

From Light to Vision: An Introduction

Yuhao Zhu

Department of Computer Science
Department of Brain and Cognitive Sciences
University of Rochester

yzhu@rochester.edu
<https://yuhaozhu.com/>
<https://horizon-lab.org/>

1 The Big Picture

Before studying the HVS, it is useful to start by discussing why we care about the HVS at all — after all, if you are computer science and/or engineering student, why would you care? We will then discuss the methodology we will use when studying the HVS.

1.1 Why Do We Study HVS?

Why do we care about studying the HVS? First and foremost, for the science itself — it is extremely satisfying to just understand “how stuff works”, is it not? Understanding the basics of the HVS will also allow us to investigate the unknowns of the HVS, and computer scientists have a lot to off. For instance, modern computational methods, especially deep (artificial) neural networks have provided us a new toolbox to better understand the biological neural networks. For instance, if a signal representation or a learning paradigm is effective in deep neural networks, would it be possible that our HVS uses a similar representation or can learning based on similar representations?

For computer scientists and engineers working on visual computing systems, there is another reason, and that is illustrated in Figure 1. The psychological experiences of the users of a computing platform, be it an AR/VR headset or a smartphone, is what we want to influence, but we, for the most part, exert that influence *indirectly*, by designing and optimizing the imaging, rendering, and computer systems. The outputs of these systems, i.e., the visual stimuli coming out of the display, become the input to the HVS of the a human whose psychological states we care to optimize. So if we understand the HVS, we could invert the HVS process, given the desired psychological states, to solve for the optimal visual stimuli, and from there we can then think about how to best design the various engineered systems.

Understanding the cellular, molecular, and neural processes in the HVS has also inspired people to better engineer systems such as imaging systems [Liao et al., 2022, Wodnicki et al., 1995] and deep neural networks, even those the output of these systems are not meant to be consumed by the HVS [Idrees et al., 2024].

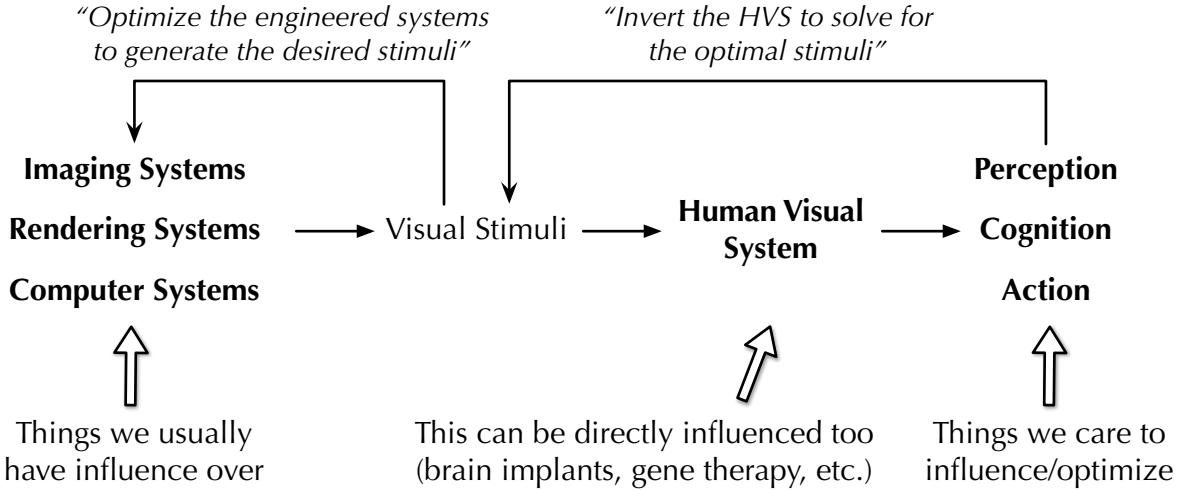


Figure 1: Understanding the HVS helps us better design rendering, imaging, and computer systems.

1.2 How Do We Study HVS?

How do photons in the real-world give rise to perception and cognitions in our brain when they enter our eyes? We want to show you that there is really no magic here. The perception and cognition we experience are fundamentally a result of the complicated (first optical and eventually electrical) signal processing in the physiological systems — our eyes and brains.

This relationship between low-level electrical signals and high-level behavioral responses in human is conceptually no different from one that we find in computers. This comparison is shown in Figure 2. For someone who is unfamiliar with computer systems and chip design, it would seem rather magical that a computer does what it does. But we know that the high-level, observable behaviors of a computer program are a result of low-level processing in the electrical circuits. Similarly, the experiences humans have in response to visual stimuli are a result of the collective behaviors of the underlying neurons in the nervous system, whose behaviors are a result of the cellular and molecular processes within and between individual neurons.

The circuits in a computer are made of engineered materials such as transistors, whereas circuits in the HVS are made of biological materials such as neurons. Fundamentally, however, it is all physics — electrons and/or ions move around and cause changes in voltage potentials and currents, and these changes are how information is propagated.

With the advancements in modern science and engineering, we can now measure, at a neuronal or even sub-neuronal level, the electrical responses of the HVS when presented with visual inputs. These measurements allow us to *correlate* electrical responses to perception and cognition, which, in turn, allow us to say something like “this part of the HVS supports or is responsible for that particular function (e.g., object detection).” It is important to note, however, that we still do not know why the electrical responses *cause* our perception and cognition.

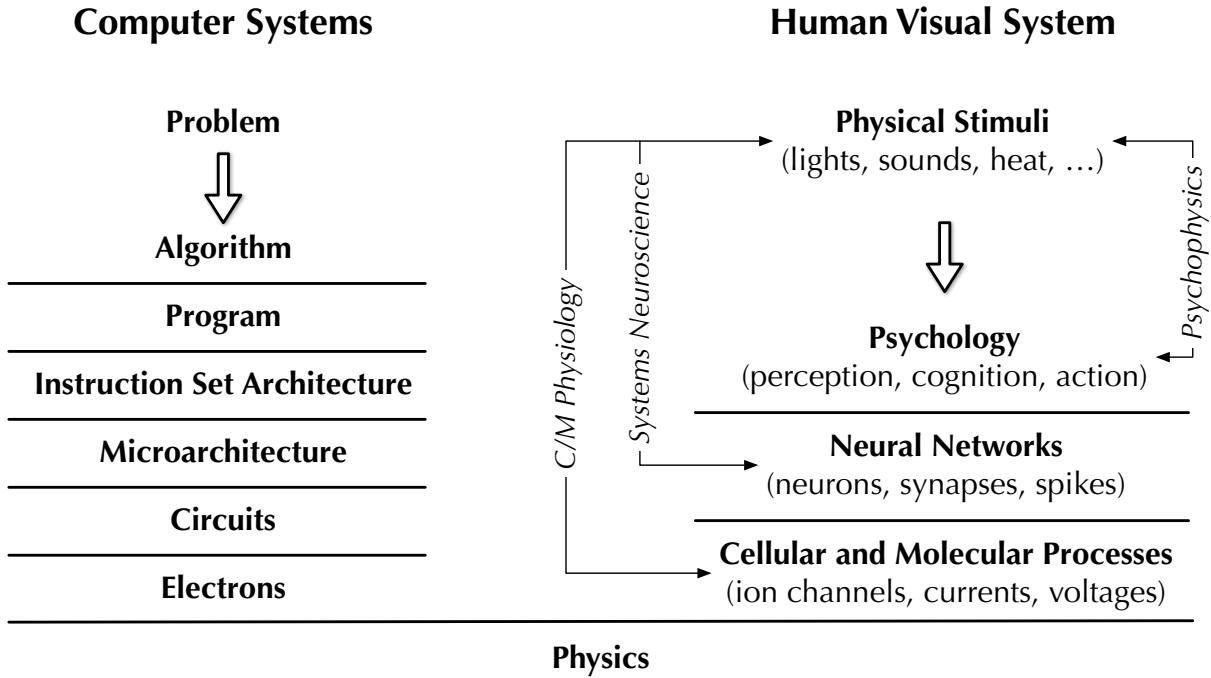


Figure 2: Just like in a computer system, HVS also involves a sequence of transformations and can be studied at different levels of abstraction.

The causation problem, for the moment, is at best a philosophical problem or, if you will, a religious one.

The goal of this Chapter is to give you an overview of the Human Visual System (HVS). We will focus on the main components and key facts of the HVS so that you can start appreciating the connections between signal processing at the physiological level and perception, cognition, and action at the behavioral level — while leaving many details to later chapters.

The signal processing in the HVS consists of three main components; this is illustrated in Figure 3. First, lights are processed in the optical domain as they enter our eyes and go through the eye optics. The optical signals then reach the retina and are first converted to electrical signals, which are further processed before exiting the retina. The retina outputs encode low-level information such wavelengths, contrast, timing of object motion, etc. The retinal outputs are then transmitted to the Lateral Geniculate Nucleus (LGN) and, for the most part, relayed to the visual cortex. Cortical processing essentially knits together the low-level, upstream information to give us vision. The retino-geniculo-cortical pathway is the main pathway for the electrical signals.

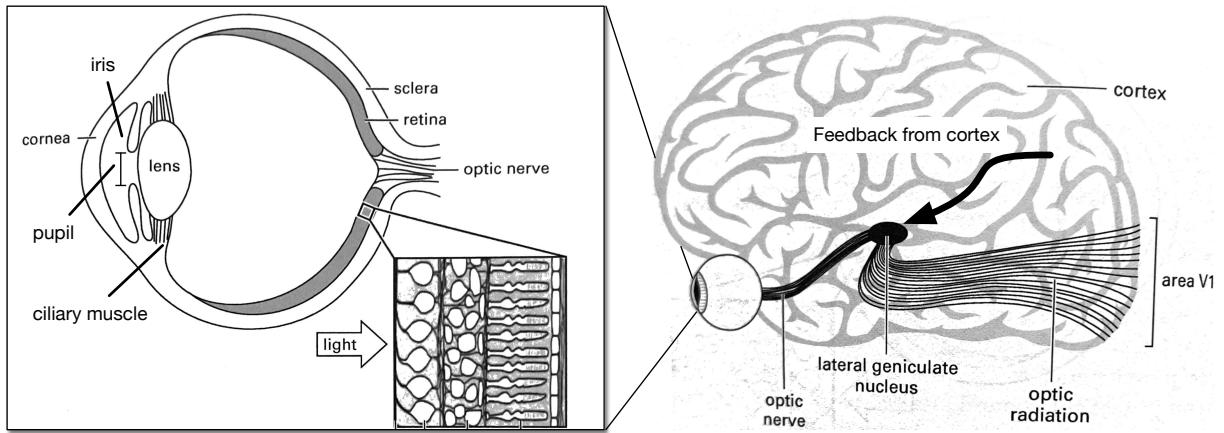


Figure 3: Pupil, under the control of the iris, let in lights. Cornea and lens focus lights with the former contributing the most optical bending power. Lens contracts and relaxes to accommodate object depth under the control of the ciliary muscle. Retina transforms optical signals to electrical signals, which are further processed and exit the retina through the optic nerve. Retinal signals go through the Lateral Geniculate Nucleus and then are projected to the visual cortex. This retino-geniculo-cortical pathway carries the main information flow in the HVS with the cortex also providing feedback to the LGN. Adapted from [Dowling and Dowling Jr \[2016\]](#), Fig. 1.1, Fig. 1.2].

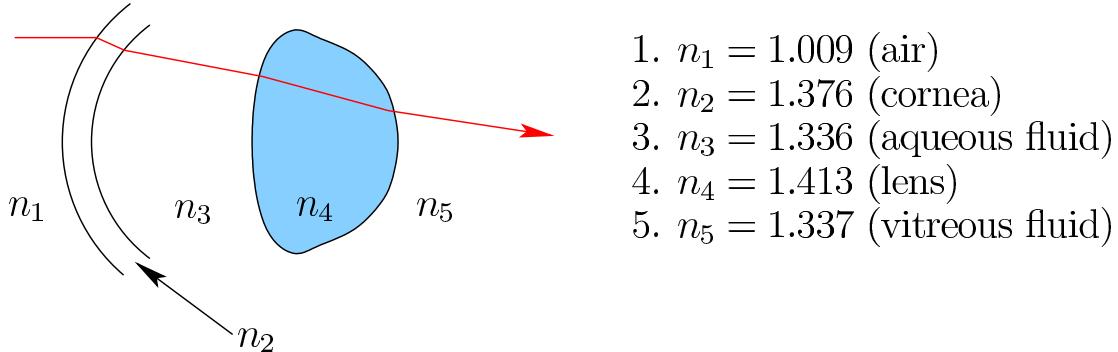
2 Eye Optics

The optical signal impinging on the retina is called the **optical image**, which is a 2D continuous signal in that at any position on the retinal surface we can ask: how much optical power there is here; that is, what is the irradiance here? Ideally, the optical image is a perfect perspective projection from the 3D physical world, with no loss of information other than the projection. The reality is much more complicated.

The Main Goal is to Focus Lights

The main goal of the eye is to focus lights on the retina. To focus lights the optics need to bend lights, which is achieved collectively by both the cornea and the lens. This is illustrated in Figure 4. Lights bend because of the difference in refractive index in adjacent ocular media. Most of the bending is done by the cornea, because there are large differences between the cornea and its adjacent media (air and aqueous fluid). The lens also contributes to light bending, albeit with a lower contribution, because the differences in refractive index between the lens and its adjacent media (aqueous fluid and vitreous fluid) are relatively small.

The cornea is fixed in shape. Lens, in contrast, is malleable in its shape. The ciliary muscle controls the contraction and relaxation of the lens, which changes the focal length, and thus bending power, of the lens, and by extension the entire eye optical system. Adjusting the focal



1. $n_1 = 1.009$ (air)
2. $n_2 = 1.376$ (cornea)
3. $n_3 = 1.336$ (aqueous fluid)
4. $n_4 = 1.413$ (lens)
5. $n_5 = 1.337$ (vitreous fluid)

Figure 4: Much of the optical bending power in eye is contributed by the cornea, which has a large refractive index difference with respect to its adjacent ocular media (Snell's law). The lens also contributes to light bending, albeit with a lower contribution. Cornea is rigid but lens is malleable, so accommodation is attributed exclusively to the lens. From [LaValle \[2023, Fig. 4.25\]](#)

length to bring an object into focus is called **accommodation**.

But if the ciliary muscle cannot properly adjust the lens, we get defocused blur, which is a form of optical **aberration**. There are a number of other optical aberrations; astigmatism and chromatic aberration are two common ones found in eyes. While not an optical aberration, diffraction also plays a role when the pupil size is very small (e.g., under strong illumination).

For our purpose, “imperfections” introduced by eye optics (aberration and diffraction) can be modeled by the Point Spread Function (PSF) of the optical system, which we will see later in the class.

Ocular Media Absorb Light Selectively

While all the ocular media are generally transparent, they still absorb some amount of lights. Critically, the absorption and, by extension, transmittance, is strongly wavelength dependent. Color vision is fundamentally tied to the power distribution of light over wavelengths, so the selective absorption of light by the ocular media significantly influences our color vision.

[Boettner and Wolter \[1962\]](#) measured the spectral transmittance of the eye, and the results are shown in Figure 5. Each curve represents the percentage of light remaining at each ocular media and the retina (including both direct transmission and forward scattering). Considering the visible range (we will discuss in the next Chapter why there are even invisible lights) roughly between 380 nm and 780 nm, we can see the ocular media significantly reduces the light power at short wavelengths.

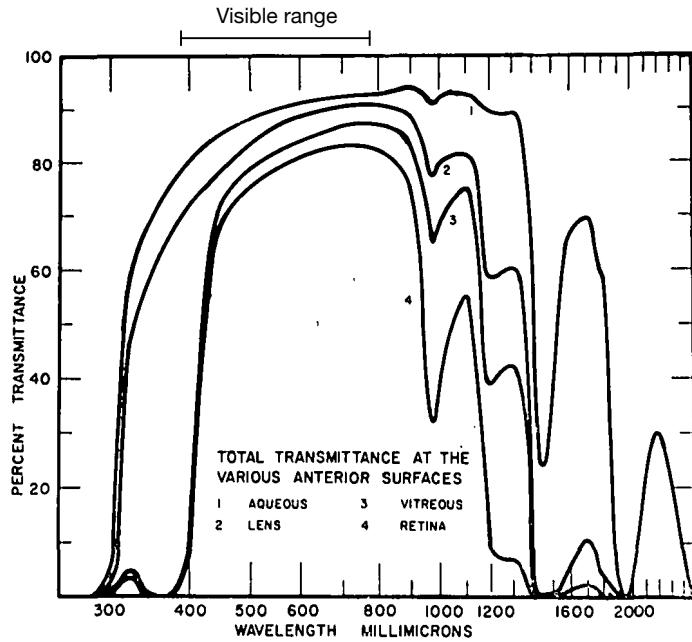


Figure 5: Much of the optical bending power in eye is contributed by the cornea, which has a large refractive index difference with respect to its adjacent ocular media (Snell's law). The lens also contributes to light bending, albeit with a lower contribution. Cornea is rigid but lens is malleable, so accommodation is attributed exclusively to the lens. Adapted from [Boettner and Wolter \[1962, Fig. 7\]](#).

3 Retina

Now the photons have arrived at the retina. The retina is where optical signals are transformed to electrical signals. The electrical signals undergo further processing on the retina and are then carried by the optic nerve to the brain. The signal transduction and processing are carried out through layers of neurons on the retina, of which there are five categories (each of which has sub-categories). They are photoreceptors, bipolar cells, horizontal cells, amacrine cells, and retina ganglion cells (RGCs).

The main information flow starts from the photoreceptors, flows through the bipolar cells, which synapse with photoreceptors and send their outputs to RGCs. Horizontal cells synapse with photoreceptors (and other horizontal cells), and amacrine cells connect with both bipolar and RGCs (and other amacrine cells). Identifying the different classes of neurons and their connections is largely due to Santiago Ramón y Cajal, who won the Nobel Prize in 1906.

Interestingly, while we might be used to neurons communicating through spikes, i.e., action potentials (which were discovered by Hodgkin and Huxley [[Hodgkin and Huxley, 1952](#)], who won the Nobel Prize in 1963), RGCs are the only type of neurons on the retina that spike. The rest of the neurons are non-spiking neurons; they communicate through graded potentials.

Optical-to-Electrical Signal Transduction Takes Place in Photoreceptors

Photoreceptors are where optical signals are transformed to electrical signals. Photoreceptors absorb incident photons; once a photon is absorbed, it could generate electrical responses through the process of **phototransduction** cascade [Wald, 1968]. George Wald won his Nobel Prize in Physiology or Medicine by essentially elucidating this process. The electrical response can be represented as photocurrents or, equivalently, photovoltages across the cell membrane of the photoreceptor. We will have a lot to say about this process later in the class.

Functional and Anatomical Organizations of the Retina are Opposite

The functional organization of the cells is opposite of the anatomical organization of the cells. Functionally, the first layer of the retina is the photoreceptor cells, which convert photons to electrical responses, and the last layer is the RGCs, which carry all the retinal output information and are directly connected to the optic nerve, which are effectively the axons of the RGCs. Anatomically, however, RGCs lie at the outermost layer of the retina and the photoreceptors are the innermost layer. Therefore, photons upon reaching the retina photons first hit the RGCs and go through other neurons before eventually hitting the photoreceptors, where the signal transduction takes place. As far as a photon is concerned, neurons before the photoreceptors are transparent and simply let the photon through without doing much about it — with an exception that we will see soon.

Blind Spot Exists Because of the Routing Issue

An implication of the anatomical organization is that the optic nerve has to be routed from the front of the retina and *through* the retina at a single location, which is called the **optic disk**. The optic disk must be free of any neurons, including photoreceptors, simply for the optic nerve to exit. Since photoreceptors sense light, the optic disk is also called the blind spot. This is illustrated in Figure 6. Some vertebrates like octopus do not have this “wiring” issue, since their retina signals exist from the back of the retina.

It is unclear whether there are evolutionary advantages of having a blind spot on our retina, but it does not seem to be a disadvantage: we clearly do not notice the blind spot in our daily life — because the downstream visual system fills in the missing information there. Our head and eye movements further mitigate the impact of the blindspot.

ipRGCs are Light-Sensitive but Do Not Contribute to Image-Forming Vision

Photoreceptors are the only type of neurons on the retina that are sensitive to light *and* contribute to image-forming vision. There is another type of neurons, a sub-type of RGCs actually, called **intrinsically photosensitive RGCs** (ipRGCs), that are also sensitive to light (i.e., they absorb photons and convert optical signals to electrical signals), but interestingly they do not (primarily) contribute to image-forming vision.

ipRGCs were discovered very recently, and it is fair to say that the discovery was a big deal for the field [Berson et al., 2002, Hattar et al., 2002]. For the past 150 years or so, human

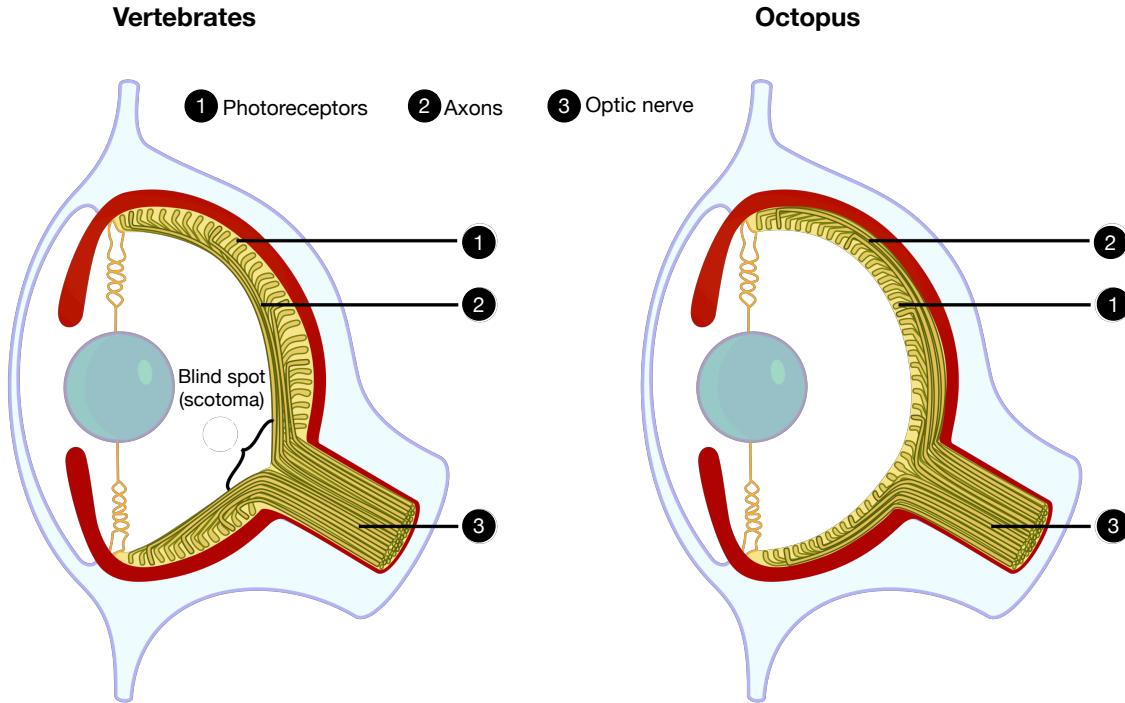


Figure 6: Vertebrate eyes have a blind spots (scotoma) because RGC axons exit the retina from the *front* of the retina. It is purely a “wiring” issue. Octopus eyes do not have this issue. Adapted from [Caerbannog \[2016\]](#).

vision can be adequately explained by photoreceptors being the only light-sensitive neurons. Now if ipRGCs are also light sensitive, do we have to re-write the science behind human vision? It turns out the while ipRGCs do respond to lights, they primarily contribute to non-image-forming vision. For instance, they are shown to impact circadian rhythms, mood, and pupillary light reflex [[Lazzerini Ospri et al., 2017](#), [Do and Yau, 2010](#)].

4 Retina Structure and Functions

Retina is organized to perform a set of low-level tasks that are crucial to vision. “Low-level” here refers to the fact that information encoded by retina are building blocks to support more complicated visual functions later in the HVS. At the risk of over-simplification, each task is achieved by a **visual stream** of neurons. These visual streams are also called **parallel pathways**. This section briefly discusses a set of basic functions of the retina and their visual streams.

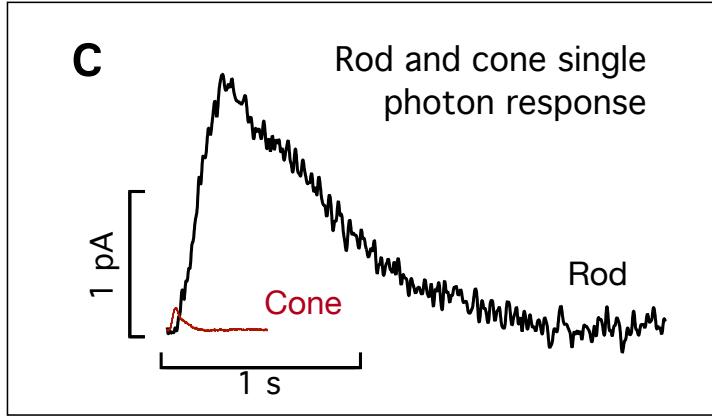


Figure 7: Single photon responses (photocurrents) of a primate rod and cone. Rods are more sensitive with a slower kinetics. Adapted from [Angueyra-Aristizábal \[2014, Fig. 1.4C\]](#).

4.1 Rod vs. Cone Specialization

Sensitivity and Kinetics

There are two types of photoreceptors: rods and cones. Perhaps the most important difference between the two is that rods are much more sensitive to light than cones. This is evident in Figure 8, which compares the single-photon response of rods and cones in primates. The response here is represented by photocurrent, which we will talk in detail later in the class.

Due to the high sensitivity, rod responses saturate quickly as the ambient light level increases, so they are primarily responsible for vision at low illumination levels (e.g., at night); rod-mediated vision is called the **scotopic vision**. Cones are much less sensitive so they are responsible for vision at normal illumination levels such as during the day. Cone-mediated vision is called the **photopic vision**. Figure 8 shows the luminance range that both the scotopic and the photopic vision are sensitive to. The sensitivity range overlap, so there is luminance range where both rods and cones contribute to vision, and that is called the **mesopic vision**.

Cones also have a faster response kinetics than rods: their responses rise and fall much faster than rods. The faster kinetics allow cones to track moving objects better than rods. Only for the purpose of understanding this, think of a camera where the expose time is very long; the capture image is blurred. Shorter exposure/shutter time captures motion better. Cones have a shorter effective “exposure time” than rods.

Spectral Sensitivity and Color Vision

Yet another important difference between rods and cones is that cone-mediated vision provides color information whereas rod-mediated vision encodes only light intensity but not color. This is because there is only one type of rods but three different types of cones, each with a different (linearly independent) wavelength sensitivity function. Fundamentally, color arises from the

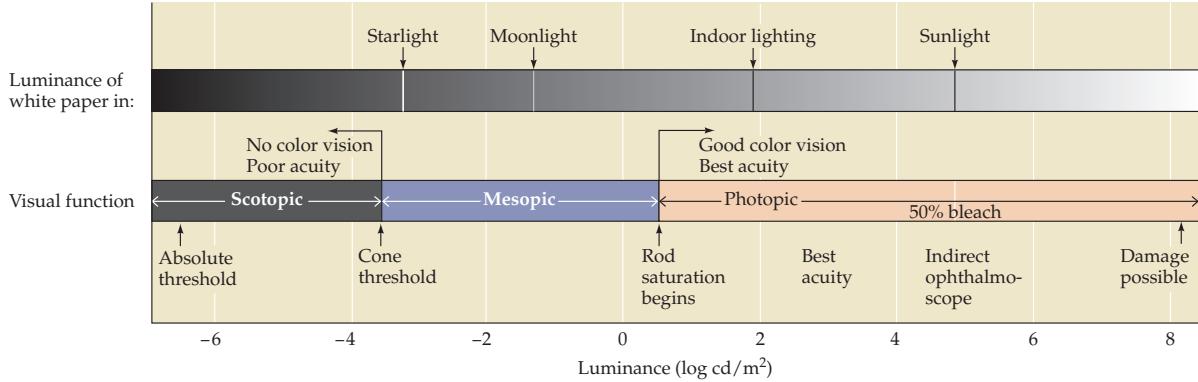


Figure 8: Sensitivity range of rod-mediated vision and cone-mediated vision. From [Purves et al. \[2017\]](#), Fig. 11.11].

wavelengths information in incident lights. Having three types allows cones to have a stronger capability of encoding wavelength information than rods.

There are two main ways to measure the sensitivity of a photoreceptor physiologically, and they are ultimately equivalent; we will have a lot to say about this later. One way is to measure the amount of light needed at each wavelength to achieve a criterion level of electrical response from a photoreceptor. An example of such data is shown in the left panel of Figure 9, collected by [Baylor et al. \[1987\]](#) on a macaque. The other way is to measure the fraction of photons at each wavelength that gets absorbed by a photoreceptor. [Dartnall et al. \[1983\]](#) collected one set of such data from human donors, shown in the right panel of Figure 9, which also shows the data for the rods in the human participants. The y -axis is *absorbance*, which is $\log(I_{\text{incident}}/I_{\text{transmitted}})$, i.e. the log value of the ratio between the incident light intensity and transmitted (i.e., unabsorbed) light intensity¹.

The three cone types peak at different wavelengths in their sensitivity; we call them Long (L), Medium (M), and Short (S) cones. It is interesting to notice that L and M cones are more similar to each other than they are to the S cones. This is a clue about the evolution of the three cone types: the divergence between L and M are fairly recent (about 30 to 35 million years ago) [[Jacobs, 2008](#)].

Spatial Distribution

There are about 120 million rods and about 6 millions cones. The left panel in Figure 10 shows the distribution of both cones and rods on the retina. Almost all the cones are concentrated at **fovea**, a small, central pit on the retina that is approximately 2 mm in diameter and subtends a visual angle of about 1°. The position in the fovea that has the peak cone density is defined

¹We can show that absorbance is proportional to absorption (i.e., fraction absorbed) using the Beer–Lambert law when the length of light path is short, which is the case here since the photoreceptors are transversely illuminated.

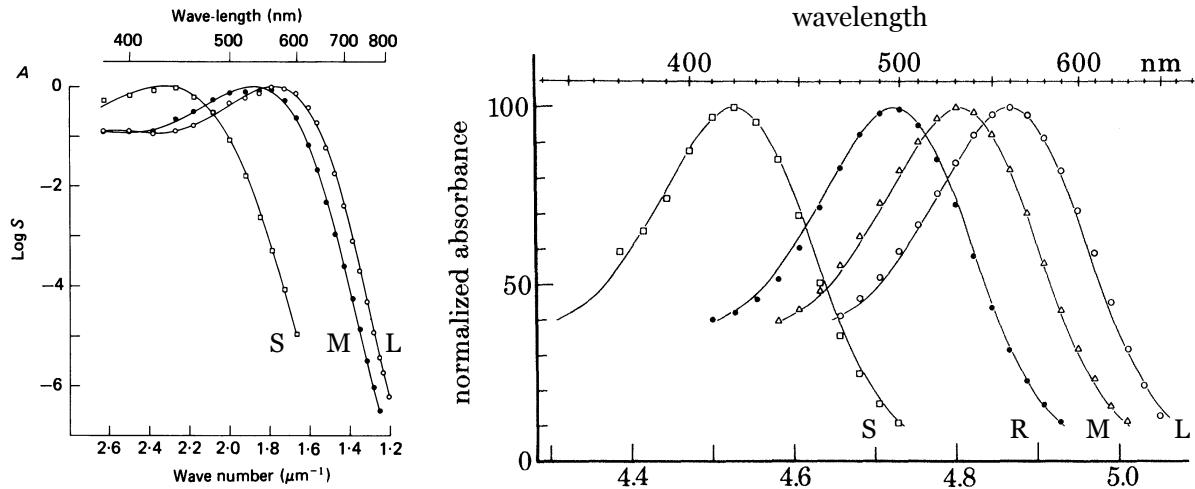


Figure 9: Left: Log-scaled spectral sensitivities of the cones (L, M, S) in a macaque; adapted from [Baylor et al. \[1987\]](#), Fig. 3A]. The spectra are normalized to peak at 1. Right: Absorbance spectra of the three cones (L, M, S) and the rod (R) in a human; adapted from [Dartnall et al. \[1983\]](#), Fig. 2]. The spectra are normalized to peak at 100.

to have an **eccentricity** of 0° . There are no rods in the fovea; all the rods are placed at the retina periphery, peaking at about 20° away from the fovea.

The right panel in Figure 10 are images of photoreceptors at the fovea and at the periphery, taken by [Curcio et al. \[1990\]](#). Cones exclusively occupy the fovea and they become sparser and larger in the periphery. Rods fill in the spaces in the periphery.

There are many important implications of the photoreceptor mosaic and distribution. First, the visual acuity decreases in the visual periphery. Think of photoreceptors as sampling the continuous optical image impinged upon the retina. A higher density leads to a higher sampling rate. In addition, larger cone sizes in the periphery are equivalent to higher degrees of blurring, since photons hitting a cone are integrated together just like by a camera pixel (although, critically, the electrical response of a cone is *not* proportional to the photon count, unlike a camera pixel; we will have a lot to say about this in the next Chapter); integration is a form of low-pass filtering.

We hasten to add that the lower acuity in the periphery is *not* exclusively attributed to the photoreceptor mosaic. As we will see shortly, how photoreceptors communicate with other neurons on the retina places an important role too.

Second, since the fovea has the highest visual acuity, our ocular motor system is evolved in such a way that when we want to see fine details of an object, we move our eyes so that lights from the objects hit the fovea. This means that we cannot see fine details of an object in dim environments if we fixate at it. Instead, we would have a better chance of seeing details if we intentionally place the object at our peripheral vision in a low-light environment.

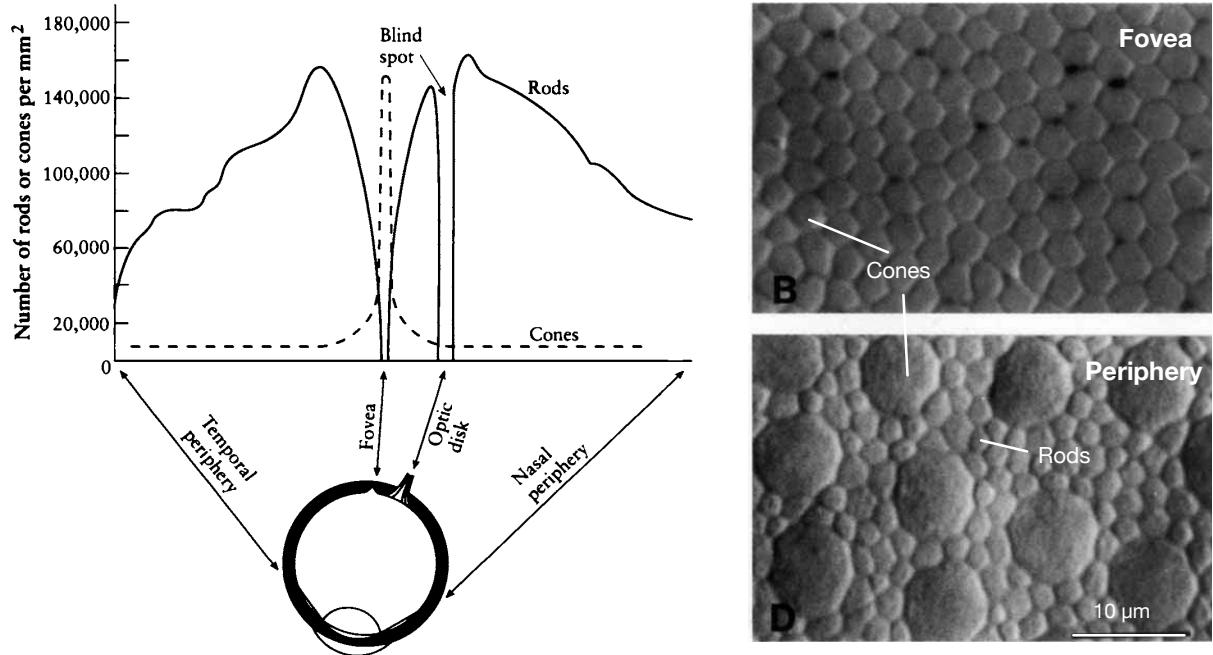


Figure 10: Left: cone and rod distribution on the retina; x-axis is the eccentricity (angular distance from the fovea, which has an eccentricity of 0°). From Glassner [2014, Fig. 1.4, p. 10]. Right: photos of photoreceptors in the fovea and periphery; rods are absent in the fovea, and cones become sparser and larger in the periphery. Adapted from Curcio et al. [1990].

Rod vs. Cone Pathways and Visual Streams

Rods and cones have their own pathways initially and merge later. This is shown in Figure 11. Both rods and cones synapse with bipolar cells, but they synapse with distinct bipolar cells. That is, an individual bipolar cell receives information from rods only or cones only. Rod pathway and cone pathway are parallel streams at this point. Bipolar cells then feed their outputs to RGCs. A RGC can mix information from both rod and cone bipolar cells. This mixing is enabled by amacrine cells, which synapse with both rod and cone bipolar cells and with RGCs. So the distinct information in the rod pathway and the cone pathway gets merged in the RGC layers.

Why are rod and cone pathways initially parallel but merge later? The initial parallel pathways allow rods and cones to extract low-level information, such as contrast, independently under different lighting conditions, but once the information is collected it is processed similarly so there is really no need to duplicate the processing circuitry.

4.2 Contrast Detection and Adaptation

Another important function of the retina is to extract contrast information. Arguably most interesting information in the physical world exists all in image contrast, i.e., local differences in

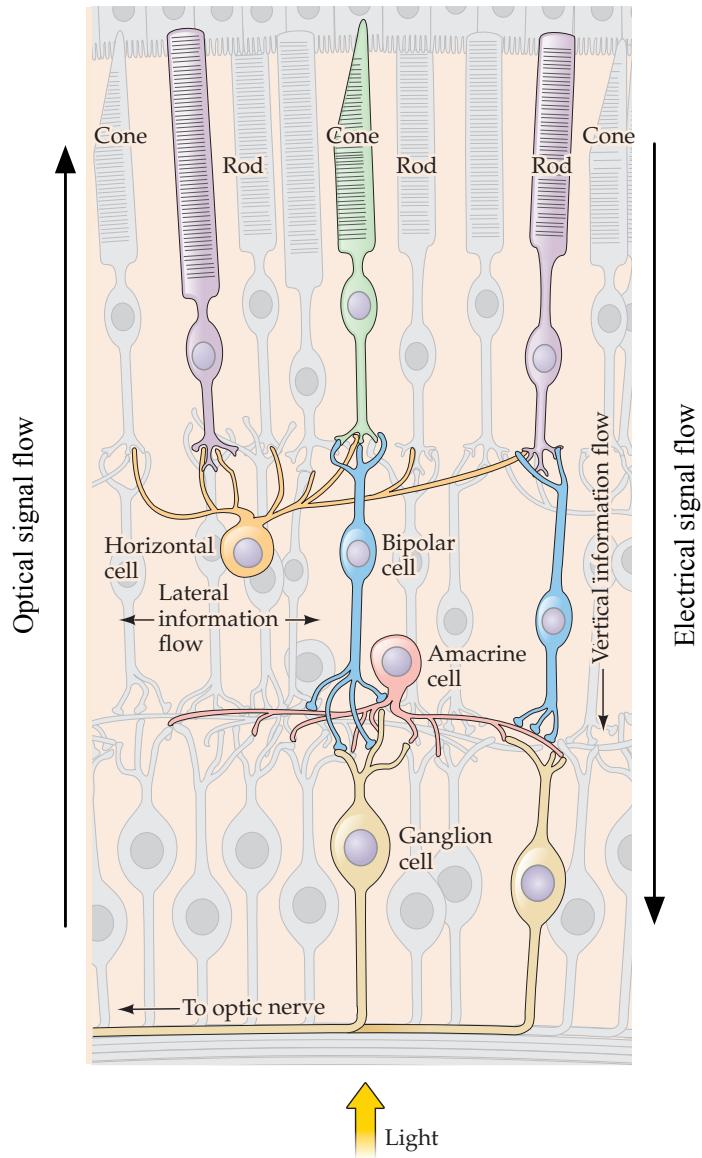


Figure 11: The basic neural network on the retina. The photoreceptors convert optical signals to electrical signals. The electrical signals go through the bipolar cells and then to the retinal ganglion cells (RGCs), which carry all the output of the retina. Horizontal and amacrine cells mediate lateral interactions, giving rise to important features such as the receptive field. Since the RGCs are at the outer most layer of the retina, the optical information and the electrical information flow in opposite directions. Adapted from Purves et al. [2017], Fig. 11.5B, p. 219]

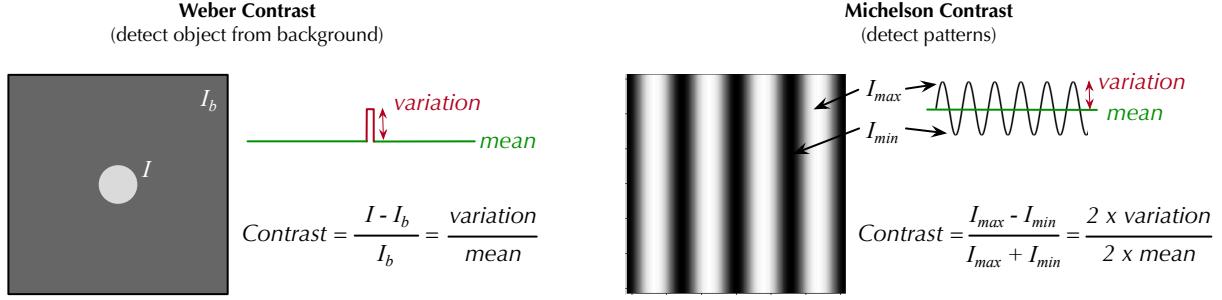


Figure 12: Weber contrast is often used for detecting objects against a uniform background, and Michelson contrast is used for detecting patterns. The two definitions are compatible: they both describe the ratio between the maximal variation of the signal over the mean.

light intensities. Take a look at your surroundings; uniform light levels where there is absolutely no change in light are rare and do not present much useful information. Fine details of an objects are really encoded in contrast.

This imposes two requirements in our visual system. First, we need to be able to extract contrasts and encode them in neural signals so as to be processed by the rest of the brain. Second, we must reliably encode contrast across a wide range of ambient light levels. Before discussing how the RGCs meet these requirements, we will first define contrast more rigorously.

Contrast is Variation Over Mean

Intuitively, contrast describes how much variation there is in a signal relatively to the average strength of the signal. There are two commonly used definitions, both of which are compatible with this intuition. They are usually used in different scenarios. Figure 12 illustrates the two definitions.

Weber contrast is often used in scenarios where there is a small object against a relatively uniform background. The contrast C_w is defined as:

$$C_w = \frac{I - I_b}{I_b}, \quad (1)$$

where I_b is the background luminance, I is the object luminance. If the object is small, the mean luminance of the entire field is approximately the background luminance, and naturally $I - I_b$ is the maximal variance over the mean.

The Michelson contrast is used in scenarios where are patterns in a scene. Taking a sinusoidal pattern as an example (and recall any arbitrary pattern can be decomposed into sinusoidal basis patterns), the contrast C_m is defined as:

$$C_m = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}, \quad (2)$$

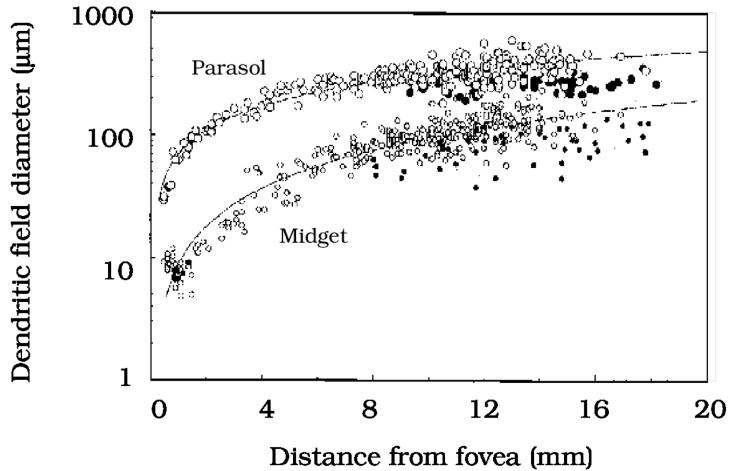


Figure 13: Dendritic field sizes (of two RGC subtypes) increase with eccentricity, indicating a higher degree of neural convergence at the periphery. From [Wandell \[1995, Fig. 5.7\]](#), which is after [Dacey and Petersen \[1992, Fig. 2A\]](#).

where I_{max} and I_{min} are the highest and lowest luminance, respectively, of the signal. We can see that C_m is also the ratio between the variation and the mean of the signal.

RGCs Pools Signals from Photoreceptors in Its Receptive Field

There are about 120 million rods, 6 million cones, and 1 million RGCs on the retina. Therefore, a single RGC *necessarily* receives signals from multiple rods and/or cones. Pooling signals from multiple neurons into a single neuron is generally called **neural convergence**, a many-to-one mapping. Evidently, there is a much higher degree of neural convergence in rods than in cones. The fovea, which recall contains only cones, is an extreme case, where there is no neural convergence; in fact, each foveal cone sends its signal to multiple RGCs, so there is a one-to-many mapping there.

The higher degree of neural convergence in the rod pathway is another reason why rod-mediated vision is more sensitive than cone-mediated vision: the responses of different rods that are pooled together to the same downstream RGC, so that RGC could generate responses faster to the brain than if the RGC receives inputs from only a single cone at the fovea. The flip side of the higher degree of convergence is that rod vision offers low spatial acuity. If a RGC generates a response, we could not resolve the source of that response since it could come from anywhere within a large group of photoreceptors being stimulated. From a signal processing perspective, summation is a form of low-pass filtering (equivalent to convolving the signal with a box filter), which naturally reduces the frequency of the signal.

The degree of neural convergence increases as the eccentricity increases. Figure 13 shows the dendritic field sizes of two RGC subtypes; the size increases with the eccentricity. The higher degree of neural convergence is another reason why peripheral acuity is much worse than that

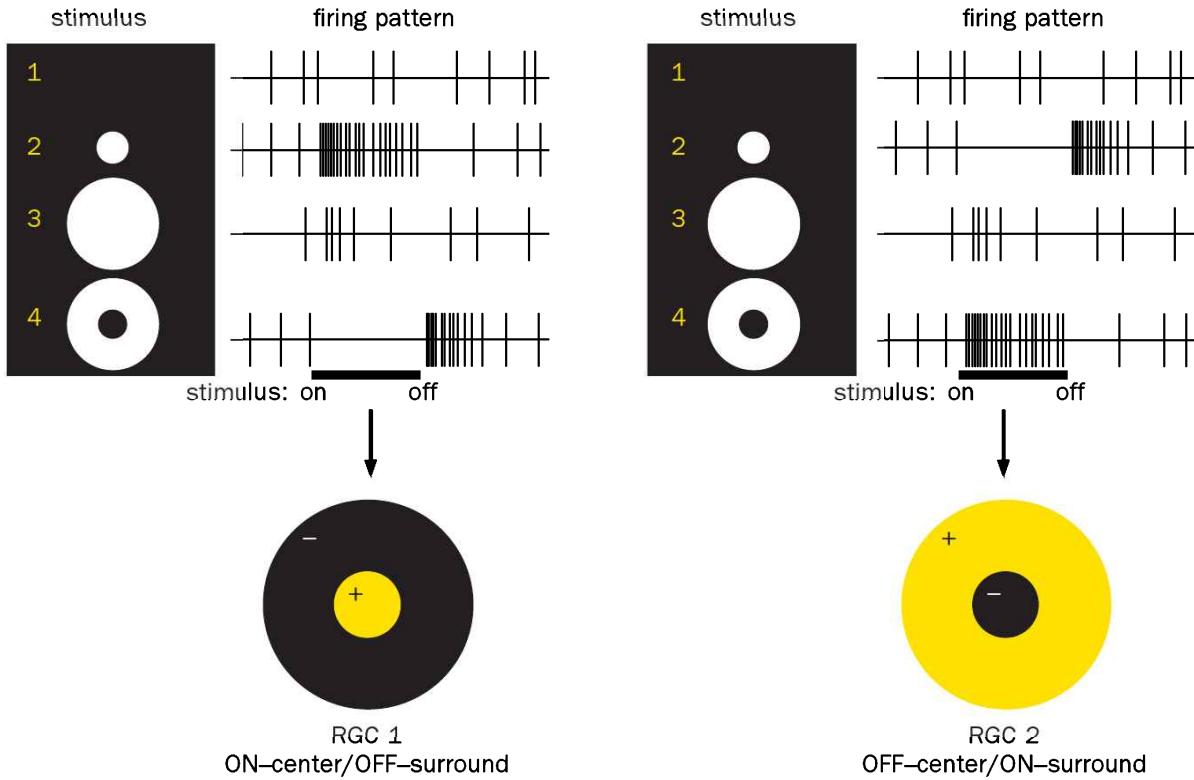


Figure 14: RGCs have a center-surround receptive field with two types. The ON-center RGCs are excited by stimuli presented at the center but inhibited by stimuli presented at the surround (stimulus 2 on the left); OFF-center RGCs have the opposite response (stimulus 4 on the right). From [Luo \[2016, Fig. 4-24\]](#), which is after [Hubel \[1995, p. 41\]](#).

at the fovea.

RGCs Have a Center-Surround Receptive Field

Neural convergence gives rises to an important concept called **receptive field**, which is central to contrast encoding. The receptive field of a neuron is the area of the retina from which activity of the neuron can be influenced. For a RGC, its receptive field is the collection of photoreceptors whose output signals converge at that RGC. Due to the one-to-mapping relationship at the fovea, the RGCs that are connected to fovea cones have a receptive field of only one cone.

The way a RGC aggregates information from the receptive field is *not* to simply sum up the signals from the individual photoreceptors. If we illuminate the entire receptive field of a RGC uniformly, the RGCs respond similarly regardless of the illumination intensity. Let's call it the spontaneous rate.

If uniformly changing the light levels does not change the RGC's response rate, what will? It turns out that you need to have *variations* in the illumination within the receptive field. RGCs

respond best to variation patterns that have a center-surround structure. For about half of the RGCs, their response rate is maximized if we present bright lights to the center photoreceptors and dark lights to the surround photoreceptors. These are called **ON-center** RGCs, since they have an excitatory center (excited by light) and inhibitory surround (inhibited by light). The other half prefer the opposite pattern: dark at the center and bright at the surround. They are the **OFF-center** RGCs, since they have an inhibitory center and an excitatory surround. RGCs are said to have a **center-surround** receptive field. Figure 14 illustrates the receptive fields of the two RGCs.

H.K. Hartline, who won the Nobel Prize in 1967, measured the RGC responses from horseshoe crabs [Hartline and Graham, 1932], using which he famously demonstrated inhibitory signals [Hartline, 1949, Hartline et al., 1956]; he was also the first to use the term receptive field [Hartline, 1938, 1939, 1940a,b]. Barlow [1953] demonstrated the inhibitory signals in a frog's RGC, and Kuffler [1952, 1953] was the first to demonstrate the center-surround receptive field structure in a mammalian (cat) RGC with Barlow making significant contributions to that too [Barlow et al., 1957].

Center-Surround Receptive Fields are Designed to Encode Contrasts

Looking at the preferred stimulus of the two RGC types in Figure 14 (stimulus 2 for ON-center and stimulus 4 for OFF-center), evidently the RGCs are designed to extract illuminant variations, i.e., contrast. If a visual field has a high (positive) Weber contrast, i.e., there is a small object that is significantly lighter than the background, the ON-center RGC would respond well to it. Similarly, an OFF-center RGC would respond well to a dark object over a uniform background.

We can also quantify the how the center-surround receptive fields respond to patterns of different Michelson contrast. A complication is that a pattern is described not only by its contrast but also frequency. At each frequency, we determine the minimal amount of contrast needed to produce a criterion level of RGC response (say 30 spikes/second). The contrast sensitivity at that frequency is the reciprocal of the threshold contrast. We then sweep the frequency and repeat this exercise for each frequency. The result of one such analysis is shown in Figure 15, which is called the **Contrast Sensitivity Function** (CSF)². We can see that the RGC's CSF is a *bandpass* filter, where there is a preferred frequency to which a RGC responds the best. When the frequency is too high or too low, the signal is equivalent to a uniform background, to which a RGC would respond weakly.

Functionally, detecting contrast allows us to detect edges and contours: information across the two sides of an edge has the highest contrast. We will see shortly how later processing stages in the HVS leverage the contrasts to extract more specific information from the visual field to aid tasks such as object recognition.

²A more complete CSF would have to consider other attributes of the stimulus, such as color, temporal frequency, and eccentricity.

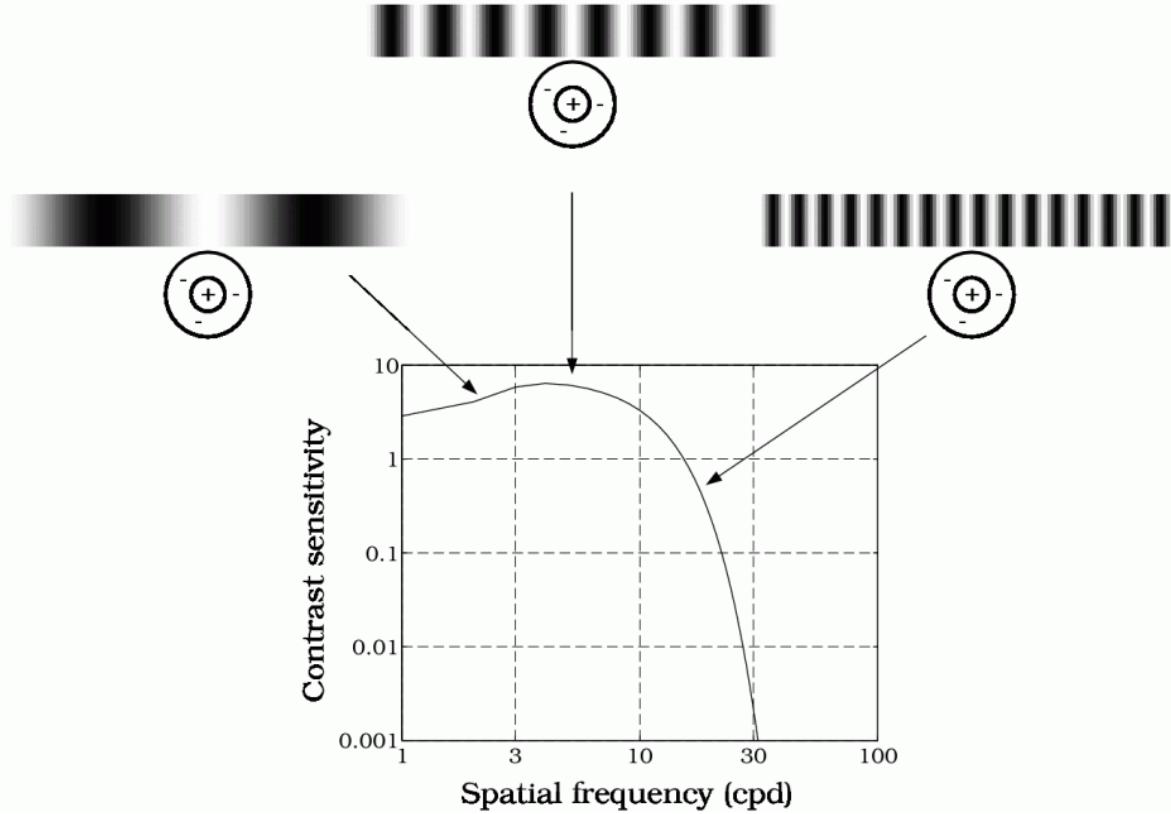


Figure 15: Contrast sensitivity function (CSF) under a center-surround RGC. CSF is usually bandpass. When the frequency is too high, the entire signal becomes almost a uniform background, to which a RGC responds weakly. From [Wandell \[1995, Fig. 5.18\]](#).

RGCs “Discount” Background Illuminations

Looking at Figure 14 again, the RGC responses do not change much with uniform illuminations (stimulus 1 and stimulus 3) no matter what the illumination level is. This is true for a wide range of illumination levels. In some sense, RGCs are able to “discount” the ambient light level so that the contrast is reliably encoded at arbitrary light levels. This is called **adaptation**.

We will return to adaptation later in the class, when we will more rigorously quantify adaptation and discuss the mechanisms behind adaptation, but for now, let’s just appreciate the significance of adaptation: being able to extract contrasts rather than absolute light levels is of significant advantage to us. The ambient level varies over several orders of magnitude, but the contrast of a scene is relatively stable regardless of the ambient light level. Consider our ape ancestors who need to find apples from a tree to survive. As the ambient light level increases, both the apple and the tree become brighter, but the contrast is relatively constant. To be able to reliably detect the apple, an ape needs to reliably extract contrast at all light levels but not

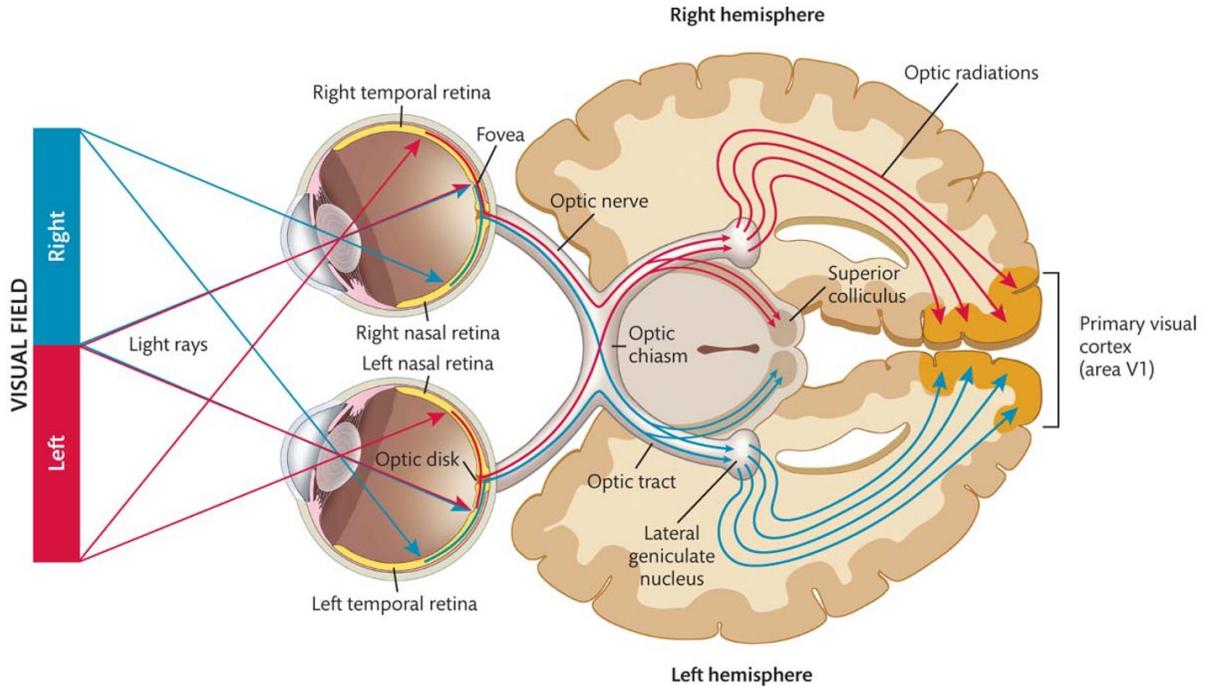


Figure 16: The main pathway from the eye to the brain. LGN collects information from one hemi-field and send it to the other half of the cortex. From [Yantis and Abrams \[2017\]](#), Fig. 3.2].

the absolute light level itself.

5 Post Retinal Processing

The signals leaving the retina are first routed to the **Lateral Geniculate Nucleus (LGN)** and then to the cortex, where vision is formed.

5.1 Lateral Geniculate Nucleus

Two the most important neurons in LGN are the magnocellular (M) cells and the parvocellular (P) cells. The magnocellular layers are connected to the midget RGCs and the parvocellular layers are connected to the parasol RGCs. Similar to RGCs, LGN neurons also have center-surround receptive fields, and their receptive field organizations are almost exact copies of that of the corresponding RGCs. This is why, by and large, LGN has been thought to be mainly a relay station, transmitting information from the retina to the brain. Interestingly, the way the LGN relays information to the brain is to gather information from one hemifield and send it to the other side of the cortex. This is achieved by wiring, as shown in Figure 16.

If LGN simply relays information, why does it exist at all? It turns out that LGN receives

about 90% of its inputs from the cortex [Sherman and Koch \[1986\]](#). This is different from retina, which is a “closed” system that doesn’t receive information from the rest of the brain. The feedback from the brain serves to regulate visual signals before they are sent to the brain. Higher-order brain regions encode cognitive information like attention, and one can imagine how attention can be used to influence what subsequent information is sent to the brain [\[O’Connor et al., 2002\]](#). If the brain were to send the feedback signals to the retina, the blind spot would have been 10 times larger, so LGN just seems like a convenient place where the feedback-driven regulation can take place.

Another Example of Parallel Pathways

Rods vs. cones is an example of parallel pathways in the HVS. The parvocellular vs. magnocellular pathway is another example; they encode different spatial/temporal frequency information. The magnocellular pathway responds to high temporal frequency well, is sensitive to low spatial frequency, and responds strongly to contrast changes. The parvocellular pathway, in large part, behaves oppositely. It is worth noting that these two visual streams start from the retina, where they start from distinct RGC cell types, and remain physically separated all the way into the primary visual cortex V1. This is different from the rod vs. cone pathway, which start at the photoreceptors and merge at the RGC layer.

5.2 Visual Cortex

Once in the cortex, the visual signals are first processed in the **primary visual cortex**, also known as visual area 1 (**V1**) or the **striate cortex**. V1 neurons primarily encode edge orientations but are also tuned to edge lengths, object motion direction, and specific colors. David Hubel and Torsten Wiesel, who won the Nobel Prize in 1981, were the first to elucidate the responses of V1 neurons and the architecture of V1 in general [\[Hubel and Wiesel, 1959, 1962, 1968\]](#).

V1 Simple Cells are Orientation Selective

Perhaps the most striking feature of V1 neurons is that they are orientation selective. The left panel of Figure 17 shows the responses of a cat V1 neuron, recorded by [Hubel and Wiesel \[1959\]](#), when presented with a slit of illumination at different orientations. This neuron responds best to a particular orientation (vertical in this case), and very weak or not at all to other orientations. The right panel in Figure 17 plots the neuron responses (spikes/second) as a function of the illumination orientation; a plot like this is called the neuron’s orientation **tuning curve**.

Why would this neuron be tuned to a specific orientation? The reason lies in its receptive field structure. Figure 18 shows the response of such a neuron when illuminated with spot lights at different locations. When the neuron is illuminated by spot lights across the vertical axis, it is inhibited, and it is excited when the spot lights are across the horizontal axis. The right panel shows the receptive field of such a neuron, where the center is inhibited and the flanking

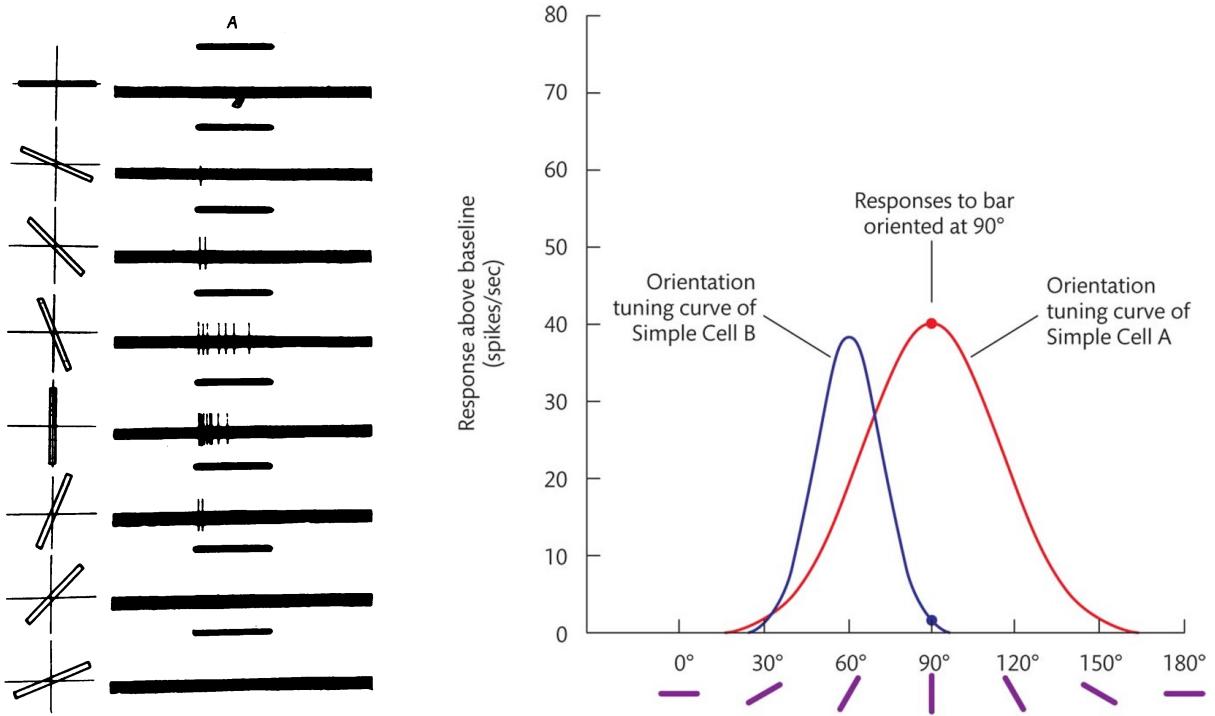


Figure 17: Left: orientation selectivity of a cat V1 simple cell; from [Hubel and Wiesel \[1959, Fig. 3\]](#); Right: orientation tuning curves of two illustrative V1 simple cells (do not necessarily correspond to the experimental data on the left); different cells can have different preferred orientations; from [Yantis and Abrams \[2017, Fig. 3.7\]](#).

areas are excitatory. There are other neurons where the excitatory and inhibitory regions are swapped.

This receptive field explains why a neuron could have an orientation selectivity: when the orientation of the stimulus coincides with the excitatory region of the receptive field the neuron is optimally stimulated³. Other orientations would involve both the excitatory and inhibitory regions, reducing or abolishing the response. V1 cells with such a receptive field are called **simple cells**. Different simple cells might have different preferred orientation; for instance, the second cell in the right panel of Figure 17 prefers a 60° orientation.

C–G in Figure 19 illustrate typical receptive fields found in V1 simple neurons. All are oriented (only one orientation is shown), but differ in arrangements. In comparison, A and B show the center-surround receptive fields found in RGCs and LGN neurons. Clearly, center-surround receptive fields simply cannot be orientation selective: try superimposing an edge and rotate it over the center-surround receptive field; will the response change much?

Why would a V1 simple neuron acquire such an oriented receptive field? This can be ex-

³Note that the receptive field in Figure 18 has an inhibitory central region and excitatory flanking areas, but the receptive field of the neuron in Figure 17 evidently has an opposite excitatory vs. inhibitory regions.

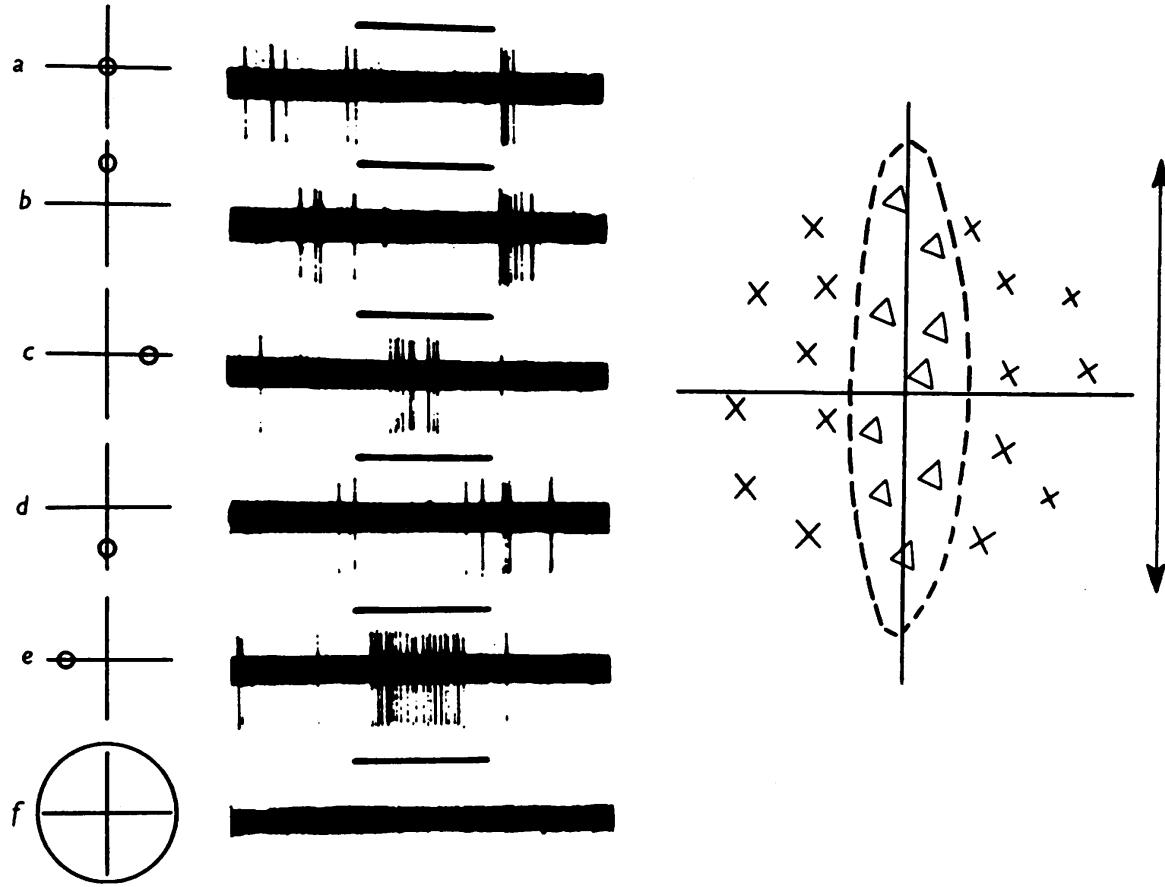


Figure 18: Responses of V1 simple cells to spot lights at different locations in the receptive field. △: inhibitory areas; ×: excitatory areas. *f* is when the entire field is illuminated uniformly. From [Hubel and Wiesel \[1959, Fig. 1\]](#).

plained by looking at how LGN neurons are connected to a V1 simple neuron. The top panel in Figure 19 illustrates the model suggested by Hubel and Wiesel, which is supported by recent electrophysiological results [[Clay Reid and Alonso, 1995](#)]. Each V1 simple cell synapses with and sums the inputs from multiple LGN neurons, whose receptive fields abut on the retina and are arranged in an oblique angle. When those receptive fields all have the same ON-center (or OFF-center) structure, the simple cell would tune for an oblique, elongated edge.

Direction, Length, and Binocular Vision Emerge from (Hyper)Complex Cells

The majority of neurons in V1 are actually not simple cells. Three-quarters of the V1 neuron population are what we call **complex cells**. Complex cells have more complex selectivities. Fundamentally, their receptive fields cannot be subdivided into excitatory and inhibitory areas.

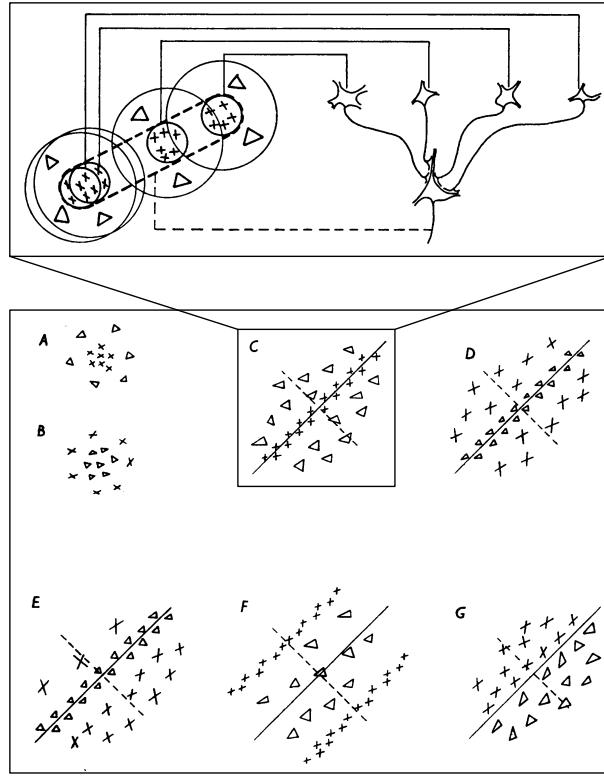


Figure 19: Bottom: typical receptive-field maps for V1 simple cells; while there are on and off regions, they are not organized in a center-surround fashion as do in RGCs/LGN. Top: this particular wiring of center-surround neurons in the LGN produce the receptive field in C at the bottom. \triangle : inhibitory areas; \times : excitatory areas. Adapted from [Hubel and Wiesel \[1962, Fig. 2, Fig. 19\]](#).

That is, they do not respond to a spot light no matter where the light is placed in the receptive field. Therefore, their responses to complicated geometries cannot be explained/predicted by their responses to spot lights, unlike those of simple cells.

Complex cells are also orientation selective, but unlike simple cells, many complex cells respond only to a properly oriented edge *sweeping* across the receptive field *as if* (but not actually) the entire receptive field is excitatory. However, they do not respond at all, or only weakly at the onset or turning off when we present a properly-oriented, stationary edge. This further shows that the responses of complex cells are not a linear superposition of responses to spot lights.

Interestingly, about one-fifth of complex cells prefer movement of a particular direction. [Hubel and Wiesel \[1968\]](#) measured the direction selectivity of V1 complex cells in monkeys, and Figure 20 shows an example, where a complex cell is optimally stimulated by a properly oriented edge moving across a particular direction, but not the opposing, orthogonal direction.

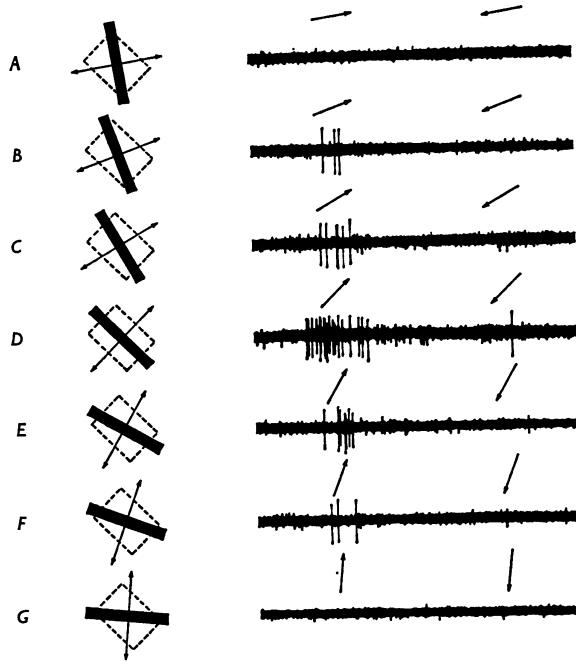


Figure 20: Some V1 complex neurons prefer properly oriented edges sweeping across its receptive field; these neurons also have direction selectivity — even under the same orientation. From [Hubel and Wiesel \[1968, Fig. 2\]](#).

This shows the **direction selectivity** of many complex cells.

[Hubel and Wiesel \[1968\]](#) also discovered a set of what they call the **end-stopping** neurons or hypercomplex cells in V1. Those neurons are tuned to properly oriented edges with a specific length, beyond which the neurons are inhibited. These neurons play a role in encoding corners, curvatures, and sudden breaks in lines [[Hubel, 1995, p. 85](#)].

Finally, Hubel and Wiesel also found that some V1 neurons respond to stimuli only from the left eye or only from the right eye, a properly termed **ocular dominance**. There are also binocular cells that can be stimulated independently by stimulus from either eye. These cells represent the first stage where information from the left and right hemi-fields converge, which is critical for depth perception.

“Be More Specific”

An obvious conclusion we can get from comparing V1 neurons and retina/LGN neurons is this: as we progress along the visual pathway, the stimulus we present to the visual system must be more specific. Put another way, our visual system increasingly extracts more specific information as signals progress in the pathway.

Being more specific is critical, as that allows us to recognize objects by their subtle details.

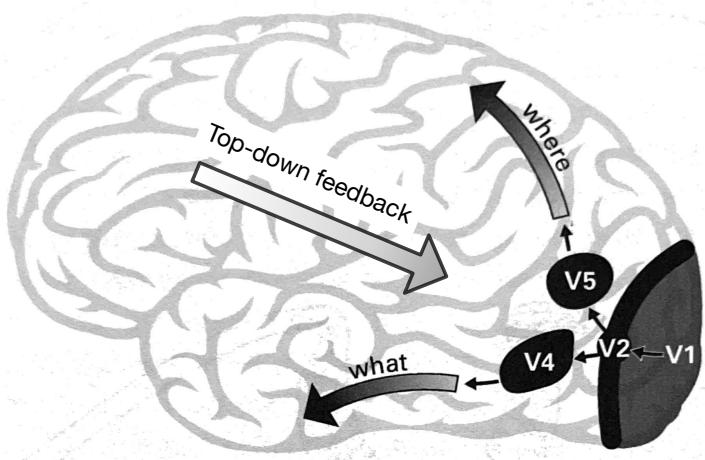


Figure 21: Once in the cortex, signals are projected from area V1 to other areas, each generally specialized in a particular information process. There are top-down feedbacks in the cortex from higher-order areas to lower-order areas. Adapted from [Dowling and Dowling Jr \[2016, Fig. 1.3\]](#).

For instance, RGCs/LGN neurons provide the contrast/edge detection capability, but virtually any object has contrasts and edges, so they are not terribly useful in recognizing specific objects. V1 simple neurons, however, allow us to detect orientations, and that is critical to our vision — because from orientations we can then infer shapes, as we recognize objects mostly by their shapes.

Critically, however, V1 simple neurons offer orientation selectivity *precisely because* RGCs/LGN neurons have contrast/edge detection capabilities, as demonstrated in Figure 19. This is why we say early visual system extracts low-level information but later visual system extracts high-level information: the former is used as the building blocks by the latter.

The Rest of the Cortex

From V1, signals are projected to other areas such as V2, V4, IT, MT, etc. There are two main projection pathways [[Nassi and Callaway, 2009](#), [Ungerleider and Mishkin, 1982](#), [Mishkin et al., 1983](#)], as shown in Figure 21. The first is the **dorsal** pathway, which is concerned with observing objects in space, such as its spatial location and motion, information that is also useful to guide actions [[Goodale et al., 1991](#)]. So this pathway is also called the “where/how” pathway. The other is the **ventral** pathway, or the “what” pathway, that carries information of the details and identify of objects and supports visual functions such as object recognition, face recognition, color perception. The two pathways interact. For instance, to guide visual action we not only need to know the position and motion of the objects but also the shape, color, etc.

The discussion so far focuses on the bottom-up information flow, the flow of information from lower-order representations in the hierarchy such as V1 to higher-order representations such as V4 and beyond. There is also the top-down information flow from the higher regions

to the lower regions. This information flow provides feedback information such as attention, knowledge, expectation, etc., to influence the early information processing in the cortex [Gilbert and Li, 2013]. Combining the bottom-up and the top-down flows, the HVS acts essentially as a self-adaptive system that automatically optimizes its performance for a given task.

6 Summary and Outlook

Eye optics mainly focus lights on the retina to form an optical image. Photoreceptors on the retinal convert optical signals to electrical signals. The rest of the retina mainly extract contrast information through the center-surround receptive field of the RGCs. The LGN mainly collects information from one hemi-field together and sends it to the other side of V1. V1 extracts more specific information such as orientation, motion direction, edge length, and higher order areas provide more specialized functions.

The rest of our discussion will be very much retina focused, both because that is what we know the most about in the HVS and because a lot of the visual functions that we care about can be largely explained by retinal behaviors.

References

- Juan M Angueyra-Aristizábal. *The Limits Imposed in Primate Vision by Transduction in Cone Photoreceptors*. PhD thesis, University of Washington Libraries, 2014.
- HB Barlow, Roo Fitzhugh, and SW Kuffler. Change of organization in the receptive fields of the cat's retina during dark adaptation. *The Journal of physiology*, 137(3):338, 1957.
- Horace B Barlow. Summation and inhibition in the frog's retina. *The Journal of physiology*, 119(1):69, 1953.
- DA Baylor, BJ Nunn, and JL Schnapf. Spectral sensitivity of cones of the monkey macaca fascicularis. *The Journal of Physiology*, 390(1):145–160, 1987.
- David M Berson, Felice A Dunn, and Motoharu Takao. Phototransduction by retinal ganglion cells that set the circadian clock. *Science*, 295(5557):1070–1073, 2002.
- Edward A Boettner and J Reimer Wolter. Transmission of the ocular media. *Investigative ophthalmology & visual science*, 1(6):776–783, 1962.
- Caerbannog. Comparison of structures in vertebrate's eye (left) with octopus' eye (right); CC BY-SA 3.0 license. [https://en.wikipedia.org/wiki/Blind_spot_\(vision\)#/media/File:Evolution_eye_2.svg](https://en.wikipedia.org/wiki/Blind_spot_(vision)#/media/File:Evolution_eye_2.svg), 2016.
- R Clay Reid and Jose-Manuel Alonso. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, 378(6554):281–284, 1995.

- Christine A Curcio, Kenneth R Sloan, Robert E Kalina, and Anita E Hendrickson. Human photoreceptor topography. *Journal of comparative neurology*, 292(4):497–523, 1990.
- Dennis M Dacey and Michael R Petersen. Dendritic field size and morphology of midget and parasol ganglion cells of the human retina. *Proceedings of the National Academy of sciences*, 89(20):9666–9670, 1992.
- Herbert JA Dartnall, James K Bowmaker, and John Dixon Mollon. Human visual pigments: microspectrophotometric results from the eyes of seven persons. *Proceedings of the Royal society of London. Series B. Biological sciences*, 220(1218):115–130, 1983.
- Michael Tri Hoang Do and KW Yau. Intrinsically photosensitive retinal ganglion cells. *Physiological reviews*, 2010.
- John E Dowling and Joseph L Dowling Jr. *Vision: how it works and what can go wrong*. MIT Press, 2016.
- Charles D Gilbert and Wu Li. Top-down influences on visual processing. *Nature reviews neuroscience*, 14(5):350–363, 2013.
- Andrew S Glassner. *Principles of digital image synthesis*. Elsevier, 2014.
- Melvyn A Goodale, A David Milner, Lorna S Jakobson, and David P Carey. A neurological dissociation between perceiving objects and grasping them. *Nature*, 349(6305):154–156, 1991.
- H Keffer Hartline. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology-Legacy Content*, 121(2):400–415, 1938.
- H Keffer Hartline. Excitation and inhibition of the “off” response in vertebrate optic nerve fibers. *Am. J. Physiol*, 126:527, 1939.
- H Keffer Hartline. The effects of spatial summation in the retina on the excitation of the fibers of the optic nerve. *American Journal of Physiology-Legacy Content*, 130(4):700–711, 1940a.
- H Keffer Hartline. The receptive fields of optic nerve fibers. *American Journal of Physiology-Legacy Content*, 130(4):690–699, 1940b.
- H Keffer Hartline. Inhibition of activity of visual receptors by illuminating nearby retinal areas in the limulus eye. *Federation Proceedings*, 8(1):69, 1949.
- H Keffer Hartline and Clarence Henry Graham. Nerve impulses from single receptors in the eye. *Journal of Cellular & Comparative Physiology*, 1932.
- H Keffer Hartline, Henry G Wagner, and Floyd Ratliff. Inhibition in the eye of limulus. *The Journal of general physiology*, 39(5):651–673, 1956.
- Samer Hattar, H-W Liao, Motoharu Takao, David M Berson, and KW Yau. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science*, 295(5557):1065–1070, 2002.

- Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500, 1952.
- David H Hubel. *Eye, brain, and vision*. Scientific American Library/Scientific American Books, 1995.
- David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *J physiol*, 148(3):574–591, 1959.
- David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106, 1962.
- David H Hubel and Torsten N Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243, 1968.
- Saad Idrees, Michael B Manookin, Fred Rieke, Greg D Field, and Joel Zylberberg. Biophysical neural adaptation mechanisms enable artificial neural networks to capture dynamic retinal computation. *Nature Communications*, 15(1):5957, 2024.
- Gerald H Jacobs. Primate color vision: a comparative perspective. *Visual neuroscience*, 25(5-6): 619–633, 2008.
- Stephen W Kuffler. Neurons in the retina: organization, inhibition and excitation problems. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume 17, pages 281–292. Cold Spring Harbor Laboratory Press, 1952.
- Stephen W Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of neurophysiology*, 16(1):37–68, 1953.
- Steven M LaValle. *Virtual reality*. Cambridge university press, 2023.
- Lorenzo Lazzerini Ospri, Glen Prusky, and Samer Hattar. Mood, the circadian system, and melanopsin retinal ganglion cells. *Annual review of neuroscience*, 40(1):539–556, 2017.
- Fuyou Liao, Zheng Zhou, Beom Jin Kim, Jiewei Chen, Jingli Wang, Tianqing Wan, Yue Zhou, Anh Tuan Hoang, Cong Wang, Jinfeng Kang, et al. Bioinspired in-sensor visual adaptation for accurate perception. *Nature Electronics*, 5(2):84–91, 2022.
- Liqun Luo. *Principles of neurobiology*. Garland Science, 2016.
- Mortimer Mishkin, Leslie G Ungerleider, and Kathleen A Macko. Object vision and spatial vision: two cortical pathways. *Trends in neurosciences*, 6:414–417, 1983.
- Jonathan J Nassi and Edward M Callaway. Parallel processing strategies of the primate visual system. *Nature reviews neuroscience*, 10(5):360–372, 2009.

- Daniel H O'Connor, Miki M Fukui, Mark A Pinsk, and Sabine Kastner. Attention modulates responses in the human lateral geniculate nucleus. *Nature neuroscience*, 5(11):1203–1209, 2002.
- Dale Purves, George J. Augustine, David Fitzpatrick, William Hall, Anthony-Samuel LaMantia, Richard D. Mooney, Michael L. Platt, and Leonard E. White. *Neurosciences*. Oxford University Press, 6 edition, 2017.
- SM Sherman and Christof Koch. The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus. *Experimental Brain Research*, 63:1–20, 1986.
- Leslie G Ungerleider and Mortimer Mishkin. Two cortical visual systems. In David J Ingle, Melvyn A Goodale, Richard JW Mansfield, et al., editors, *Analysis of visual behavior*, pages 549–586. Mit Press Cambridge, MA, 1982.
- George Wald. Molecular basis of visual excitation. *Science*, 162(3850):230–239, 1968.
- Brian A. Wandell. *Foundations of vision*. Sinauer Associates, 1995.
- Robert Wodnicki, Gordon W Roberts, and Martin D Levine. A foveated image sensor in standard cmos technology. In *Proceedings of the IEEE 1995 Custom Integrated Circuits Conference*, pages 357–360. IEEE, 1995.
- Steven Yantis and Richard A Abrams. *Sensation and Perception*. Worth Publishers, 2 edition, 2017.