optical focus to retinal dysfunction to cortical lesions. Psychologists aim to identify and quantify the systematic rules of perception, including models of visual sensitivity, image quality, and the laws that predict percepts such as brightness, color, motion, size, and depth. Systems neuroscientists seek to relate visual experience and performance to the neural signals in the visual pathways, and computational investigators seek principles and models of perceptual and neural processes. Image systems engineers ask how to design sensors and processing to provide effective artificial vision systems.

Vision science draws upon findings from many fields, including biology, computer science, electrical engineering, neuroscience, psychology, and physics. Clear communication among people trained in different disciplines is not always straightforward. One of the ways that vision science has flourished is by using the language of mathematics to communicate core ideas. Vision science uses many types of mathematics; here we describe methods that have been used for many decades. These are certain linear methods, descriptions of noise distributions, and Bayesian inference. Many other linear methods (e.g., principal components, Fourier and Gabor bases, and independent components analysis) and nonlinear methods (e.g., linear–nonlinear cascades, normalization, information theory, and neural networks) can be found throughout the vision science literature. For this chapter, we focus on a few core mathematical methods and the complementary role of computation.

Physics – the field that quantifies the input to the visual system – provides mathematical representations of the light signal and definitions of physical units. The field of physiological optics quantifies the optical and biological properties of the lens. These properties are summarized as a mathematical transformation that maps the physical stimulus to the image focused on the retina, generally

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

referred to as the retinal image. At each retinal location the image is characterized as the spectral irradiance (power per unit area as a function of wavelength). Retinal anatomy and electrophysiology identify the properties of the rod and cone photoreceptors, enabling us to calculate the photopigment excitations from the retinal image using linear algebraic methods.

Perhaps the most famous use of mathematics in vision science is at the intersection of physics and psychology: the laws of color matching formalize the relationship between the physics of light and certain aspects of color appearance. The mathematical principles of color matching are also deeply connected to Thomas Young's biological insight that there are only three types of cone photopigment (Young, 1802). This insight implies a low-dimensional biological encoding of the high-dimensional spectral light. The linear algebraic techniques used to describe the laws of color matching were developed by the mathematician Hermann Grassmann. Indeed, he developed vector spaces in part for this purpose (Grassmann, 1853). The mathematics he introduced remains central to color imaging technologies and throughout science and engineering.

While acknowledging the importance of mathematical foundations, it is also important to recognize that there is much to be gained by building computational methods that account for specific system properties. The added value of computations is clear in many different fields, not just vision science. The laws of gravity are simple, but predicting the tides at a particular location on earth is not done via analytic application of Newton's formulas. Similarly, that color vision is three-dimensional is a profound principle, yet precise stimulus control requires accounting for many factors, such as variations of the inert pigments across the retinal surface (CIE, 2007; Whitehead, Mares, & Danis, 2006) and the wavelength-dependent blur of chromatic aberration (Marimont & Wandell, 1994). The mathematical principles guide, but we need detailed computations to predict precisely how color matches vary from central to peripheral vision.

We hope this chapter helps the reader value principles expressed by equations and computations embodied in software. Establishing the principles first provides a foundation for implementing accurate computations. Historically, our knowledge about vision has been built up by developing principles, testing them against experiments, and combining them with computation; this remains a useful and important approach. Indeed, we believe the goal of vision science includes not only producing models that account for performance and enable engineering advances, but also leveraging those models to extract new principles that help us think about how visual circuits work.

There are competing views: some would argue that large data sets combined with analyses using machine learning provide the best way forward to understanding, and recent years have seen impressive engineering advances achieved with this approach (D. D. Cox & Dean, 2014). We are certainly interested in the performance of such models as a point of departure, but here we emphasize principles and data-guided computational implementations of these principles.

This chapter begins by describing the representation of the visual stimulus, and how light rays in the scene pass through the optics of the eye and arrive at the retina. Next, we explain how the retinal photoreceptors (a) transform the retinal spectral irradiance into photoreceptor excitations, and (b) spatially sample the retinal image. Each of these steps can be expressed by a crisp mathematical formulation. To describe the real system with quantitative precision, we implemented software that models specific features of the scene, optics, and retina (ISETBio; Cottaris *et al.*, 2019, 2020; https://github.com/isetbio/isetbio/wiki), and we illustrate the use of these models in several examples.

The frontiers of vision science use mathematics to understand visual percepts, which provide a useful basis for thought and action. The information provided by light-driven photopigment excitations is used to create these percepts, but knowledge of the excitations alone falls far short of describing visual perception. The brain makes inferences about the external world from the retinal encoding of light, and throughout the history of vision science many investigators have suggested that the role of neural computation is to implement the principles that underlie these inferences. This point was emphasized as early as Helmholtz, who wrote:

> The general rule determining the ideas of vision that are formed whenever an impression is made on the eye, is that such objects are always imagined as being present in the field of vision as would have to be there in order to produce the same impression on the nervous mechanism. (Helmholtz, 1866; English translation Helmholtz, 1896)

Within psychology this idea is called unconscious inference, a phrase that emphasizes that we are not aware of the neural processes that produce our conscious experience, an idea that was important to Helmholtz. Perhaps more important in this context is the principle that the percepts represent critical properties of external objects in the field of view, such as depth, reflectance, shape, and motion.

The mathematics of perceptual inference can take many forms, and in common scientific practice the mathematics of inference depend on what is known about the input signal. If the scene properties are not uniquely determined by the sensory measurements, such as when only three spectral classes of cones sample the spectral irradiance of the retinal image, probabilistic reasoning about the likely state of the world is inevitable. In vision science, linear methods combined with the mathematical tools of probabilistic inference are commonly used to understand how the brain interprets the mosaic of photoreceptor excitations to see objects, depth, and color. In the final part of this chapter we close the loop between sensory measurements and perceptual inference by introducing the mathematics of such inferences, focusing on two specific examples relevant to the study of the initial visual encoding. The principles we introduce, however, apply generally.

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

1.2 Scene to Retinal Image

1.2.1 Light Field

Light is the most important visual stimulus.[1] The word light means the electromagnetic radiation that is visible to the human eye.[2] The mathematical representation of light has been developed over many centuries through a series of famous experiments, and these experiments provide several different ways to think about light. Many properties of how light is encoded by the eye can be understood by treating light as comprising rays of many different wavelengths.

In a passage in his 1509 notebook (Da Vinci, 1970), Leonardo da Vinci noted that an illuminated scene is filled with rays that travel in all directions.[3] As evidence, he described a pinhole camera (camera obscura) made by placing a small hole in a wall of a windowless room (Figure 1.1). The wall is adjacent to a brightly illuminated piazza; an image of the piazza (inverted) appears on a wall within the room. Leonardo noted that an image is formed wherever the pinhole is placed, and he concluded that the rays needed to form an image must be present at all of these

Figure 1.1 *Light field geometry. The complete set of rays in the environment is the light field. The rays that arrive at the imaging system, in this figure a large pinhole camera, are the incident light field. If the imaging system includes a lens, rather than just a pinhole, the incident light field is described by the positions and angles of the rays at the lens aperture. Figure reproduced from Ayscough (1755).*

1 Mechanical force on the retina (pressure phosphenes) and injecting current into the retina or brain (electrical phosphenes) can also cause a visual sensation.

2 www.merriam-webster.com/dictionary/light

3 From the section prove how all objects, placed in one position, are all everywhere.

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

positions. Leonardo compared the space-filling light rays to the traveling waves that arise after dropping a rock in a pond.

The Russian physicist, Andrey Gershun, provided a mathematical representation of the geometry of these rays, which he called the light field (Gershun, 1939). The mathematical representation of the light field quantifies the properties of the light rays at each position in space [Equation (1.1)]. Each ray travels from a location (x, y, z) in a direction (α, β) and has a wavelength and polarization (λ, ρ). To know these parameters and the intensity of every ray is to know the light field at a given moment in time:

$$LF(x, y, z, \alpha, \beta, \lambda, \rho). \tag{1.1}$$

The light field representation does not capture some phenomena of electromagnetic radiation such as interference (waves) or the Poisson character of light (photon) absorption by the photoreceptors. Even so, the light field representation provides an excellent model to describe the ways in which light interacts with surfaces, and the geometric description of the light field is important in the mathematics of computer graphics, a technology that is important for illumination engineering, photography, and cinema (Pharr, Jakob, & Humphreys, 2016; Wald *et al.*, 2003, 2006).

1.2.2 The Incident Light Field

An eye – or a camera – records a small subset of the light field, those rays arriving at the pupil or entrance aperture. We call these the incident light field. In Figure 1.1 the dashed and solid lines are the light field and the solid lines are the incident light field. The natural parameterization of the incident light differs from the general light field. We can represent the incident light field using only the position (u, v) and angle (α, β) of the rays at the entrance aperture of the imaging system:

$$ILF(u, v, \alpha, \beta, \lambda, \rho, t). \tag{1.2}$$

Equation (1.2) also represents time (t) explicitly, which allows it to describe effects of motion both in the scene and by the eye.

1.2.3 Spectral Irradiance and the Plenoptic Function

The eye and most cameras do not measure the full incident light field. Rather, the rays are focused to an image at the retina or sensor, and the photodetectors respond to the sum across all directions of the image rays. To be explicit about this, Adelson and Bergen (1991) introduced the term plenoptic function, a simplified version of the incident light field, that was chosen to guide thinking about the computations carried out in the human visual pathways [their Equation (2)]. First, they approximated the eye as a pinhole camera; with this approximation all rays have the same entrance position **p**. Additionally, the retina/sensor surface defines the direction (**d**) of the rays that pass through the pinhole. For the pinhole case,

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

specifying two angles of a ray at the pinhole is equivalent to specifying the location where a ray will intersect the retina/sensor surface, (r_x, r_y). Finally, Adelson and Bergen ignored polarization as unimportant for human perception. With these restrictions, the plenoptic function for human vision is simply the retinal spectral irradiance, over time (t):

$$E(r_x, r_y, \lambda, t; \mathbf{p}, \mathbf{d}). \tag{1.3}$$

In Equation (1.3) we have explicitly reintroduced position and direction, but these are often implicit [as in the formulation of Equation (1.2) above]. Understanding the progression from light field to incident light field to retinal spectral irradiance is useful for understanding how the information available for visual processing relates to the complete set of potential information that could be sensed by a visual system.

Adelson and Bergen note that by placing the pinhole at many different positions and viewing directions, we can estimate the full light field from the set of spectral irradiances. It is possible to be more efficient and estimate the incident light field by using a lens, rather than a pinhole, inserting a microlens array over the photodetector array and placing multiple detectors behind each microlens. Both cameras and microscopes have employed this technology to support depth estimation (Adelson & Wang, 1992) and control focus and depth of field in post-processing (Ng *et al.*, 2005). Cameras that estimate the full incident light field are not currently in wide use (Wikipedia contributors, 2021); but, the widely used dual pixel autofocus technology obtains a coarse measure of the incident light field (Canon U.S.A., Inc., 2017; Mlinar, 2016). This is accomplished by inserting a microlens array over pairs of photodetectors. With this design rays from, say, the left and right sides of the lens are captured by adjacent detectors. This coarse estimate of the light field is useful for setting the lens focus and estimating depth.

1.2.4 The Initial Visual Encoding

Computational models of the early visual pathways define a series of transformations that characterize how the incident light field becomes a neural response. In this chapter, we introduce the mathematics used to characterize the initial visual encoding in the context of the first few of these transformations (Figure 1.2; see also Brainard & Stockman, 2010; Packer & Williams, 2003; Rodieck, 1998; Wandell, 1995). We focus on the encoding of the spectral radiance by the photoreceptors – subsequent neural processing operates on this visual encoding.

A visual scene's light field is generated by the properties and locations of the light sources and objects, and how the rays from the light sources are absorbed and reflected by the objects. Here we consider the special case of scenes presented on a flat display, so that in the idealized case where the display is the only object and there are no other light sources, the full light field is determined just by the spectral radiance emitted at each location of the display. Elsewhere, we consider

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

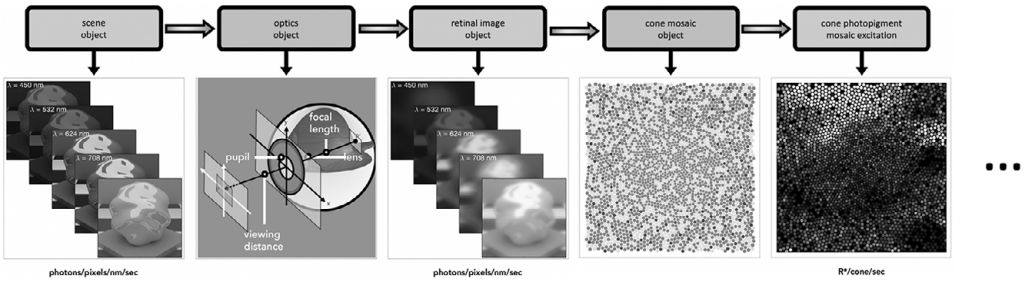

Figure 1.2 *The initial encoding of light by the visual system.* Scene*: An image on a display surface is characterized by the spectral radiance at each display location. Images of the display spectral radiance are shown at a few sample wavelengths, along with a rendering of the image.* Optics*: The incident light field enters the pupil of the eye and a spectral irradiance image is formed on the retina. The* retinal image *is blurred relative to the displayed image, and the spectral irradiance is affected by lens and macular pigment absorptions.* Cone mosaic*: The retinal image is spatially sampled by the L-, M-, and S-cone mosaics.* Cone excitations*: The retinal image irradiance, spectrally weighted by each cone photopigment absorptance function, is integrated within the cone's aperture and temporally integrated over the exposure duration to produce a pattern of cone excitations. This figure should be viewed in color. The color version is available at https://color.psych.upenn.edu/supplements/earlyencoding/computationsColorFig.pdf. We thank Nicolas Cottaris for the figure.*

the more general case of modeling the formation of the retinal spectral irradiance, given a description of the light sources and objects in a three-dimensional scene (Lian *et al.*, 2019).

The optics of the eye collect the incident light field and focus the rays to produce the spectral irradiance arriving at the retina. Factors such as diffraction and aberrations in the eyes optics mean that this image is blurred relative to the displayed image. In addition, wavelength-selective absorption of short-wavelength light by the lens and inert macular pigment also affect the spectral irradiance. Of note (but not illustrated in Figure 1.2), the density of the macular pigment is high in the central area of the retina and falls off rapidly with increasing eccentricity.

Photoreceptors spatially sample the retinal image. Excitations of photopigment molecules in these photoreceptors provide the information available to the visual system for making perceptual inferences about the scene. Here we consider the cone photoreceptors, which operate at light levels typical of daylight. There are three spectral classes of cones, each characterized by its own spectral sensitivity. That there are three classes leads to the trichromatic nature of human color vision. Figure 1.2 illustrates a patch of cone mosaic from the central region of the human retina. The properties of the mosaic are quite interesting. For example, there are no S-cones in the very center of the retina, and many properties of the

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

mosaic (e.g., cone density, cone size, cone photopigment optical density) vary systematically with eccentricity (Brainard, 2015; Hofer & Williams, 2014).

Not considered here is a separate mosaic of highly sensitive rod photoreceptors that is interleaved with the cone mosaic. The rods mediate human vision at low light levels (Rodieck, 1998). We also ignore the melanopsin containing intrinsically sensitive retinal ganglion cells (Gamlin *et al.*, 2007; Hattar *et al.*, 2002; Van Gelder & Buhr, 2016). The principles we develop, however, also apply to modeling the excitations of these receptors.

Modeling of the initial visual encoding is well understood, and we explain the key linear systems principles next, using a simplified representation of the light stimulus. Advanced modeling of the subsequent neural processes includes non-linearities; the mathematical principles and computational methods we introduce are a fundamental part of the full description. After explaining the mathematical principles, we illustrate how to extend them through computational modeling that harnesses the power of computers to characterize biological reality in more detail than is possible with analytic calculations alone.

1.3 Mathematical Principles

1.3.1 Linear Systems

Linear systems and the tools of linear algebra are the most important mathematical methods used in vision science. Indeed, when trying to characterize a system, the scientist's and engineer's first hope is that the system can be approximated as linear. A system, L, is linear if it follows the superposition rule:

$$L(x + y) = L(x) + L(y). \tag{1.4}$$

Here x and y are two possible inputs to the system and $x + y$ represents their superposition. The homogeneity rule of linear systems follows from the superposition rule. Consider that

$$\begin{aligned} L(x + x) &= L(2x) \\ &= L(x) + L(x) \\ &= 2L(x). \end{aligned}$$

This is easily generalized for any integer m to show that:[4]

$$L(mx) = mL(x). \tag{1.5}$$

4 It is an exercise for the reader to show that a system that follows the superposition rule also obeys the homogeneity rule, not just for integers, but for any real scalar. If x is a real-valued scalar, homogeneity also implies superposition. When $\mathbf{x}$ is a real-valued vector with entries x_n, however, a system can obey homogeneity but not superposition. For example, $f(\mathbf{x}) = \sqrt[3]{\sum x_n^3}$ satisfies homogeneity but not superposition. The reader may find it of interest to consider why we used an exponent of three rather than two for this example.

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

No physical system can be linear over an infinite range – if you put enough energy into a system it will blow up! But many systems are linear over a meaningful range of input values.

1.3.2 Linearity Example: Cone Excitations and Color Matching

Vision is initiated when a photopigment molecule absorbs a photon of light. The absorption can cause the photopigment, a protein, to change conformation, an event we refer to as a photopigment excitation. The excitation initiates a molecular cascade inside the photoreceptor that changes the ionic currents at the photoreceptor membrane. The change in current modulates the voltage at the photoreceptor synapse and causes a release of neurotransmitter (Rodieck, 1998).

The transformation from the spectral energy of light, $E(\lambda)$, incident upon a cone to the number of photopigment excitations, n, produced by that light is an important, early vision, linear system. Consider two different spectra, denoted by $E_1(\lambda)$ and $E_2(\lambda)$. Let L represent the system that describes the transformation between spectra and excitations. This system obeys the superposition rule:

$$L(E_1 + E_2) = L(E_1) + L(E_2). \tag{1.6}$$

This linearity holds well over a wide range of light levels typical of daylight natural environments (Burns *et al.*, 1987).

An important feature of photopigment excitations is that their effect on the membrane current and transmitter release does not differ with the wavelength of the exciting photon. Such differences might have existed because different wavelengths are preferentially absorbed at different locations within the cone outer segment, or because photons of different wavelengths carry different amounts of energy. The observation that all excitations have the same impact is called the Principle of Univariance. As Rushton wrote:

> The output of a receptor depends upon its quantum catch, but not upon what quanta are caught. (Rushton, 1972)

The color-typical human retina contains three distinct classes of cones, which are referred to as the L (long-wavelength sensitive), M (middle-wavelength sensitive), and S (short-wavelength sensitive) cones. While the effects of photopigment excitations are univariant, the probability of a photopigment excitation is wavelength-dependent. The wavelength-dependent probability that an incident photon leads to an excitation is characterized by the pigment's spectral absorptance.[5] The absorptance depends on the density of the photopigment within the

5 The absorptance spectrum is the probability that a photon is absorbed. Not all absorbed photons lead to an excitation, so an additional factor specifying the quantal efficiency (probability of excitation given absorption) needs to be included in the calculation. Current estimates put the quantal efficiency of human cone photopigment near 67%. In addition, the calculation of cone excitations from spectral irradiance requires taking into account the size of the cone's light-collecting aperture.

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

cone's outer segment, as well as on the outer segment length; details are elaborated elsewhere (Rodieck, 1998; see also Packer & Williams, 2003; Pugh, 1988).

It is difficult to measure the light at the retinal surface in the living eye, but it is straightforward to measure the light incident at the cornea. Hence, it is typical to specify the absorptance with respect to the spectrum of the light incident at the cornea. This convention effectively combines the effects of the lens, the inert retinal macular pigment, the photopigment absorptance, and quantal efficiency. For simplicity, vision scientists call the cornea-referred spectral excitation curve the cone fundamental.

The three (L-, M-, and S-) cone fundamentals define for each cone type the probability of excitation given the spectrum of light entering the eye. The human cone fundamentals have been carefully measured and tabulated (Figure 1.3; Stockman & Sharpe, 2000; Stockman, Sharpe, & Fach, 1999; www.cvrl.org) and are the subject of an international standard (CIE, 2007).

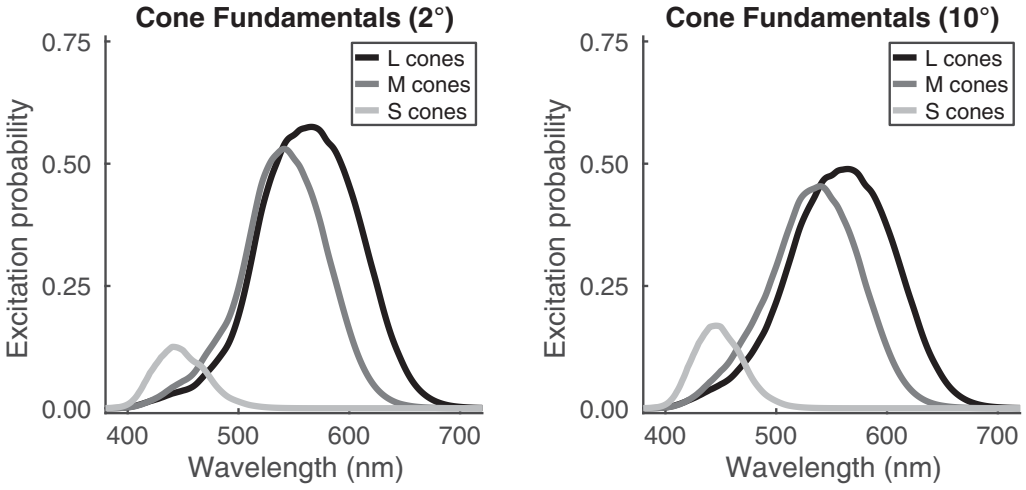

Figure 1.3 *Human cone fundamentals. The left panel shows estimates of the L-, M-, and S-cone fundamentals for foveal viewing. The fundamentals are the probability of excitation per photon entering the cone's entrance aperture, but with pre-retinal absorption taken into account. Note the large difference between the L- and M-cone fundamentals compared to the S-cone fundamental. This difference is due partly to the selective absorption of short-wavelength light by the lens and macular pigment. The right panel shows estimates for cones at 10° eccentricity. The S-cone fundamental is relatively higher at 10°, because there is little or no macular pigment at that eccentricity; and for the same reason there is a slight change in the relative values of the L- and M-cone fundamentals. In addition, the cone outer segment lengths decrease with eccentricity, leading to the lower peak probability of excitation in the periphery. This reduction, however, is more than compensated for by an increase in the size of the cone apertures with eccentricity. The impact of the aperture is not shown in these plots, but see Figure 1.4.*

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

To compute the number of cone excitations we use linear formulas. Suppose that a cone's fundamental is given by $C(\lambda)$. Using linearity and continuous mathematics, we compute the number of excitations at a single location as

$$N(r_x, r_y) = \int C(\lambda)E(r_x, r_y, \lambda)d\lambda. \tag{1.7}$$

The discrete form of this integral, commonly used in computational methods, is the inner product of the cone fundamental with the cornea-referred spectral irradiance incident upon a retinal location:[6]

$$N(r_x, r_y) = \sum_{\lambda_i} C(\lambda_i)E(r_x, r_y, \lambda_i)\Delta\lambda. \tag{1.8}$$

Here the λ_i are a set of w discretely sampled wavelengths, and $\Delta\lambda$ is the wavelength sample spacing.

1.3.3 Matrix Formulation of Linearity

We can calculate cone excitations by a matrix multiplication. The matrix $\mathbf{C}$ combines the three discretized cone fundamentals $C_L(\lambda_i)$, $C_M(\lambda_i)$, and $C_S(\lambda_i)$ into its rows, so that its dimension is $3 \times w$. Similarly, we write the spectral irradiance at a position, $E(r_x, r_y, \lambda)$, as a $w \times 1$ vector $\mathbf{e}(r_x, r_y)$. The L-, M-, and S-cone excitations available at a retinal location are described by a three-dimensional column vector:

$$\mathbf{n}(r_x, r_y) = \mathbf{C}\mathbf{e}(r_x, r_y). \tag{1.9}$$

The vector field $\mathbf{n}(r_x, r_y)$ describes the potential information available to the visual system from the cones at a moment in time. This representation replaces the dependence of the spectral irradiance on wavelength with the excitations of the three classes of cones. As we describe in more detail below, not all of this potential information is sensed by the visual system, since the cones discretely sample $\mathbf{n}(r_x, r_y)$.

It is worth reflecting on the implication of the linearity expressed by Equation (1.9). If we measure the cone fundamentals at each of the sample wavelengths λ_i, we can predict the cone excitations to any spectrum $E(r_x, r_y, \lambda_i)$. Thus, linearity implies that we can compute the system response to any input after making enough measurements to determine the system matrix $\mathbf{C}$. The ability to delineate the set of measurements required for complete system characterization is an important consequence of linearity, and this observation applies to linear systems in general, not just to computation of cone excitations.

A second implication of Equation (1.9) concerns which spectral radiances appear to be the same; these pairs are called metamers. Young (1802) had proposed that metamers arise if two lights produce the same set of cone excitations. This

6 In this formulation, we do not make the spatial extent of the cone acceptance aperture explicit. This aperture introduces additional blur into the retinal image. Computational models (see Figure 1.7) account for this factor; it is significant.

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

implies that the difference between a metameric pair is in the null space of the matrix, $\mathbf{C}$.[7] That is, $\mathbf{e_1}$ and $\mathbf{e_2}$ must satisfy

$$\begin{aligned} \mathbf{Ce_1} &= \mathbf{Ce_2} \\ \mathbf{0} &= \mathbf{C}(\mathbf{e_2} - \mathbf{e_1}). \end{aligned} \tag{1.10}$$

Wyszecki (1958; see also Wyszecki & Stiles, 1982) referred to vectors in the null space of $\mathbf{C}$ as metameric black spectra. Adding a nonzero metameric black to any spectrum produces a metamer.

Displays and printers do not reproduce the original physical stimulus; rather, they create lights designed to be metamers to the original. Thus, calculating metamers is central to color reproduction technologies. Practical aspects of the computation of metamers for color reproduction applications, including limitations based on the spectra a device can produce, are discussed in detail elsewhere (Brainard & Stockman, 2010; Hunt, 2004).

1.3.4 Color-Matching Functions

James Clerk Maxwell (1860) was the first to measure pairs of spectral irradiance functions, $\mathbf{e_1}$ and $\mathbf{e_2}$, that appear the same to humans despite being physically different. These data place constraints on estimates of the matrix $\mathbf{C}$, but do not uniquely determine it. To understand why, note that the null space of $\mathbf{C}$ is the same as the null space of $\mathbf{T} = \mathbf{MC}$, for any invertible 3×3 matrix $\mathbf{M}$. Thus, any such matrix $\mathbf{T}$ predicts the same set of matches.

The rows of $\mathbf{T}$, when viewed as functions of wavelength, are referred to as a set of color-matching functions. We say that the color-matching functions are only unique up to a linear transformation. The technology for creating metamers relies on color-matching functions which were chosen as an international standard (CIE, 1986, 2007). How color-matching functions may be obtained directly from perceptual color-matching experiments, without explicit reference to the cone fundamentals, is treated in many sources (Brainard & Stockman, 2010; Wandell, 1995; Wyszecki & Stiles, 1982). Indeed, high-quality measurements of behavioral color matching (e.g., Stiles & Burch, 1959) provide key data that constrain modern estimates of human cone fundamentals.

There are a number of properties of the eye that must be modeled if we are to compute a true estimate of cone mosaic excitations. For example, only one type of cone is present at each position, so we must specify a cone spatial sampling scheme. That is the reason that we use the term *potential information* to describe cone excitations $\mathbf{n}(r_x, r_y)$ as a function of retinal location – not all of that information is sampled by the cone mosaic. Also, as noted above, the density of both inert pigments and photopigments varies with retinal location, as does the size of the

7 The null space of a matrix $\mathbf{C}$ is the space of vectors $\mathbf{v}$ such that $\mathbf{Cv} = 0$. If a matrix has column dimension n and rank r, its null space has dimension $n - r$.

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

cone apertures. Enough is known about these properties to enable us to compute a reasonable approximation to the cone mosaic excitations across the retina.

1.3.5 Noise in the Sensory Measurements

Measurement noise is fundamental in the physical sciences and engineering. Two types of noise are used throughout the sensory sciences: Gaussian (normal) noise and Poisson noise. Gaussian noise has two parameters (a mean and variance) but the Poisson distribution has a single parameter (the Poisson mean equals its variance). The formulas for the Gaussian density function and Poisson probability mass function, along with example draws from these distributions, are provided in Figure 1.4.

The Gaussian and Poisson distributions can be compared by setting the Gaussian mean equal to its variance. For small values, the Gaussian has values below zero. As the Poisson mean increases, the matched Gaussian is extremely similar (Figure 1.4).

There is an important conceptual difference between how these noise distributions are used in applications. There are many theorems about additive Gaussian noise, and thus it is common to introduce noise in a model with such noise using a fixed mean (μ) and standard deviation (σ). The added noise has the same distribution for all values of the signal (signal-independent noise).

For typical sensor measurements, including the cone excitations, the noise depends on the signal. Specifically, for the cones and many other measurement devices, the noise is Poisson distributed, with the Poisson parameter equal to the mean number of excitations (signal-dependent noise). The difference between signal-independent and signal-dependent noise can be quite significant (Figure 1.4).

1.3.6 Image Formation

The linear system principles described for one-dimensional spectral functions can be extended to two-dimensional functions, such as images. We use linear system methods to analyze how the cornea and lens form the retinal image. An important, but simple, case occurs for an image confined to a plane, such as a visual display or an optometrist's eye chart. For such images we can estimate the spectral irradiance at the cone apertures using a two-dimensional linear system computation.

The image emitted from a visual display is a function of position (x, y) and wavelength λ. As a first approximation, the display emits the same density of rays over a wide angle, which is why the display appears to be approximately the same when seen from different positions. The image from the display is called the spectral radiance, $I(x, y, \lambda)$, and it has units of W/sr/m^2/nm.

The spectral irradiance at the retina, $E(r_x, r_y, \lambda)$, is formed from the cone of rays that are captured by the pupil. In this case, r_x and r_y specify retinal location and the units of the image are those of spectral irradiance, W/m^2/nm, which result from integration over the solid angle of the pupil.

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

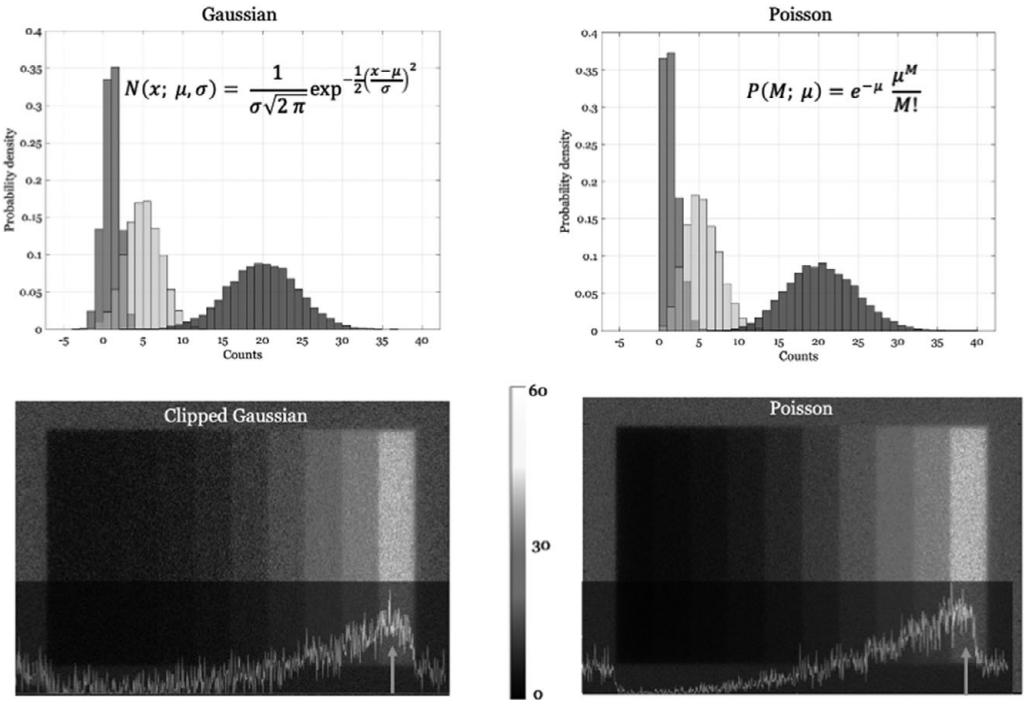

Figure 1.4 *The number of cone excitations is inescapably noisy, following a signal-dependent Poisson distribution. (Top) For mean values greater than 10, the Poisson distribution is reasonably approximated by a Gaussian distribution with a mean equal to the variance. For smaller values, it is necessary to clip the negative values for the Gaussian to achieve a good approximation. Low excitation rates are common under low-light conditions and for nearly all conditions when assessing the S-cones and rods (Baylor* et al.*, 1979; Hecht, Schlaer, & Pirenne, 1942). (Bottom) The signal-dependent nature of Poisson noise is important; simply adding Gaussian noise with a fixed mean is not a good approximation if there is a substantial range in the mean excitation values. The images illustrate the excitations in response to a series of bars spanning a large range of mean excitation using a signal-independent clipped Gaussian noise (left) and a Poisson noise (right). The Gaussian distribution added to the signal has zero mean and variance equal to the number of excitations in the brightest bar (arrows); this approximates Poisson noise for that bar. The inset trace, which shows excitations across a row of the image, illustrates that the Gaussian noise is too large for the dark bars. Had the variance been set to match the noise at the dark bar, the clipped Gaussian would be too small for the brightest bar. The simulation was created for an array of M-cones in the central fovea, a 2 ms exposure duration, achromatic bars of increasing intensity, and a bright bar luminance of 300 cd/m^2. This figure should be viewed in color. The color version is available at https://color.psych.upenn.edu/supplements/earlyencoding/noiseColorFig.pdf.*

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

The key linear system idea [Equation (1.6)] holds for retinal image formation (Wandell, 1995). If two input images, $I_1(x, y, \lambda)$ and $I_2(x, y, \lambda)$, produce two retinal images, $E_1(r_x, r_y, \lambda)$ and $E_2(r_x, r_y, \lambda)$, then the superposition of the input images, $I_1(x, y, \lambda) = I_1(x, y, \lambda) + I_2(x, y, \lambda)$, produces the superposition of the retinal images:

$$E(r_x, r_y, \lambda) = E_1(r_x, r_y, \lambda) + E_2(r_x, r_y, \lambda). \tag{1.11}$$

It follows that if the input image is the weighted sum of two input images, $I(x, y, \lambda) = \alpha I_1(x, y, \lambda) + \beta I_2(x, y, \lambda)$, the output retinal image will be the weighted sum of the two corresponding retinal images:

$$E(r_x, r_y, \lambda) = \alpha E_1(r_x, r_y, \lambda) + \beta E_2(r_x, r_y, \lambda). \tag{1.12}$$

As noted above, an important consequence of linearity is that it tells us how to generalize. When we know the response to an image I_k, measuring the response to a second image, I_j, enables us to predict the responses to an entire class of new images, all images of the form $\alpha I_k + \beta I_j$.

1.3.7 Shift-Invariance and Convolution

To characterize color matching we used the fact that a discrete linear system may be expressed as a matrix multiplication [Equation (1.9)]. A matrix can also be used to express retinal image formation, but in this case the number of measurements required to determine the requisite matrix is very large. For this reason, we consider an additional special and simplifying property linear systems can have: shift-invariance. These are linear systems such that shifting the position of the input correspondingly shifts the position of the output, without changing its form.[8] It is possible to measure whether a system is shift-invariant by a simple experiment. For an input image, say $I(x, y, \lambda)$, measure the retinal image $E(r_x, r_y, \lambda)$. Then shift the input, $I(x - \delta x, y - \delta y, \lambda)$, and measure the retinal image again. If for all choices of $(\delta x, \delta y)$ in the image domain, the output is shifted equivalently, $E(r_x - \delta r_x, r_y - \delta r_y, \lambda)$ in the retinal image domain, then the system is shift-invariant. Here the retinal image shifts $(\delta r_x, \delta r_y)$ differ from their image counterparts $(\delta x, \delta y)$ by the factor that converts the positional units of the image to those of the retinal image.

We can express linearity and shift-invariance using the convolution formula. For simplicity, we choose one wavelength and suppress λ. Suppose $P(r_x, r_y)$ is the retinal image from an image that is just a single point. The image $P(r_x, r_y)$ plays a central role in the characterization of convolutional optical systems: it is called the point spread function.[9] The point spread function is all we need to compute the retinal image for any input image. The idea is to treat the input image as a set of points, and to add shifted copies of the point spread function, each weighted by the input image intensity:

8 When describing optics, a shift-invariant region within the visual field is called an isoplanatic region.

9 The point spread function is the spatial analog of the impulse response function used to characterize time-invariant linear systems.

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

$$E(r_x, r_y) = \int_u \int_v I(u,v)P(r_x - u, r_y - v)dudv. \quad (1.13)$$

The importance of linear shift-invariance is that we characterize the system fully by one measurement, $P(r_x, r_y)$. We use the convolution formula and this measurement to compute the responses of a linear shift-invariant system to any input.

While shift-invariance and convolution are important concepts, the eye's optics deviate significantly from this ideal. Shift-invariance is a good approximation of human retinal image formation in local regions, say spanning a few degrees of visual angle and a change in wavelength of 20–50 nm. Properties of the photoreceptor sampling mosaic further limit the accuracy of the shift-invariant approximation of the visual encoding (see Figure 1.6). Thus the convolutional approximation is helpful for thinking about encoding over small regions, but it is not an accurate depiction when one considers a larger field of view. A realistic approximation requires computational modeling.

1.4 Computational Model of the Initial Encoding

The mathematical principles described above tell us how to compute the retinal image and the noisy cone excitations from a displayed image; the calculations are straightforward for a single retinal location. But an accurate model of the visual system must account for variations in the optics, pigments, and sampling properties of the cone mosaic with visual field location. These are substantial and impact the information available to the brain for making perceptual inferences about the visual scene. Parameters with significant spatial variation across the visual field include the optical point spread function, density and size of the cones in the mosaic, the distribution of different cone types within the overall mosaic, and the cone fundamentals. To make a realistic calculation requires implementing a computational model of the visual transformations.

1.4.1 The Value of Computational Modeling

Carefully validated computer simulation of the initial visual encoding has the potential to support advances in understanding many aspects of visual function. We use image-computable models to build upon the mathematical characterizations – earned through 400 years of experimental and theoretical work in vision science – and estimate the initial visual signals. Such knowledge is an essential foundation to use when modeling less well understood visual processes. The models help us separate effects attributable to known factors of the initial encoding from effects of factors that arise in later processing. For example, understanding cortical visual processing requires representing the input to the cortex. Without accurate modeling of the input, we risk attributing features of the cortical signals to the wrong neural mechanisms.

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

Because of the central role computational modeling plays in understanding vision, we have invested in developing a set of freely available software tools to model retinal image formation and cone excitations (Image Systems Engineering Tools for Biology – ISETBio; https://github.com/isetbio/isetbio.git; Cottaris *et al.*, 2019, 2020). The tools can be used for images presented on planar displays and for full three-dimensional descriptions of the objects and light sources in the scene (Image Systems Engineering Tools 3D; https://github.com/iset/iset3d.git; Lian *et al.*, 2019). In this section we briefly illustrate some basic calculations enabled by ISETBio. We are not advocating for our implementation in particular, but we do believe that the field needs to develop trusted open-science tools for computational modeling.

1.4.2 Shift-Varying and Wavelength-Dependent Point Spreads

The point spread functions from a single subject, measured at different retinal locations and wavelengths, differ significantly (Figure 1.5). The variation with retinal location occurs because the optical aberrations depend on the direction of the rays incident at the retina. The ISETBio tools can explicitly represent the full

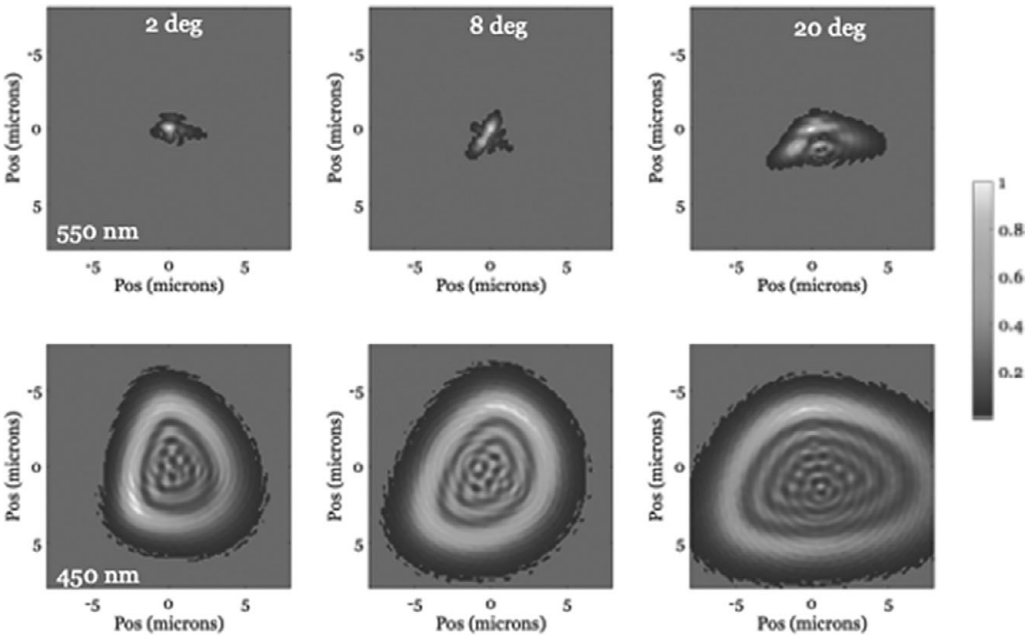

Figure 1.5 *The human point spread function. The images in the top row show the point spread functions at 550 nm from a typical subject measured at three different visual eccentricities. The point spread increases with eccentricity. The bottom images show the point spread but for light at 450 nm. The human eye cannot focus these two wavelengths at the same time because the index of refraction in the lens and cornea is wavelength-dependent. For many people, chromatic aberration is the largest aberration. A diagram showing simple ways to estimate degrees of visual angle is available from Branwyn (2016).*

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

incident light field and calculate these effects from a model eye (Lian *et al.*, 2019). Improvement of eye models is an active area of investigation, and in some cases ISETBio relies on empirical measurements of the eye's optics to predict responses over a range of retinal field locations (Jaeken & Artal, 2012; Polans *et al.*, 2015).

The point spread function varies with pupil diameter and wavelength in addition to visual field position. The dependence on pupil diameter, which varies with the light level of the scene, occurs for two reasons. As the pupil opens, the aberrations vary because more of the imperfectly shaped corneal and lens surfaces refract the light. As the pupil closes, diffraction starts to be a significant factor. The wavelength dependence is explained by the refractive indices of the cornea and lens. These chromatic aberrations are the largest of all the aberrations (Thibos *et al.*, 1990; Wandell, 1995).

1.4.3 Shift-Varying Sampling

Figure 1.6 shows the spatial arrangement of cones at different locations within the retina. The cone density is highest in the central fovea where the cones are tightly packed. Moving away from the center, cone density falls off and the cone apertures become larger. As cone density decreases, rod photoreceptors (the smaller receptors in the peripheral images) appear and fill the gaps between the cones. In addition, not apparent in the figure, cones become shorter away from the fovea. The shortening reduces the spectral absorptance.

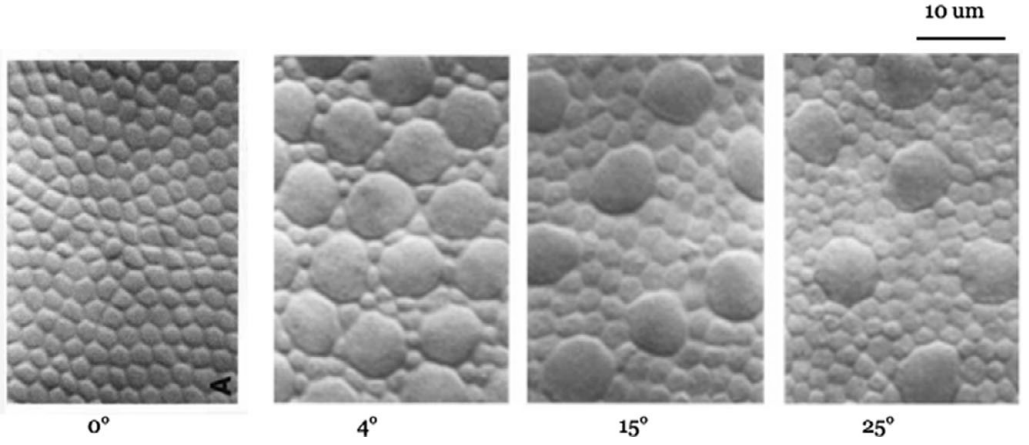

Figure 1.6 *Human cone and rod sampling mosaics. The en face images show the photoreceptor inner segments, where light enters the cones, at four retinal eccentricities. In the central region, all of the receptors are cones. At 4° and beyond, the large apertures are the cones and the smaller apertures are the rods. The cone sampling density and cone aperture sizes differ substantially between the central fovea and other visual eccentricities. The reduced sampling density limits the spatial resolving power of the eye. The larger cone apertures increase the rate of photon excitations per cone. Scale bar is 10 μm. Recomposited from figures in Curcio* et al. *(1990).*

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

The sampling density reduction means that less spatial information about the retinal image is extracted at retinal locations away from the central fovea. The relative density of the different cone types also varies with eccentricity. Indeed, as noted above, there are no S-cones in the very central fovea (Williams, MacLeod, & Hayhoe, 1981), so that vision in this small retinal region is dichromatic rather than trichromatic. Perhaps this region is specialized for high-resolution vision and omitting a few S-cones, which see a blurry retinal image at short wavelengths because of the chromatic aberrations, maximizes the information transmitted to the brain about spatial structure (Brainard, 2015; Garrigan *et al.*, 2010; Hofer & Williams, 2014; Williams *et al.*, 1991; Zhang, Cottaris, & Brainard, 2021).

The impact of the cone size and density, along with variations in the inert pigments described above, mean that calculating the cone excitations is shift-varying: the calculation is linear, but the parameters change with eccentricity. These eccentricity-dependent calculations are included in the ISETBio simulations. There is little value in expressing the full complexity of these calculations in pure mathematical form.

The impact of the several eccentricity-dependent factors on the cone excitations is substantial and illustrated in Figure 1.7. The images in the left column illustrate calculations in the central fovea and the images in the right column illustrate the same calculations at 10° in the periphery. The top image shows the differences in the size and density of the cone photoreceptor apertures. Also, notice the absence of S-cones in the small region of the very central fovea. The images inset in the top show the size of the point spread function for an in-focus wavelength: there are many more cones within the foveal point spread than within the 10° point spread.

The images in the middle row represent the number of cone excitations in response to a relatively low-frequency grating pattern. There are more excitations per cone at 10° than in the fovea, and there are many more cones representing the stimulus in the fovea. The third row shows the effect of increasing the stimulus spatial frequency. The foveal mosaic samples densely enough to preserve the regular pattern, but at 10° the spatial samples look like a wobbly representation of the stimulus.

Finally, notice that many cones have relatively low excitation levels to this achromatic stimulus. These cones appear as the quasi-regular array of black dots that are easy to see at 10°. They are also present, but harder to see, in the excitations for the central location. These cones are the S-cones, which absorb many fewer photons than the L- and M-cones. This lower excitation rate is partly due to the spectral transmission of the lens (and in the central region the macular pigment), which absorbs a great deal of short-wavelength light.

In summary, the principles of linearity and shift-invariance are useful guides for reasoning about cone excitations. These principles were part of our toolkit as we built a specific model of the human eye, and so they would be for any model. However, in the human eye deviations from shift-invariance are substantial. In addition, there are significant differences between people that may be important for explaining between-subject differences. Thus, an essential ingredient for building

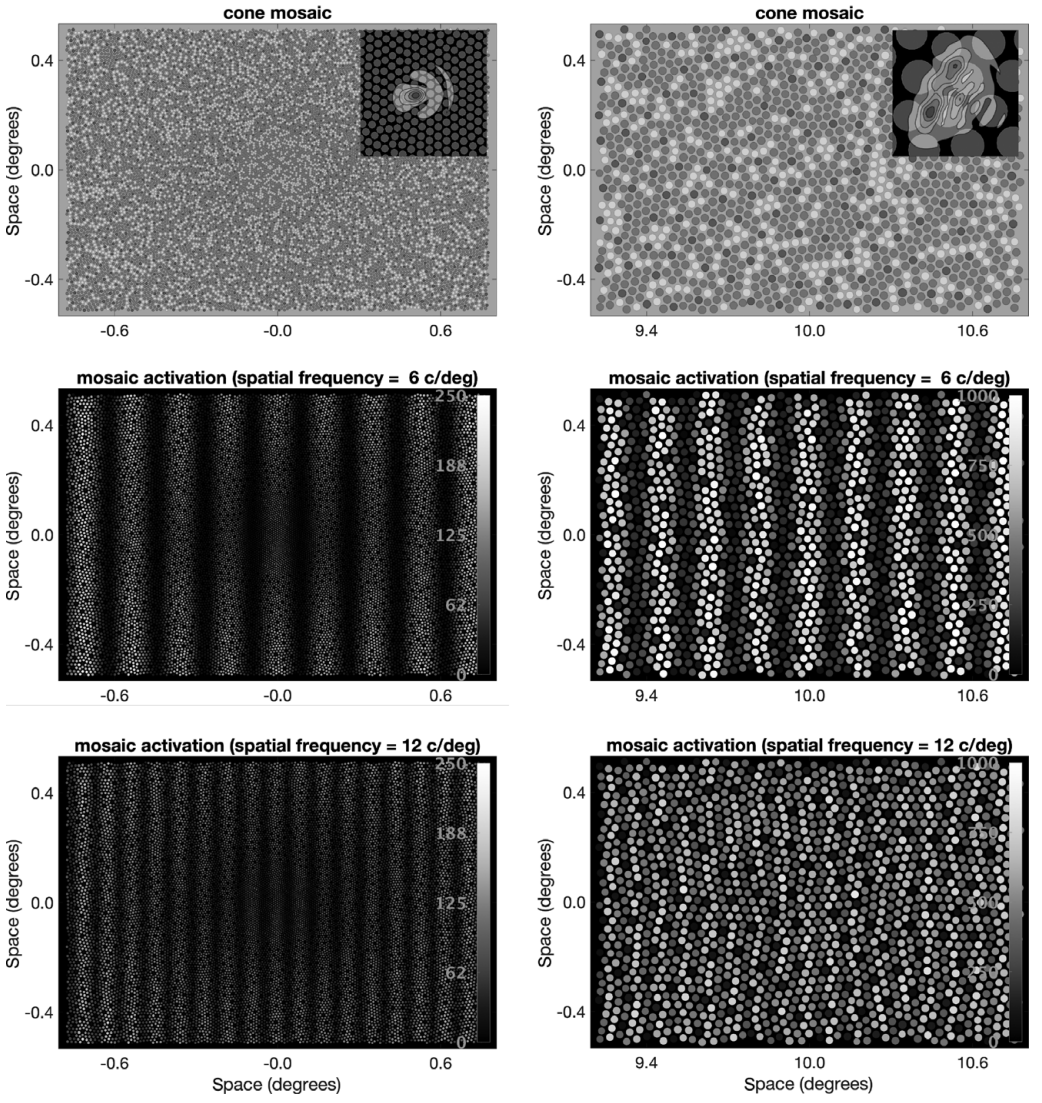

Figure 1.7 *Excitation calculations. The two columns represent two retinal eccentricities, each about* $1^{\circ 2}$. *(Top) The interleaved L-, M-, and S-cone mosaics, shown as red, green, and blue dots, are shown at the top. The inset shows an expanded view of the point spread function in the same region. The rods are not represented. (Middle and bottom) The gray level in these images shows the estimated cone excitations for a 6 c/deg harmonic and a 12 c/deg harmonic. The scale for the foveal location runs between 0 and 250, while that for the peripheral location runs from 0 to 1000. Peripheral cones have more excitations to the same stimulus because the cone apertures are larger. This figure should be viewed in color. The color version is available at https://color.psych.upenn.edu/supplements/earlyencoding/excitationsColorFig.pdf. Figure courtesy of Nicolas Cottaris.*

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

a computational model are data sets that quantify the critical model parameters (e.g., how the optical point spread function and cone density vary with visual field position) and how these parameters vary across individuals. For these reasons, a computational model is essential for applications that aim to create realistic estimates of the cone excitations for a population.

The computational implementation has benefited from mathematical principles and from data collected and shared by many investigators. Conversely, the exercise of building computational models often highlights the need for data sets that do not yet exist (e.g., across individuals, are optical quality and cone density independent, or do they covary in some systematic way?) At this point in the chapter, the reader might find it useful to re-read the quote at the start of this chapter, which was written by von Kries, Helmholtz's greatest disciple (Cahan, 1993), more than a century ago.

1.4.4 Spatial Derivatives of the Cone Excitations Mosaic

Adelson and Bergen (1991) observed that the partial derivatives of the spectral irradiance correspond to computations performed by neurons in the early visual system. Figure 1.8 illustrates these derivatives for several cases: derivatives with

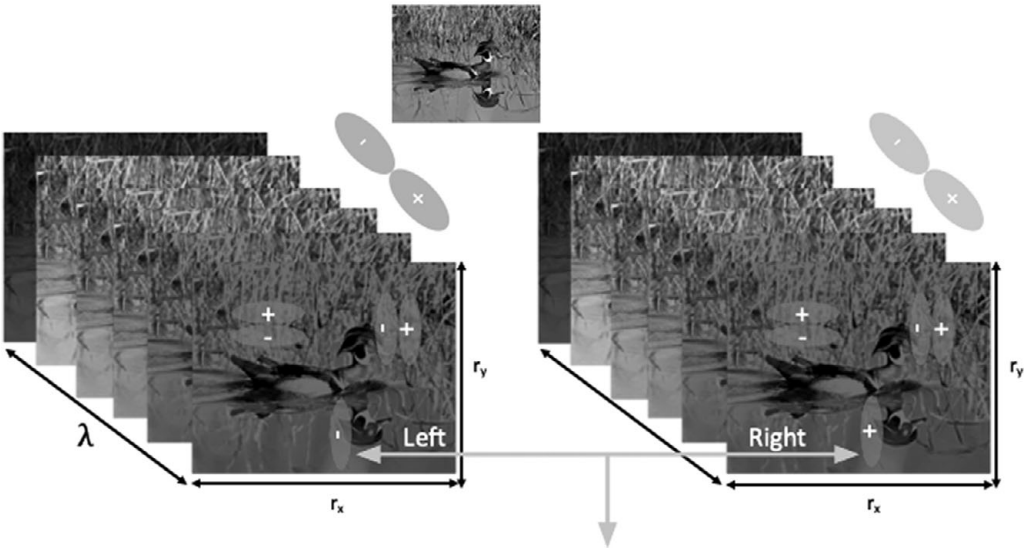

Figure 1.8 *Derivatives of the retinal image. A scene (top) is represented as spectral irradiance hypercubes for the left and right eye. The responses of neurons that compute the local differences, as indicated by several oval pairs with ±, approximate local partial derivatives. Differences can be taken across spatial location, across wavelength, across the spectral radiance measured by the two eyes, and across time (not shown). This figure should be viewed in color. The color version is available at https://color.psych.upenn.edu/supplements/earlyencoding/derivativesColorFig.pdf. The original color image was kindly provided by David Sparks.*

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

respect to spatial position, wavelength, and viewpoint (i.e., across the viewpoints provided by the left and right eyes). Receptive fields that respond to these derivatives include neurons that are pattern-selective (Priebe, 2016; Shapley & Lennie, 1985), cone-opponent (Shevell & Martin, 2017; Solomon & Lennie, 2007), and stereo disparity-selective (Cumming & DeAngelis, 2001). Partial derivatives with respect to time describe motion-selective neurons (Pasternak & Tadin, 2020; Wei, 2018).

The emphasis that Adelson and Bergen (1991) place on these derivatives is consistent with the generally accepted idea that it is the local change (contrast) in the spectral irradiance, not the absolute level of that irradiance, that provides the critical information used for perception (Shapley, 1986). Later in the chapter, we analyze psychophysical measurements of contrast sensitivity, which characterize quantitatively how small changes in spatial contrast are encoded by human vision.

An additional advantage of representations based on derivatives is that they are a highly compressible representation of naturally occurring spectral irradiance. The reason for this is that natural radiances tend to vary slowly, and thus many of the partial derivatives are near zero. A distribution with many repeated values may be compressed by coding the repeated values with tokens specified with a small number of bits, reserving tokens specified with a large number of bits for rarely occurring values (Cover & Thomas, 1991; Wandell, 1995).

1.5 Perceptual Inference

1.5.1 Ambiguity and Perceptual Processing

An important and consistent take-away from the analysis of sensory encoding is that the information available to the brain about the state of the external world is ambiguous: many different physical configurations produce the same sensory representation. A classic example is metamerism: there are only three classes of cone photoreceptors and different spectra produce identical triplets of responses in the L-, M-, and S-cones. Another well-known example is depth reconstruction: the three spatial dimensions of the light field are projected onto a two-dimensional retina, and many 3D shapes produce the same retinal image. Such many-to-one mappings are a reason why Helmholtz (1866, 1896) emphasized perceptual inference: the brain decodes the sensory representation to produce perceptions that are a likely guess about the state of the external world. Perception is an unconscious inference.

1.5.2 Mathematical Principles of Inference

The mathematical formulation of perceptual inference can be developed within a Bayesian probabilistic framework. Suppose $\mathbf{x}$ is a vector that describes some aspect of a scene. The entries of $\mathbf{x}$ might represent the spectral power density of a light entering the eye at a set of discretely sampled wavelengths, the pixel values of a displayed stimulus image, the optical flow vectors corresponding to a

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

viewed dynamic scene, or a full 3D scene description input to a computer graphics package. Now, suppose $\mathbf{y}$ is the sensory representation at some stage of the visual system produced when an observer views the scene described by $\mathbf{x}$. The entries of $\mathbf{y}$ might describe the retinal image, the excitations of each cone in the retinal mosaic, or the action potentials in a class of retinal ganglion cells.

Because sensory measurements are noisy, the relation between $\mathbf{y}$ and $\mathbf{x}$ is described by a conditional probability distribution, $p(\mathbf{y}|\mathbf{x})$. This distribution is referred to as the likelihood function. The likelihood can be thought of as a forward model that relates the scene parameters $\mathbf{x}$ to the sensory representation $\mathbf{y}$.

Within the Bayesian framework, the perceptual representation results from a choice the brain makes about the most likely scene given the observed sensory representation. Indeed, we can reverse the likelihood function, $p(\mathbf{y}|\mathbf{x})$, to obtain a conditional probability distribution $p(\mathbf{x}|\mathbf{y})$, which is called the posterior distribution. The posterior defines which are the more or less likely scenes, given the sensory measurements. To obtain the posterior, we use Bayes' rule (Bishop, 2006; Lee, 1989):

$$p(\mathbf{x}|\mathbf{y}) = K(\mathbf{y})p(\mathbf{y}|\mathbf{x})p(\mathbf{x}), \tag{1.14}$$

where $K(\mathbf{y})$ is a normalizing factor that depends on $\mathbf{y}$ but not $\mathbf{x}$. This factor ensures that the posterior integrates to 1 for any value of $\mathbf{y}$. For many applications, our interest is in how the posterior depends on $\mathbf{x}$, and it is not necessary to compute $K(\mathbf{y})$.

Critically, $p(\mathbf{x})$ is a prior distribution that describes the statistical regularities of the scenes; how likely it is *a priori* that the world is in the state $\mathbf{x}$. A prior is essential because many scenes might have produced the same sensory measurements. Bayes' rule specifies how to combine the prior with the likelihood. Sometimes little is known about the prior. In these cases, using the Bayesian formulation directs our attention to learn more about it. The Bayesian formulation also forces us to make the forward model explicit in the form of the likelihood.

The posterior is a distribution over possible $\mathbf{x}$. We need a means of selecting a specific value, say $\hat{\mathbf{x}}$, to generate the percept. One common way to make a choice is to select a value $\hat{\mathbf{x}}$ that is most likely: the maximum *a posteriori* (MAP) estimate. Other possibilities, such as the mean of the posterior, are also commonly used. The interested reader is referred to the literature on Bayesian decision theory for more on this topic (e.g., Berger, 1985).

It is helpful to consider a simple example. Above we explained that the mean cone excitations at a location are a linear function of the radiance of a displayed image. Suppose we treat the spectral radiance on a display as the state of the world $\mathbf{x}$, with the entries of $\mathbf{x}$ appropriately ordered, and we denote the noisy cone excitations as $\mathbf{y}$. Then

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \boldsymbol{\epsilon} \tag{1.15}$$

for an appropriately arranged matrix $\mathbf{C}$, and where the noise in cone excitations is represented by the random variable $\boldsymbol{\epsilon}$. If we approximate $\boldsymbol{\epsilon}$ with a signal-independent zero-mean Gaussian distribution, we have

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

$$p(\mathbf{y}|\mathbf{x}) = norm(\mathbf{C}\mathbf{x}, \sigma_y^2 \mathbf{I}_y), \tag{1.16}$$

where *norm*() denotes the multivariate Gaussian distribution, σ_y^2 is the variance of the noise added to each mean cone excitation under the Gaussian approximation to the Poisson noise. The symbol $\mathbf{I}_y$ denotes the identity matrix with the same dimensionality as the vector $\mathbf{y}$.

We can also use a Gaussian distribution to describe a prior over $\mathbf{x}$:

$$p(\mathbf{x}) = norm(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x), \tag{1.17}$$

where the vector $\boldsymbol{\mu}_x$ and matrix $\boldsymbol{\Sigma}_x$ represent the mean and covariance of the prior.

Given the Gaussian likelihood and prior, the posterior is also Gaussian; its mean and covariance matrix may be computed analytically from the mean and covariance matrices of the likelihood and prior. This result follows from a standard identity that the product of two multivariate Gaussian distibutions is also a multivariate Gaussian (see Rasmussen & Williams, 2006; Brainard, 1995 provides the derivation in the context of the Bayesian posterior). In the case where the posterior is a multivariate Gaussian, its mean $\boldsymbol{\mu}_{x|y}$ provides the estimate of $\mathbf{x}$ that corresponds to both the posterior mean and the MAP estimate.

Figure 1.9 illustrates the idea for a simple example case. Suppose that the display has only two pixels and emits at only one wavelength. Then $\mathbf{x} = [x_1, x_2]^T$. We will assume that the radiance at each pixel of the display can range between 0 and 1. For natural images, there is a strong correlation between the radiance at neighboring pixels at the same wavelength (Burton & Moorehead, 1987; Tkacik *et al.*, 2011). A bivariate Gaussian prior distribution with this property is illustrated in the left panel of Figure 1.9. The mean of the prior is $\mathbf{x} = [0.5, 0.5]^T$ while the covariance matrix $\boldsymbol{\Sigma}_x$ corresponds to a common standard deviation of 0.127 and a correlation across the two pixels of 0.89. The strong correlation in the prior restricts the best guesses about the values of $\mathbf{x}$ relative to the full available range.

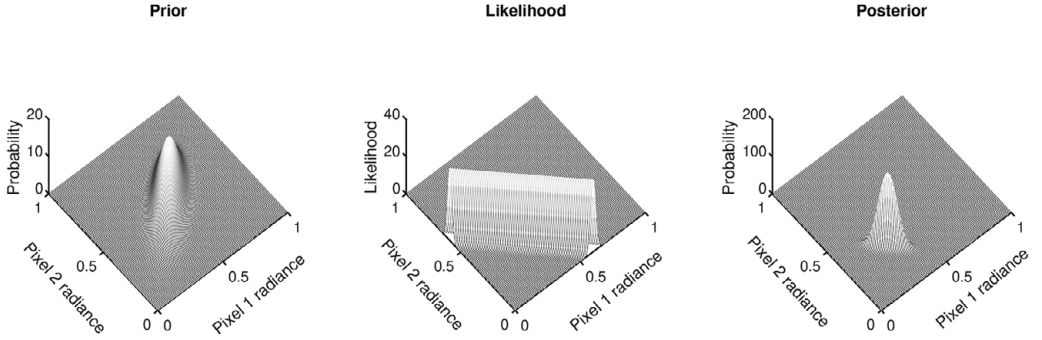

Figure 1.9 *Bayes' reconstruction. See description in text. For the prior and posterior, probability is given as the probability mass for a region of size* 0.01^2 *in the pixel radiance plane. Matlab code to produce this figure is available at https://github.com/DavidBrainard/BrainardFigListings.git (subdirectory scripts/MathPsychChapter/FigLinBayesExample, script Example.m).*

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

To compute a likelihood we need to know the nature of the sensory measurements. We suppose that there is just one cone and that it is equally sensitive to the radiance at the two display pixels. This gives us $\mathbf{C} = [0.5, 0.5]$. We assume that the mean excitation of the cone is perturbed by zero-mean Gaussian noise with standard deviation $\sigma_y = 0.01$. The middle panel of Figure 1.9 illustrates the likelihood for the specific cone excitation $\mathbf{y} = 0.3$: the likelihood $p(\mathbf{y} = 0.3|\mathbf{x})$ is plotted as a function of x_1 and x_2. This likelihood is highest along the ridge where the weighted sum of the pixel radiances sums to the observed cone excitation of 0.3. The likelihood falls off away from this ridge, with the rate of falloff determined by the magnitude of the noise. If the noise were smaller, the falloff would be faster and the likelihood ridge thinner, and conversely if the noise were larger, the falloff would be slower and the likelihood ridge wider. The likelihood alone tells us that $\mathbf{x}$ is unlikely to lie far from the ridge. At the same time, the likelihood makes explicit the ambiguity about $\mathbf{x}$ remaining after observing $\mathbf{y}$, with many values of $\mathbf{x}$ equally likely.

Bayes' rule specifies that the prior and likelihood should be combined using point-by-point multiplication over the pixel radiance plane [Equation (1.14)], and then normalized to form the posterior. The right panel of Figure 1.9 illustrates the result of this multiplication. The same result may be obtained directly by application of the analytic formulas for the posterior.

The posterior makes intuitive sense: it is large where both the prior and likelihood are large, and the resulting distribution is more concentrated than either the prior or likelihood alone. Although there is still uncertainty remaining in the posterior, it captures what we know about the scene when we combine the statistical regularities of the displayed images with the sensory measurement provided by the cone excitation.

1.5.3 Thresholds and Ideal Observer Theory

In this and the next sections, we show how ideas of perceptual inference as implemented through Bayes' rule help us understand perceptual processing. We begin with analysis of threshold measurements. A threshold is the minimum difference required for an observer to correctly discriminate between two stimuli; threshold measurements are a fundamental psychophysical tool. They are used to characterize perceptual performance and guide inferences about the neural mechanisms underlying this performance.

Consider, for example, discrimination between a uniform field and a contrast grating (see Figure 1.10). In a typical experiment, the observer is shown the uniform field and the grating in sequence, with the order randomized on each trial. The observer's task is to indicate which was presented first. In the experiment the stimulus contrast is titrated to a level at which the observer is correct, say, 80% of the time. The estimated contrast is the threshold.

Threshold measurements quantify the information needed by the visual system to make a basic perceptual decision: namely, that two stimuli differ. They involve small perturbations of the visual stimulus, and they may be thought of as assessing

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

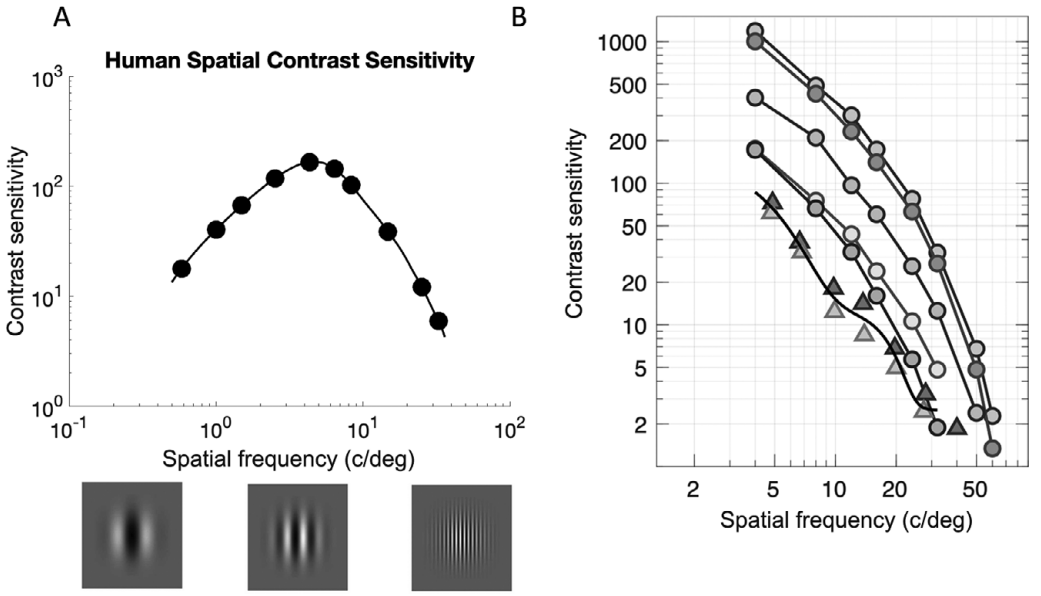

Figure 1.10 *Modeling the human contrast sensitivity function. (A) Sensitivity, defined as the inverse of threshold contrast, is plotted as a function of spatial frequency. The stimuli were small, equal-sized patches of contrast gratings. Replotted from De Valois, Morgan, and Snodderly (1974). The smooth curve replots the smooth curve in the original figure, while the solid points show the spatial frequencies on the smooth curve at which contrast sensitivity was measured. See the original figure for the actual sensitivity measurements through which the smooth curve was drawn. The thumbnails below the plot illustrate contrast grating patches at different spatial frequencies, but are not otherwise matched to the spatial frequency of the plot. (B) Triangles and black line: Human contrast sensitivity function for two observers, data from Banks, Geisler, and Bennett (1987). Grey circles/line: Contrast sensitivity of an ideal observer implemented at the level of the Poisson limited cone excitations, from Banks, Geisler, and Bennett (1987). Red circles/line: Ideal observer CSF with recent estimates of optics and mosaic properties. Blue circles/line: Computational observer CSF with decision rule determined using supervised machine learning. Green circles/line: Computational observer CSF additionally accounting for fixational drift. Purple circles/line: Computational observer CSF additionally incorporating a model of the transformation from excitations to photocurrent. This figure should be viewed in color. The color version is available at https://color.psych.upenn.edu/supplements/earlyencoding/csfColorFig.pdf. If you are nonetheless viewing a grayscale version of the figure, the order of the colors of the ideal/computational observer CSFs from top to bottom is: gray, red, blue, green, purple. After Figure 6 of Cottaris* et al. *(2020).*

sensitivity to derivatives of the retinal image. In this way, thresholds are connected to the ideas introduced above about the importance of derivatives of the spectral radiance as a basis for visual processing.

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

Figure 1.10 shows the threshold for contrast gratings measured as a function of grating spatial frequency: this is called the spatial contrast sensitivity function (CSF). When measured with static or very slowly moving gratings, the human CSF has an inverted U-shape: the highest contrast sensitivity is between three and six cycles per degree, with lower sensitivity at higher and lower spatial frequencies. Because any image may be synthesized by a weighted superposition of sinusoidal gratings (Bracewell, 1978), the CSF characterizes the sensitivity to basic stimulus components. Because the visual system as a whole is neither shift-invariant nor linear, however, the CSF is a useful but incomplete description of sensitivity.

We would like to understand how the human contrast sensitivity function is limited by the properties of the visual components described in this chapter. Bayes' rule provides a way to build this understanding by linking the initial encoding to performance on the psychophysical threshold detection task. Analyses of this sort are called ideal observer theory (Geisler, 1989). Ideal observer theory allows us to estimate the extent to which discrimination performance is limited by the early visual encoding. Relevant factors include blurring by the eye's optics, which reduces the retinal contrast of a grating stimulus, spatial sampling by the cone mosaic, and the Poisson variability in the cone excitations. Of particular interest is separating aspects of visual performance that are tightly coupled to these factors from aspects that are limited by processes not incorporated into the ideal observer calculation.

So, how do we use Bayes to predict performance in the two-interval forced choice task described above? We use the terms reference stimulus and comparison stimulus to describe the two stimuli being discriminated. In this example the reference stimulus is a spatially uniform field and the comparison stimulus is a patch of contrast grating with known spatial frequency, orientation, size, and contrast; but, the ideas we develop here apply to any two stimuli being discriminated.

Using the computational methods described in this chapter, we compute the mean cone excitations to the reference and comparison stimuli. Let $\mathbf{u}_r$ be the vector of mean cone mosaic excitations in response to the reference stimulus and let $\mathbf{u}_c$ be the vector of mean cone excitations in response to the comparison stimulus. In the two-interval forced choice task, the observer must indicate whether the reference came first followed by the comparison, or the other way around. We thus form two concatenated vectors, $\mathbf{u}_1 = [\mathbf{u}_r, \mathbf{u}_c]$ and $\mathbf{u}_2 = [\mathbf{u}_c, \mathbf{u}_r]$.

To apply Bayes' rule to this problem, we think of the scene as described by a binary random variable. This variable, x, can take on value 1 or 2. These values represent the reference first and reference second possibilities that can occur on each trial. The prior probability $p(x)$ is given by

$$p(x = 1) = 0.5;\ p(x = 2) = 0.5. \tag{1.18}$$

The data available to the observer to make a response of $x = 1$ or $x = 2$ are the pattern of observed cone excitations across the two intervals, which we will denote by $\mathbf{y}$. We know that for $x = 1$, each entry of $\mathbf{y}$ is an independent Poisson random

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

variable with mean given by the corresponding entry of $\mathbf{u}_1$, while for $x = 2$ the means are given by the corresponding entries of $\mathbf{u}_2$. From this, we have for the posterior:

$$p(x = 1|\mathbf{y}) = Kp(\mathbf{y}|x = 1)p(x = 1) = K\prod_i p(y_i|x = 1)p(x = 1), \tag{1.19}$$

where y_i denotes the ith entry of $\mathbf{y}$ and we have explicitly expressed the joint distribution of independent random variables as the product of their individual distributions. K is a normalizing constant whose value we need not calculate.

We substitute the expression for the probability mass function of a Poisson random variable and the value of $p(x = 1)$ to obtain

$$p(x = 1|\mathbf{y}) = K\prod_i \frac{{u_{1i}}^{y_i}e^{-u_{1i}}}{y_i!}0.5, \tag{1.20}$$

where u_{1i} denotes the ith entry of $\mathbf{u}_1$. Similarly, we have

$$p(x = 2|\mathbf{y}) = K\prod_i \frac{{u_{2i}}^{y_i}e^{-u_{2i}}}{y_i!}0.5. \tag{1.21}$$

To maximize the percent correct on the task, the observer should compare $p(x = 1|\mathbf{y})$ with $p(x = 2|\mathbf{y})$ and indicate 1 or 2 according to which is larger. It is instructive to implement this comparison in terms of the difference of the logs of $p(x = 1|\mathbf{y})$ and $p(x = 2|\mathbf{y})$, with a response of 1 corresponding to a difference greater than or equal to 0 and a response of 2 corresponding to a difference less than 0. Writing the difference of logs explicitly and simplifying, we have decision variable

$$\delta = \sum_i y_i \log\left(\frac{u_{1i}}{u_{2i}}\right) + \sum_i (u_{2i} - u_{1i}). \tag{1.22}$$

An observer who responds according to the sign of δ will maximize the percent correct. The value of the percent correct depends on how δ is distributed when $x = 1$ and $x = 2$. Geisler (1984) provides a Gaussian approximation to these distributions, which may be used to obtain the corresponding percent correct. As with the human psychophysical experiment, contrast may be titrated to find the ideal observer threshold contrast, that which leads to the ideal observer having the criterion percent correct.

Figure 1.10B shows the ideal observer contrast sensitivity for human foveal viewing (gray circles/line), along with psychophysical measurements of human contrast sensitivity at spatial frequencies increasing from 5 cpd, and with the measurements (triangles/black line) made with stimuli matched to those used in the ideal observer calculations (Banks, Geisler, & Bennett, 1987). As with the human data at higher spatial frequencies, the ideal observer contrast sensitivity function falls off as spatial frequency increases; the slope of this falloff closely resembles that of the human observer. This correspondence suggests that the factors that cause the human falloff share basic features with those included in

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

the ideal observer calculation. Here the primary factor is blur from the eye's optics and cone apertures, both of which reduce the contrast captured by spatial variation in the cone excitations.

The ideal observer CSF also differs from the human measurements. One difference is that the overall sensitivity of the ideal observer is markedly higher than that of the human observer. The Poisson noise in the cone excitations limits the ideal observer sensitivity. The fact that incorporating only this noise source leads to an ideal observer more sensitive than the human tells us that additional factors limit human sensitivity and motivates study of what these additional factors are.

One approach is to define a single "efficiency" parameter representing an omnibus loss of information by the actual visual system relative to an ideal observer calculation. This is often sufficient to bring ideal observer predictions into alignment with measured human performance (Burge, 2020), as is true in the case of the ideal and human CSF rolloff at high spatial frequencies. The efficiency parameter can be thought of as capturing the effect of additional noise in the human visual system, not included in the ideal observer calculation, whose effect on performance is stimulus-independent.

It is important to note, however, that the difference between ideal and human performance is not fully explained by a single efficiency parameter. For example, the ideal observer CSF does not roll off at low spatial frequencies but the human CSF does. The factors that produce the measured low-spatial-frequency rolloff are not included in the ideal observer calculations presented here. As with the difference in overall sensitivity, the difference between ideal and human CSF at low spatial frequencies motivates investigation of what additional factors in the human visual system account for the difference.

1.5.4 Computational Observers

The ideal observer calculation used by Banks, Geisler, & Bennett, (1987) employed a simplified model of the eye's point spread function and cone mosaic, and this simplification enabled efficient computation of ideal observer performance. In two recent papers, Cottaris *et al.* (2019, 2020) employed computational methods to examine the effect of more recent estimates of the point spread function (Thibos *et al.*, 2002) and a more detailed model of the foveal mosaic on performance. These had only a modest effect on the predictions (Figure 1.10B, red circles/line).

The ideal observer developed above has full knowledge of mean cone excitations and Poisson structure of the noise, so that the observer's performance is not degraded by stimulus uncertainty (Geisler, 2018; Pelli, 1985). Cottaris *et al.* (2019) relaxed this assumption by replacing the ideal observer decision rule with a decision rule based on a trained linear classifier (C. D. Manning, Raghavean, & Schutze, 2008; Schölkopf *et al.*, 2002). The classifier measured the match of the data to a template that had the same spatial structure as the stimuli. The decision boundary was optimized in the presence of noise. The need to partially learn the decision rule reduced the absolute level of ideal observer performance

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

while retaining the same CSF shape (blue circles/line in Figure 1.10B). Cottaris *et al*. (2020) then introduced a computational model of fixational eye movements (Mergenthaler & Engbert, 2007; see also Engbert & Kliegl, 2004) and showed that an approach to handling the stimulus motion blur introduced by these movements further reduced performance (green circles/line). Finally, Cottaris *et al*. (2020) introduced a computational model of the transformation from excitations to electrical photocurrent, which included both gain control and additional noise. Accounting for this transformation brought computational observer performance into approximate alignment with the human measurements at the higher spatial frequncies (purple circles/line).

This analysis outlines a set of factors that together provide an account of the high-spatial-frequency limb of the human spatial CSF, capturing both the shape and absolute level of this important measure of performance. For the purposes of the present chapter, we emphasize less the specific elements of the account, which will surely be refined by future research, but rather the way the mathematical principles are combined with computational modeling with the goal of accounting for the full richness of the visual system. The combination of principles and computations accounts for factors that are beyond what is possible using analytic calculations alone.

1.5.5 Image Reconstruction

The ideal observer and computational observer development above applies Bayesian inference to the analysis of threshold measurements. Thresholds characterize the limits of visual performance, and the analyses illustrate how threshold performance can be linked to quantitative measurements of physiological optics, retinal anatomy, and retinal physiology. Not all vision is threshold vision, however. Sometimes we are interested in predicting what clearly visible stimuli look like (e.g., "that apple looks red") or how similar easily distinguishable objects appear (e.g., "the color of the apple appears more similar to the color of the tomato than it does to the color of the banana"). There are a number of methods for studying suprathreshold vision. These include asymmetric matching (Brainard & Wandell, 1992; Burnham, Evans, & Newhall, 1957; Wandell, 1995) and various scaling techniques (T. F. Cox & Cox, 2001; Knoblauch & Maloney, 2012; Maloney & Yang, 2003). We will not treat these methods here. Below, however, we illustrate how Bayesian methods can be used to understand how the initial visual encoding shapes the perceptual inferences that can be made about suprathreshold stimuli.

In our introduction to Bayes' rule, we illustrated the core ideas by considering reconstruction of a two-pixel image from the excitations of a single cone, using both a Gaussian and a Gaussian likelihood. As computer power has increased, these same Bayesian principles have been applied to increasingly large perceptual problems. As we illustrate here, it is now possible to reconstruct an estimate of a full displayed color image from a realistic model of cone excitations using the Poisson likelihood (Zhang, Cottaris, & Brainard, 2021).

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

The forward computation starts with the displayed image $\mathbf{x}$ and computes the cone excitations $\mathbf{y}$. The vector $\mathbf{x}$ can be thought of as the concatenation of the linearized and rasterized pixel values for each of the red, green, and blue channels of the display. Using the Poisson noise model of the cone excitations, we compute the likelihood of observed cone excitations $p(\mathbf{y}|\mathbf{x})$. Here the vector $\mathbf{y}$ is simply a list of the excitations of each cone in the mosaic. Because the mean cone excitations are a linear function of the display pixel values, we can write for these mean excitations

$$\bar{y} = \mathbf{R}\mathbf{x} \tag{1.23}$$

for some matrix $\mathbf{R}$. Each column of this matrix may be computed as the vector of cone excitations produced when one pixel is at its maximum value for one color channel, with the display values for all other pixels and color channels set to zero, and these computations may be implemented in software such as ISETBio to determine explicitly the matrix $\mathbf{R}$ (Zhang, Cottaris, & Brainard, 2021). This yields for the likelihood

$$p(\mathbf{y}|\mathbf{x}) = Poisson(\mathbf{R}\mathbf{x}), \tag{1.24}$$

where $Poisson()$ denotes the result of Poisson noise applied independently to its vector argument by taking each entry of the argument as the corresponding Poisson mean.

Next, we specify a prior distribution $p(\mathbf{x})$ for natural images. Natural images have a great deal of structure (Simoncelli, 2005), and a full statistical description of this structure is not currently available. There are two robust regularities of natural images, however, that can be described by a multivariate Gaussian. The first is that within a single wavelength band, the spectral radiances at nearby image locations are highly correlated (Field, 1987; Pratt, 1978; Ruderman, Cronin, & Chiao, 1998). The second regularity is that at a single position, values in nearby wavelength bands are highly correlated (Burton & Moorehead, 1987; Tkacik *et al.*, 2011). This is a consequence of the relatively smooth spectral functions one observes in nature (Cohen, 1964; Maloney, 1986; Vrhel, Gershon, & Iwan, 1994). These two observations may be used to construct a covariance matrix for a multivariate Gaussian that describes the second-order statistics of natural images. Together with the average image, these provide a Gaussian image prior.

With the likelihood and prior, we can construct an estimate of the image given a vector of cone excitations. As with many calculations described in this chapter, the principles of Bayesian estimation guide the way, but once we introduce the Poisson likelihood, we turned to numerical computational methods to find the solution.

We used ISETBio to reconstruct images from cone excitations, with the Poisson likelihood and Gaussian image prior described above. We reconstructed images for retinal patches at various visual field eccentricities. As visual field eccentricity increases, the point spread of the retinal image becomes more blurred and the density with which the cones sample the image decreases (Figures 1.5, 1.6, and 1.7). Thus, less information becomes available to the visual system in the

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

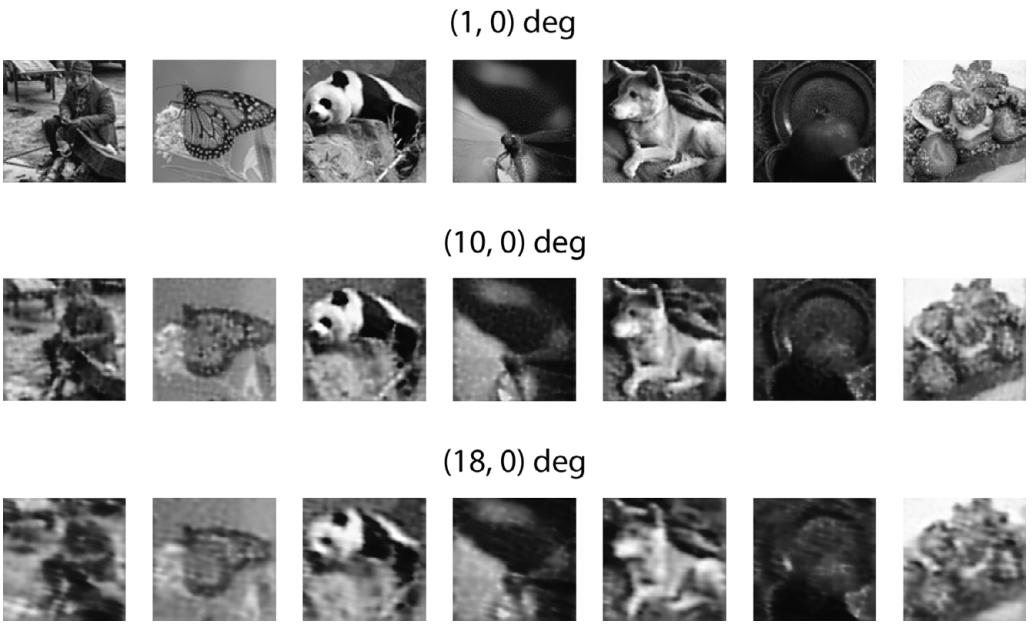

Figure 1.11 *Image reconstructions from cone excitations at three retinal eccentricities. Each row shows reconstructions of seven images using the Bayesian method and Poisson likelihood and multivariate Gaussian prior. The reconstructions at 1° eccentricity are close to veridical, with increasing distortions seen at the 10 and 18° locations. Each original and reconstructed image was represented at a pixel resolution of 128 × 128, and the extent of each image on the retina was 1° × 1°. The mean excitation of the cones was 10^5 excitations per cone, so the simulation corresponds to a relatively high signal-to-noise regime. The parameters of the Gaussian prior were fit to 16 × 16 pixel patches of images from the ImageNet ILSVRC data set (www.image-net.org), and extended in an overlapping blockwise fashion to the higher image pixel resolution. This figure should be viewed in color. The color version is available at https://color.psych.upenn.edu/supplements/earlyencoding/GaussianReconColorFig.pdf. Figure courtesy Lingqi Zhang. See Zhang, Cottaris, and Brainard (2021) for a more extended discussion of Bayesian image reconstruction and the general methods used to produce this figure.*

peripheral visual field. The effect of this loss for reconstruction depends on the prior. Although the information loss means that two images whose cone excitations are different in the fovea can produce the same cone excitations in the periphery, this ambiguity need not degrade the reconstructions if the probability that one of the two images will occur is small.

The reconstructions in Figure 1.11 show the effect of information loss at the level of the cone excitations, in the context of the Gaussian image prior. The reconstructed image quality in the periphery is worse than in the fovea, but many objects remain recognizable from the peripheral reconstructions. Moreover, there

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

are interesting interactions between the likelihood and prior. For example, the recovery of color can be better in the fovea and more peripheral locations than it is in the mid-periphery (see images of strawberries in Figure 1.11, for example). Zhang, Cottaris, and Brainard (2021) describe the reconstruction approach to analyzing the initial visual encoding in more detail, extending the ideas to a more realistic prior than the Gaussian, and showing a number of calculations that use image reconstruction to examine how prior and likelihood interact to support both color and spatial vision (see also Brainard, Williams, & Hofer, 2008).

Image reconstruction computations provide useful insights about how statistical regularities in natural scenes interact with the sensory measurements to guide perception. But, it is important to bear in mind that reconstruction of displayed images is not the task for which visual perception evolved. Rather, we view the task of perception to reconstruct the properties and positions of objects in the three-dimensional environment. The Bayesian ideas presented here have applicability to this task as well (Knill & Richards, 1996), but a computational solution that is as effective as human vision currently remains elusive. This is an area where recent progress in machine learning and deep neural networks may provide new insights.

1.5.6 Optimizing Sensory Measurements

Earlier in this chapter, we explained that the visual system appears to extract information about motion, color, and pattern from the pattern of cone excitations by estimating the local derivatives of various quantities (Adelson & Bergen, 1991). The Bayesian framework provides a quantitative framework for addressing how to optimize which signals should be transduced by a sensory system when the goal is reconstruction of the state of the environment, as well as how the sensory signals should be summarized (e.g., in the form of local derivatives) for further processing. Indeed, the Bayesian image reconstruction methods developed here point towards the ingredients required for a full analysis of such questions. To know what measurements we should make, we first need to know the prior distribution over the environmental states that an organism will encounter. We then need a parameterized set of candidate likelihood functions, each of which describes a feasible arrangement of the sensory apparatus and (if desired) associated early processing. This information allows us to compute the posterior over the environmental states for any candidate likelihood function, and we can ask how well different sensory measurements constrain the posterior, averaging this information over the environmental states described by the prior. Developing a parameterized set of candidate likelihood functions requires an understanding of what biological constraints apply to the sensory system. Also required is an understanding of the cost of different types of error in the resultant perceptual representation (the loss function; Berger, 1985), as well as how the cost of error should be balanced against the energetic cost of making and processing the sensory measurements (Balasubramanian, Kimber, & Berry, 2001; Koch *et al.*, 2004; Laughlin, 2001). A number of authors have pursued questions of optimizing sensory measurements in this manner (Garrigan *et al.*, 2010; Levin, Durand, &

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

Freeman, 2008; J. R. Manning & Brainard, 2009; Zhang, Cottaris, & Brainard, 2021).[10] It would be interesting to compare the results of an analysis of this sort to the Adelson/Bergen conjecture that approximations to local derivatives represent an optimal measurement set.

1.6 Summary and Conclusions

To focus on the mathematics of the initial visual encoding, we introduce vision science from the point of view of a forward calculation: physics of the stimulus, image formation, and quantitative system modeling. The key mathematical principles are linear algebra, shift-invariant linear systems, and specification of sensory noise. The mathematics of vision science shares much in common with the mathematics of many fields of science and engineering.

After expressing and implementing the forward calculations, we explore the mathematics of Helmholtz's hypothesis: people perceive a stimulus that is the most likely explanation of the cone excitations. We use Bayesian inference methods to clarify the uncertainty about the encoded signal. This approach requires that we confront the problem of establishing priors on the signal. There is a close connection between Helmholtz's unconscious inference and Bayesian inference; the latter may be thought of as a quantitative implementation of Helmholtz's idea.

The approach we describe has a long and accomplished tradition. But, it is not the only valid way to make progress in vision science; several other approaches are important. A quantitative study of behavioral rules can be very informative. For example, color appearance matching was a largely behavioral exploration at first; an understanding of the physics of the signal and the biological underpinnings followed later. Also, neurobiological measures can be helpful. Anatomical and functional measurements that characterize the properties of multiple pathways within the visual system – including multiple types of retinal ganglion cells and multiple pathways through the visual cortex – are useful guides to understanding visual specializations and computations, particularly for stages of vision beyond the initial encoding. Finally, engineering work to build functional artificial visual systems continues to be very helpful in understanding vision: a classic principle states that the best way to demonstrate you understand a system is to build one that does the same thing. Engineering efforts continue to clarify features that we might look for in the nervous system, as well as why certain behavioral patterns emerge.

The field of vision science is large and vigorous enough that there is no need to choose a single approach. We are inspired by the fact that different investigators adopt different approaches, all seeking to gain understanding. To the student thinking about how to approach vision science, we offer advice from an American philosopher who commented about making difficult decisions: "When you come to a fork in the road, take it" (Yogi Berra).

10 The formalism used in these analyses is interestingly similar to that underlying Bayesian adaptive psychophysical procedures (Watson, 2017; Watson & Pelli, 1983).

https://doi.org/10.1017/9781108902724.002 Published online by Cambridge University Press

1.7 Related Literature

This chapter introduces key mathematical and computational approaches to understanding the initial visual encoding. A number of the mathematical ideas we present here are developed in more detail by Wandell (1995), and the classic treatment of visual perception by Cornsweet (1970) remains a valuable introduction to the field, as does Rodieck (1998). Principles of ray tracing are introduced in many computer graphics texts (e.g., Pharr, Jakob, & Humphreys, 2016); similarly many texts introduce optics (e.g., Hecht, 2017). In the context of the retinal image and cone excitations specifically, Packer and Williams (2003), Pugh (1988), and Yellott, Wandell, and Cornsweet (1984) are useful. Brainard and Stockman (2010) elaborate in more detail on using linear algebra in support of colorimetric applications. Although we do not treat the Fourier transform and frequency domain representations in this chapter, the reader who wishes to specialize in this field will want to learn about these ideas. Two useful sources are Bracewell (1978) and Pratt (1978). Useful introductions to statistical inference include Bishop (2006) and Duda, Hart, and Stork (2001).

Acknowledgments

We thank Nicolas Cottaris and Lingqi Zhang for providing figures for this chapter. We also thank Amy Ni, Nicolas Cottaris, Lingqi Zhang, Joyce Farrell, Eline Kupers, Heiko Schutt, and Greg Ashby for useful comments on the manuscript.

References

Adelson, E. H., & Bergen, J. (1991). The plenoptic function and the elements of early vision. In M. Landy & J. Movshon (Eds.), *Computational models of visual processing* (pp. 3–20). Cambridge, MA: MIT Press.

Adelson, E. H., & Wang, J. Y. A. (1992, February). Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(2), 99–106.

Ayscough, J. (1755). *Light field image.* https://commons.wikimedia.org/wiki/File:1755_james_ayscough.jpg. (Accessed: July 24, 2021.)

Balasubramanian, V., Kimber, D., & Berry, M. J. (2001). Metabolically efficient information processing. *Neural Computation*, *13*(4), 799–815.

Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Research*, *27*(11), 1915–1924.

Baylor, D. A., Lamb, T. D., & Yau, K. W. (1979). Responses of retinal rods to single photons. *Journal of Physiology*, *288*(Mar), 613–634.

Berger, T. O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer Science + Business Media LLC.

Bracewell, R. (1978). *The Fourier transform and its applications*. New York: McGraw-Hill.

Brainard, D. H. (1995). *An ideal observer for appearance: Reconstruction from samples* (Tech. Rep. No. 95-1). UCSB Vision Labs Technical Report.

Brainard, D. H. (2015). Color and the cone mosaic. *Annual Review of Vision Science*, *1*, 519–546.

Brainard, D. H., & Stockman, A. (2010). Colorimetry. In M. Bass et al. (Eds.), *The Optical Society of America handbook of optics, 3rd edition, Volume III: Vision and vision optics* (pp. 10.1–10.56). New York: McGraw-Hill.

Brainard, D. H., & Wandell, B. A. (1992). Asymmetric color-matching: How color appearance depends on the illuminant. *Journal of the Optical Society of America A*, *9*(9), 1433–1448.

Brainard, D. H., Williams, D. R., & Hofer, H. (2008). Trichromatic reconstruction from the interleaved cone mosaic: Bayesian model and the color appearance of small spots. *Journal of Vision*, *8*(5), 1–23.

Branwyn, G. (2016, September). *Sky angles.* https://makezine.com/2016/09/16/measuring-tip-ruler/sky_angles/. (Accessed: August 8, 2021.)

Burge, J. (2020). Image-computable ideal observers for tasks with natural stimuli. *Annual Review of Vision Science*, *6*, 491–517.

Burnham, R., Evans, R., & Newhall, S. (1957). Prediction of color appearance with different adaptation illuminations. *Journal of the Optical Society of America*, *47*(1), 35–42.

Burns, S. A., Elsner, A. E., Lobes, L. A., Jr, & Doft, B. H. (1987, April). A psychophysical technique for measuring cone photopigment bleaching. *Investigative Ophthalmology & Visual Science*, *28*(4), 711–717.

Burton, G. J., & Moorehead, I. R. (1987). Color and spatial structure in natural images. *Applied Optics*, *26*(1), 157–170.

Cahan, D. (1993). *Hermann von Helmholtz and the foundations of nineteenth-century science*. Berkeley, CA: University of California Press.

Canon U.S.A., Inc. (2017, April). *Introduction to dual pixel autofocus.* www.usa.canon.com/internet/portal/us/home/learn/education/topics/article/2018/July/Intro-to-Dual-Pixel-Autofocus-(DPAF)/Intro-to-Dual-Pixel-Autofocus-(DPAF). (Accessed: July 8, 2021.)

CIE (1986). *Colorimetry, second edition* (Report No. 15.2). Vienna: Bureau Central de la CIE.

CIE (2007). *Fundamental chromaticity diagram with physiological axes - parts 1 and 2. technical report 170-1*. Vienna: Bureau Central de la CIE.

Cohen, J. (1964). Dependency of the spectral reflectance curves of the Munsell color chips. *Psychonomic Science*, *1*, 369–370.

Cornsweet, T. (1970). *Visual perception*. New York: Academic Press.

Cottaris, N. P., Jiang, H., Ding, X., Wandell, B. A., & Brainard, D. H. (2019). A computational observer model of spatial contrast sensitivity: Effects of wavefront-based optics, cone mosaic structure, and inference engine. *Journal of Vision*, *19*(4), 8.

Cottaris, N. P., Wandell, B. A., Rieke, F., & Brainard, D. H. (2020). A computational observer model of spatial contrast sensitivity: Effects of photocurrent encoding, fixational eye movements, and inference engine. *Journal of Vision*, *20*(7), 17.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: John Wiley & Sons.

Cox, D. D., & Dean, T. (2014, September). Neural networks and neuroscience-inspired computer vision. *Current Biology*, *24*(18), R921–R929.

Cox, T. F., & Cox, M. A. A. (2001). *Multidimensional scaling*, 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.

Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Review of Neuroscience*, *24*(1), 203–238.

Curcio, C., Sloan, K., Kalina, R., & Hendrickson, A. (1990). Human photoreceptor topography. *Journal of Comparative Neurology*, *292*(4), 497–523.

Da Vinci, L. (1970). *The notebooks of Leonardo da Vinci* (Vol. 1; J. P. Richter, Ed.). New York: Dover.

De Valois, R. L., Morgan, H., & Snodderly, D. M. (1974). Psychophysical studies of monkey vision–III. Spatial luminance contrast sensitivity tests of macaque and human observers. *Vision Research*, *14*(1), 75–81.

Duda, R., Hart, P., & Stork, D. (2001). *Pattern classification and scene analysis,* 2nd ed. New York: John Wiley & Sons.

Engbert, R., & Kliegl, R. (2004). Microsaccades keep the eyes' balance during fixation. *Psychological Science*, *15*(6), 431–436.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*(12), 2379–2394.

Gamlin, P. D. R., McDougal, D. H., Pokorny, J., Smith, V. C., Yau, K. W., & Dacey, D. M. (2007). Human and macaque pupil responses driven by melanopsin-containing retinal ganglion cells. *Vision Research*, *47*(7), 946–954.

Garrigan, P., Ratliff, C. P., Klein, J. M., Sterling, P., Brainard, D. H., & Balasubramanian, V. (2010). Design of a trichromatic cone array. *PLoS Computational Biology*, *6*(2), e1000677.

Geisler, W. S. (1984). Physical limits of acuity and hyperacuity. *Journal of the Optical Society of America A*, *1*(7), 775–782.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*(2), 267–314.

Geisler, W. S. (2018). Psychometric functions of uncertain template matching observers. *Journal of Vision*, *18*(2), 1.

Gershun, A. (1939). The light field. *Journal of Mathematical Physics*, *18*(1–4), 51–151.

Grassmann, H. (1853). Zur Theorie der Farbenmischung. *Annalen der Physik und Chemie*, *165*, 69–84.

Hattar, S., Liao, H. W., Takao, M., Berson, D. M., & Yau, K. W. (2002). Melanopsin-containing retinal ganglion cells: Architecture, projections, and intrinsic photosensitivity. *Science*, *295*(5557), 1065–1070.

Hecht, E. (2017). *Optics*, 5th ed. Boston, MA: Pearson.

Hecht, S., Schlaer, S., & Pirenne, M. (1942). Energy, quanta and vision. *Journal of the Optical Society of America*, *38*(6), 196–208.

Helmholtz, H. (1866). *Handbuch der physiologischen Optik II* (3rd ed., 1911) (pp. 243–244). Hamburg: Voss.

Helmholtz, H. (1896). *Physiological optics*. New York: Dover.

Hofer, H., & Williams, D. R. (2014). Color vision and the retinal mosaic. In L. M. Chalupa & J. S. Werner (Eds.), *The new visual neurosciences* (pp. 469–483). Cambridge, MA: MIT Press.

Hunt, R. W. G. (2004). *The reproduction of colour*, 6th ed. Chichester: John Wiley & Sons.

Jaeken, B., & Artal, P. (2012). Optical quality of emmetropic and myopic eyes in the periphery measured with high-angular resolution. *Investigative Ophthalmology and Visual Science*, *53*(7), 3405–3413.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.

Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in R (use R!)*. New York: Springer-Verlag.

Koch, K., McLean, J., Berry, M., Sterling, P., Balasubramanian, V., & Freed, M. A. (2004). Efficiency of information transmission by retinal ganglion cells. *Current Biology*, *14*(17), 1523–1530.

Laughlin, S. B. (2001). Energy as a constraint on the coding and processing of sensory information. *Current Opinion in Neurobiology*, *11*(4), 475–480.

Lee, P. M. (1989). *Bayesian statistics*. London: Oxford University Press.

Levin, A., Durand, F., & Freeman, W. T. (2008). *Understanding camera trade-offs through a Bayesian analysis of light field projections* (Report). Cambridge, MA: MIT Press.

Lian, T., MacKenzie, K. J., Brainard, D. H., Cottaris, N. P., & Wandell, B. A. (2019). Ray tracing 3D spectral scenes through human optics models. *Journal of Vision*, *19*(12), 23.

Maloney, L. T. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America A*, *3*(10), 1673–1683.

Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision*, *3*(8), 573–585.

Manning, C. D., Raghavean, P., & Schutze, H. (2008). *Introduction to information retrieval*. Cambridge: Cambridge University Press.

Manning, J. R., & Brainard, D. H. (2009). Optimal design of photoreceptor mosaics: Why we do not see color at night. *Visual Neuroscience*, *26*(1), 5–19.

Marimont, D. H., & Wandell, B. A. (1994, December). Matching color images: The effects of axial chromatic aberration. *Journal of the Optical Society of America A*, *11*(12), 3113–3122.

Maxwell, J. (1860). On the theory of compound colours and the relations of the colours of the spectrum. *Philosophical Transactions of the Royal Society of London*, *150*, 57–84.

Mergenthaler, K., & Engbert, R. (2007). Modeling the control of fixational eye movements with neurophysiological delays. *Physical Review Letters*, *98*(13), 138104.

Mlinar, M. (2016, May). *Image processing methods for image sensors with phase detection pixels* (No. 9338380).

Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., & Hanrahan, P. (2005). *Light field photography with a hand-held plenoptic camera* (Unpublished doctoral dissertation, Stanford University).

Packer, O., & Williams, D. R. (2003). Light, the retinal image, and photoreceptors. In S. K. Shevell (Ed.), *The science of color,* 2nd ed. (pp. 41–102). Oxford: Optical Society of America/Elsevier Ltd.

Pasternak, T., & Tadin, D. (2020). Linking neuronal direction selectivity to perceptual decisions about visual motion. *Annual Review of Vision Science*, *6*, 335–362.

Pelli, D. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, *2*(9), 1508–1532.

Pharr, M., Jakob, W., & Humphreys, G. (2016). *Physically based rendering: From theory to implementation*. San Francisco, CA: Morgan Kaufmann.

Polans, J., Jaeken, B., McNabb, R. P., Artal, P., & Izatt, J. A. (2015). Wide-field optical model of the human eye with asymmetrically tilted and decentered lens that reproduces measured ocular aberrations. *Optica*, *2*(2), 124–134.

Pratt, W. K. (1978). *Digital image processing*. New York: John Wiley & Sons.

Priebe, N. J. (2016). Mechanisms of orientation selectivity in the primary visual cortex. *Annual Review of Vision Science*, *2*, 85–107.

Pugh, J. (1988). Vision: Physics and retinal physiology. In R. Atkinson, R. Herrnstein, G. Lindsey, & R. Luce (Eds.), *Stevens' handbook of experimental psychology*, 2nd ed. (Vol. 1, pp. 75–163). New York: John Wiley & Sons.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.

Rodieck, R. (1998). *The first steps in seeing*. Sunderland, MA: Sinauer.

Ruderman, D. L., Cronin, T. W., & Chiao, C. C. (1998). Statistics of cone responses to natural images: Implications for visual coding. *Journal of the Optical Society of America A*, *15*(8), 2036–2045.

Rushton, W. (1972). Pigments and signals in colour vision. *Journal of Physiology*, *220*(3), 1P–31P.

Schölkopf, B., Smola, A. J., Bach, F., *et al.* (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. Cambridge, MA: MIT Press.

Shapley, R. M. (1986). The importance of contrast for the activity of single neurons, the VEP and perception. *Vision Research*, *26*(1), 45–62.

Shapley, R. M., & Lennie, P. (1985). Spatial frequency analysis in the visual system. *Annual Review of Neuroscience*, *8*(1), 547–581.

Shevell, S. K., & Martin, P. R. (2017). Color opponency: Tutorial. *Journal of the Optical Society of America A*, *34*(7), 1099–1108.

Simoncelli, E. P. (2005). Statistical modeling of photographic images. In A. Bovik (Ed.), *Handbook of image and video processing* (pp. 431–441). New York: Academic Press.

Solomon, S., & Lennie, P. (2007). The machinery of colour vision. *Nature Reviews Neuroscience*, *8*(4), 276–286.

Stiles, W., & Burch, J. (1959). NPL colour-matching investigation: Final report (1958). *Optica Acta*, *6*, 1–26.

Stockman, A., & Sharpe, L. (2000). Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype. *Vision Research*, *40*(13), 1711–1737.

Stockman, A., Sharpe, L. T., & Fach, C. C. (1999). The spectral sensitivity of the human short-wavelength cones. *Vision Research*, *39*(17), 2901–2927.

Thibos, L. N., Bradley, A., Still, D. L., Zhang, X., & Howarth, P. A. (1990). Theory and measurement of ocular chromatic aberration. *Vision Research*, *30*(1), 33–49.

Thibos, L. N., Hong, X., Bradley, A., & Cheng, X. (2002). Statistical variation of aberration structure and image quality in a normal population of healthy eyes. *Journal of the Optical Society of America A*, *19*(12), 2329–2348.

Tkacik, G., Garrigan, P., Ratliff, C., Milcinski, G., Klein, J. M., Sterling, P., … Balasubramanian, V. (2011). Natural images from the birthplace of the human eye. *PLoS ONE*, *6*(6:e20409).

Van Gelder, R. N., & Buhr, E. D. (2016). Ocular photoreception for circadian rhythm entrainment in mammals. *Annual Review of Vision Science*, *2*.

von Kries, J. (1902). Chromatic adaptation. In *Sources of color vision* (pp. 109–119). Cambridge, MA: MIT Press.

Vrhel, M., Gershon, R., & Iwan, L. (1994). Measurement and analysis of object reflectance spectra. *Color Research And Application*, *19*(1), 4–9.

Wald, I., Dietrich, A., Benthin, C., Efremov, A., Dahmen, T., Gunther, J., … Slusallek, P. (2006, September). Applying ray tracing for virtual reality and industrial design. In *2006 IEEE symposium on interactive ray tracing* (pp. 177–185). ieeexplore.ieee.org.

Wald, I., Purcell, T. J., Schmittler, J., Benthin, C. *et al*. (2003). Realtime ray tracing and its use for interactive global illumination. *Eurographics, State of the Art Reports*.

Wandell, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer.

Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision*, *17*(3), 10.

Watson, A., & Pelli, D. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics*, *33*(2), 113–120.

Wei, W. (2018). Neural mechanisms of motion processing in the mammalian retina. *Annual Review of Vision Science*, *4*, 165–192.

Whitehead, A., Mares, J., & Danis, R. (2006). Macular pigment. A review of current knowledge. *Archives of Ophthalmology*, *124*(7), 1038–1045.

Wikipedia contributors. (2021, July). *Lytro.* https://en.wikipedia.org/w/index.php?title=Lytro&oldid=1032099081. (Accessed: July 8, 2021.)

Williams, D. R., MacLeod, D., & Hayhoe, M. (1981). Foveal tritanopia. *Vision Research*, *19*(9), 1341–1356.

Williams, D. R., Sekiguchi, N., Haake, W., Brainard, D. H., & Packer, O. (1991). The cost of trichromacy for spatial vision. In *From pigments to perception* (pp. 11–22). New York: Plenum Press.

Wyszecki, G. (1958). Evaluation of metameric colors. *Journal of the Optical Society of America*, *48*(7), 451–454.

Wyszecki, G., & Stiles, W. (1982). *Color science: Concepts and methods, quantitative data and formulae*, 2nd ed. New York: John Wiley & Sons.

Yellott, Jr., J. I., Wandell, B. A., & Cornsweet, T. N. (1984). The beginnings of visual perception: The retinal image and its initial encoding. In I. Darien-Smith (Ed.), *Handbook of physiology: The nervous system* (Vol. III, pp. 257–316). New York: Easton.

Young, T. (1802). On the theory of light and colours. *Philosophical Transactions of the Royal Society of London*, *92*, 20–71.

Zhang, L., Cottaris, N. P., & Brainard, D. H. (2021). An image reconstruction framework for characterizing early vision. *bioRxiv*. doi: 10.1101/2021.06.02.446829