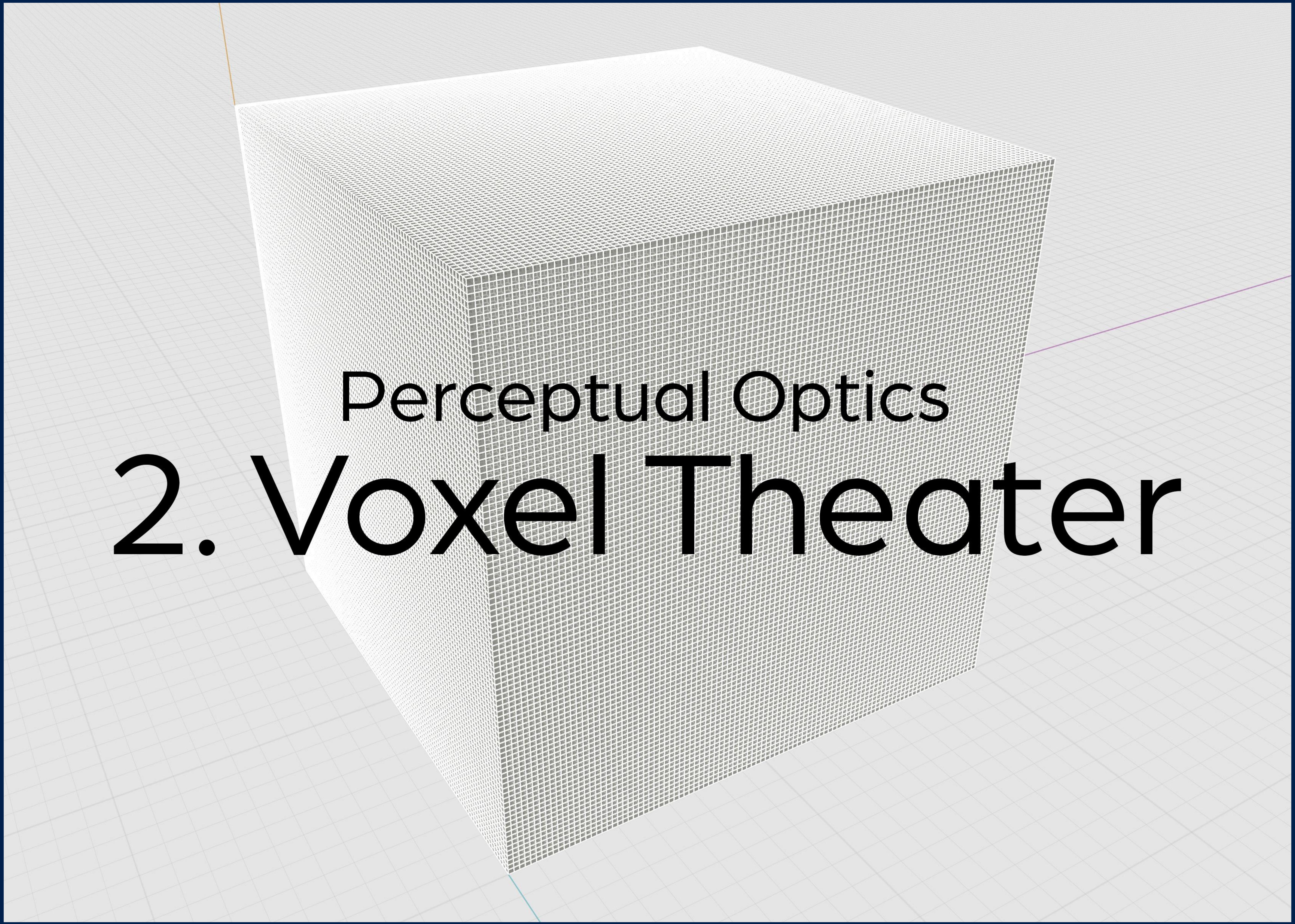
Perceptual Optics

2. Voxel Theater



Perceptual Optics

2. Voxel Theater

2. Voxel Theater

Objectives:

- Definitions. What are pixels and voxels?
- Visual Paths through Brain. Completion of physical optics: physical paths/processing in the brain (*RGC → LGN → V1 → what/where paths → PFC → premotor → motor cortex* [response to visual stimulus]; *pretectal path; SC path*).
- Requirements from Experience. Start from perception, and work downwards, considering observed phenomena that will require some substrate/mechanism (a 3D voxel theater appears "attached" to the opening of each eye; left and right hemifield are much more merged than left and right eyeballs; how many voxels can we discriminate?).
- Stereoscopic Vision. Combining left and right eye's views.
- Computing and Representing Depth; Vision as Point vs. Large Field.

What Is a Pixel?

A pixel (short for "picture element") is the smallest unit of an image that can be displayed or represented on a screen. It is a tiny square that contains a specific color and brightness value. When many pixels are arranged together in a grid, they form an image. The resolution of an image is determined by the number of pixels in the grid. The more pixels there are, the higher the resolution and the more detail the image can display. Pixels are used in various display technologies, such as computer monitors, televisions, digital cameras, and mobile devices.

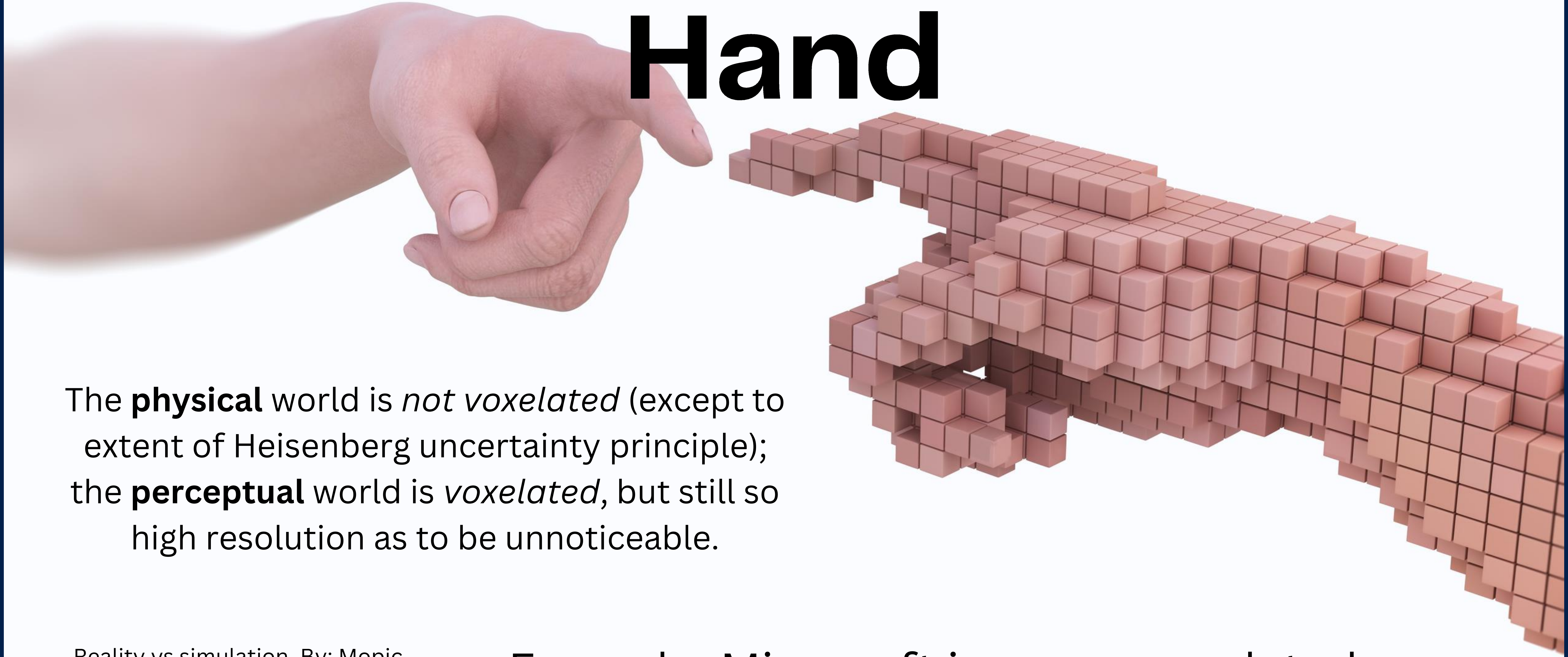
Your smart phone has some resolution of **pixels** (smallest increments of width and height)

What Is a Voxel?!

A voxel (short for "volume element") is the 3D equivalent of a pixel. It represents a tiny cube of space in a three-dimensional volume or grid. Like pixels, voxels are used to represent visual information, but they are used in 3D graphics and imaging applications to represent 3D objects and their properties, such as color, transparency, and texture. The resolution of a 3D model is determined by the number of voxels in the grid, similar to how the resolution of a 2D image is determined by the number of pixels in its grid. Voxels are commonly used in medical imaging, computer-aided design, and video game development.

In Star Wars, R2D2's hologram message from Leia had some resolution of **voxels** (smallest increments of width, height, and depth)

A "Voxelated" Hand



The **physical** world is *not voxelated* (except to extent of Heisenberg uncertainty principle); the **perceptual** world is *voxelated*, but still so high resolution as to be unnoticeable.

Reality vs simulation. By: Mopic. https://Dreamstime.com. Paid use.

Example: Minecraft is a very voxelated game

Physical Pixels & Voxels

Pixels

Images of the *physical world* can be divided into pixels. In perspective images captured by cameras or seen by eyes, pixels represent the angles of incoming light and can be utilized to reconstruct a 3D scene. On the other hand, in orthographic drawings, pixels represent the spatial relations of objects along a particular plane.

Voxels

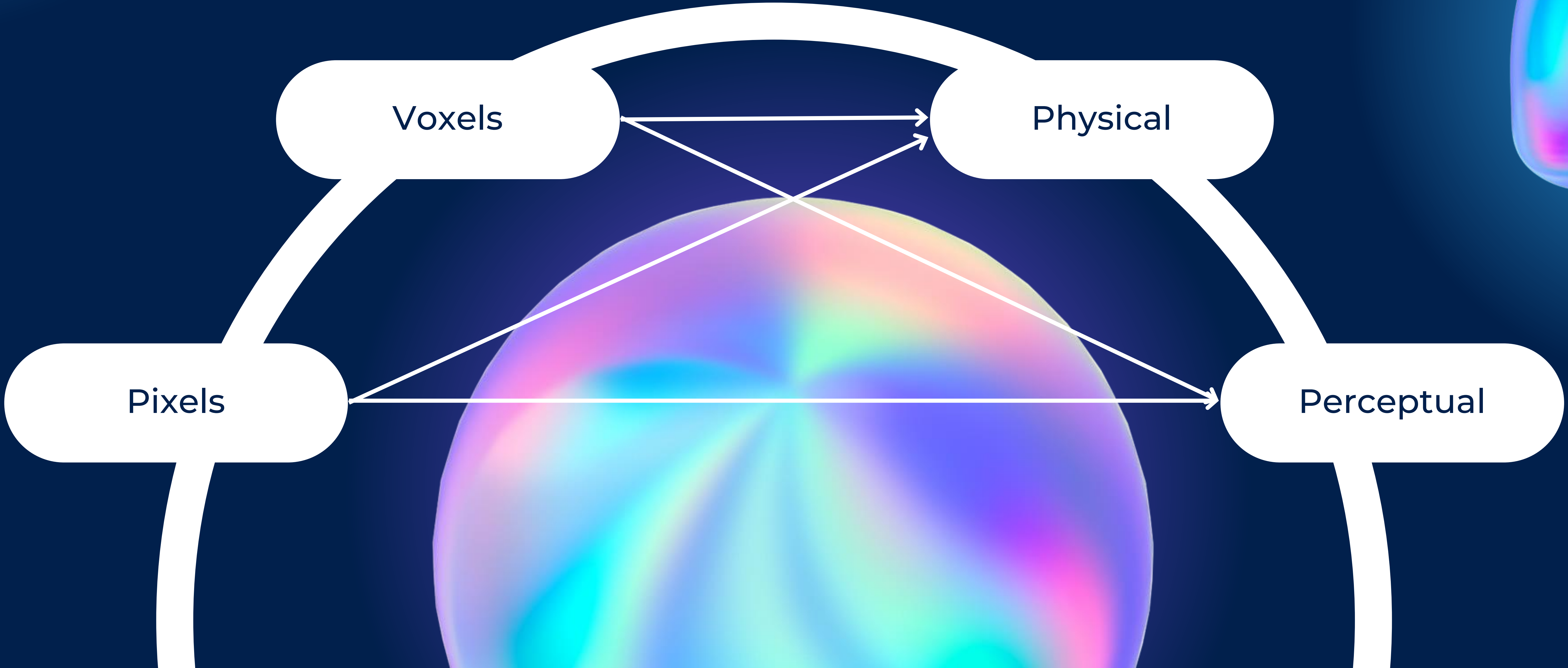
The *physical world* could be divvied up into any desired resolution of voxels for mapping purposes. A one-meter cube could be subdivided into a million (100 units w, h, d) or ten trillion (10,000 units w, h, d) voxels.

Perceptual Pixels & Voxels

The perceptual world is a term used to describe the brain's representation of the physical world. It includes a 3D reconstruction or model of the surrounding environment, which is drawn projecting onto the avatar's eyes in a 2D perspective image. If the depth is inaccurately perceived, it can result in a misplacement of objects within the 3D model.

Recap

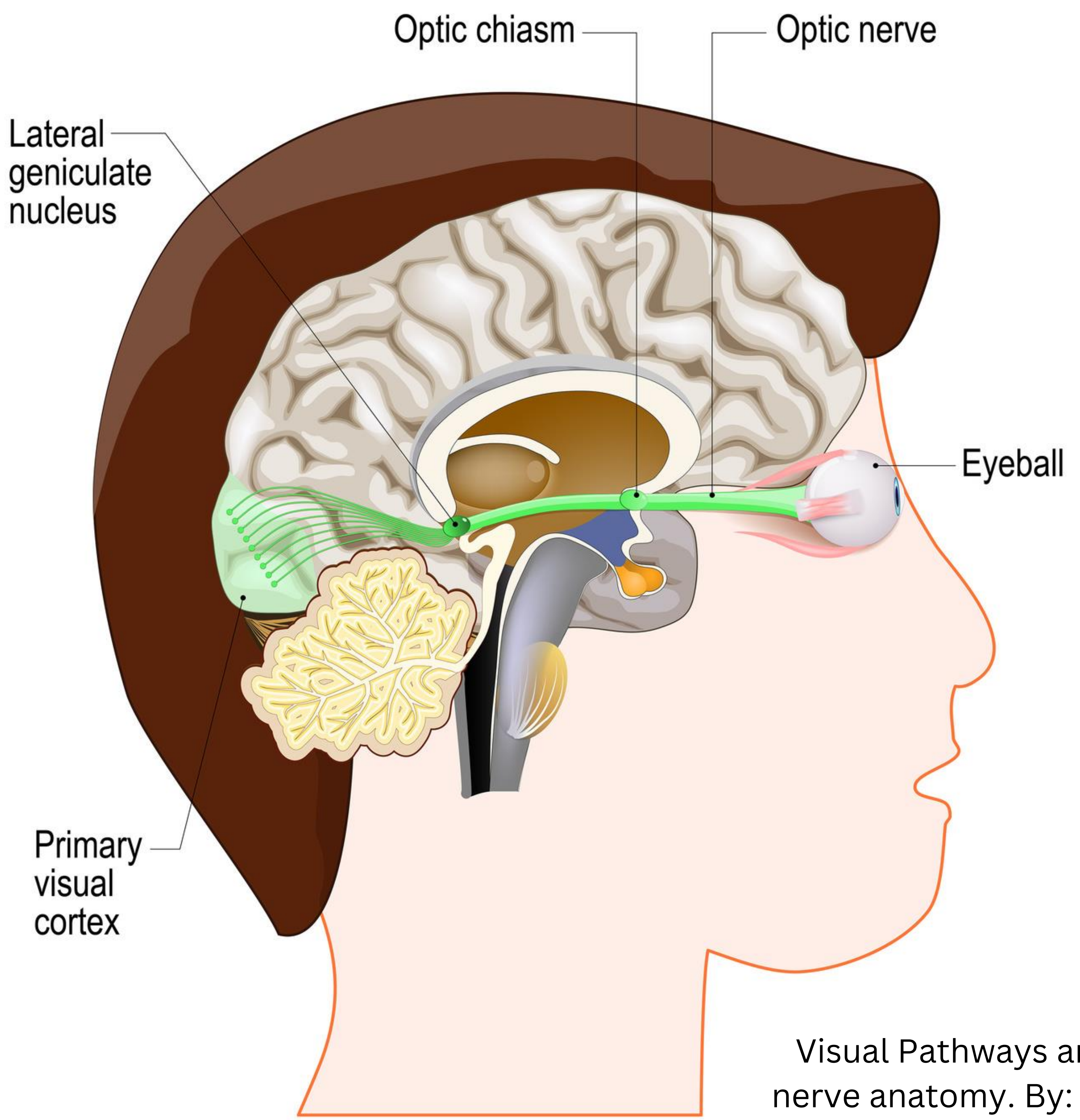
We can (for utility) divide the physical world into pixels and voxels, and often do (for use in technology). The perceptual world can only display a certain level of resolution. As such, we can explore its limits.



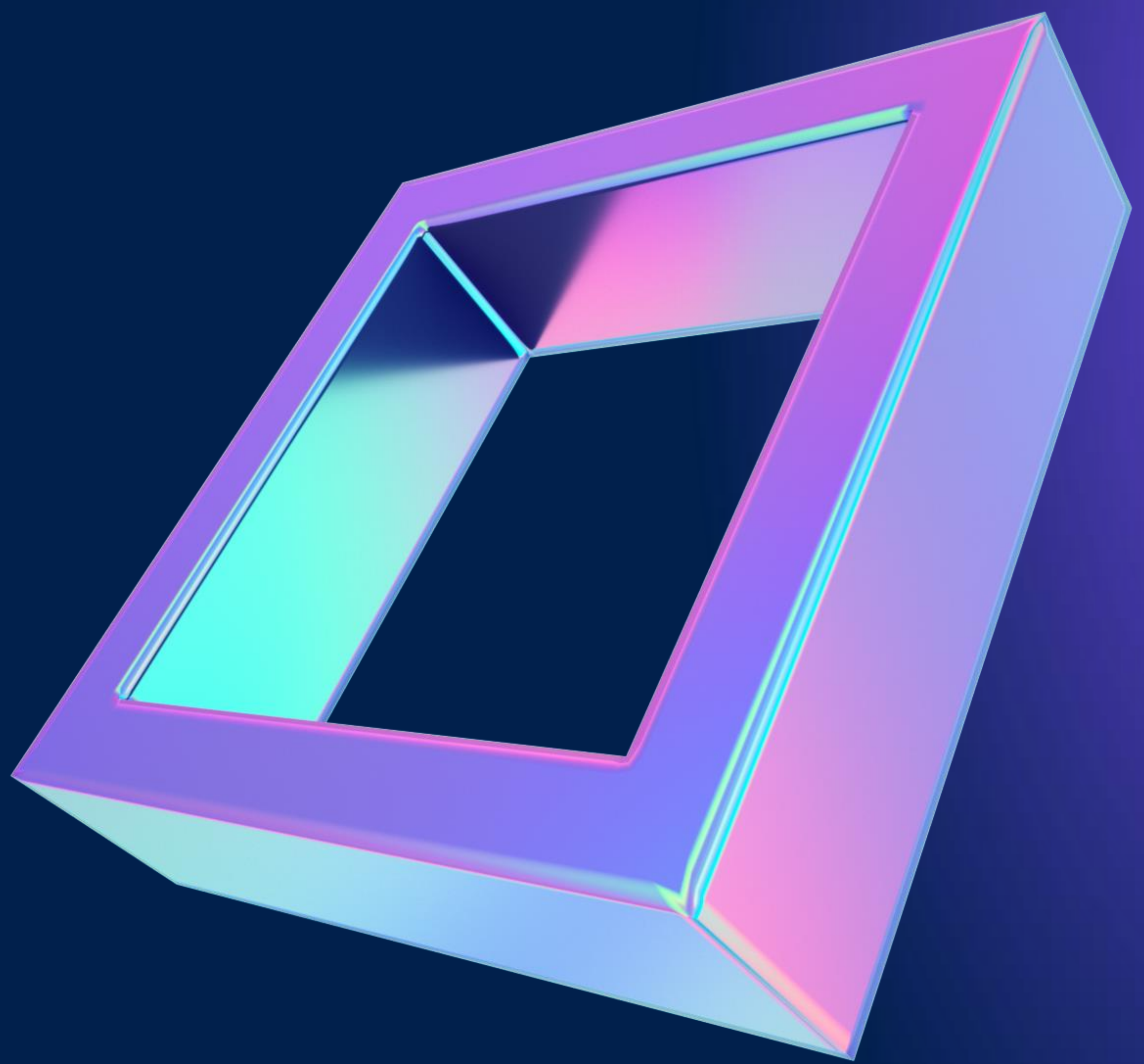
Visual Paths

Visual pathways

Cortical path (displayed below), pretectal path (accommodation, pupil size), superior colliculus path (saccade direction/amount), suprachiasmatic nucleus path (circadian rhythms).



Visual Pathways and Optic nerve anatomy. By: Designua. https://Dreamstime.com. Paid use.



1. Primary Visual Path: Conscious Sight Perception

1. The primary visual pathway is responsible for processing and interpreting visual information, ultimately leading to conscious sight and perception of color, shape, location, and motion.
2. Visual information from the retina is transmitted through the optic nerve to the lateral geniculate nucleus (LGN) in the thalamus.
3. The LGN acts as a relay station, processing and organizing visual information before sending it to the primary visual cortex (V1) located in the occipital lobe.
4. The primary visual cortex processes basic visual features, and then sends the information to higher-order visual areas for more complex processing and integration.
5. This intricate neural network allows us to perceive, interpret, and react to the visual world around us, resulting in the conscious experience of sight.

2. Pretectal Path:
Lens Accommodation (Focus); Pupil Constriction

1. The pretectal pathway is responsible for regulating lens accommodation (focusing) and pupil constriction in response to changes in light intensity and viewing distance.
2. Visual information from the retina is transmitted to the pretectal nucleus in the midbrain, which processes light intensity and depth information.
3. For lens accommodation, the pretectal nucleus sends signals to the Edinger-Westphal nucleus, which in turn stimulates the ciliary muscles to adjust the shape of the lens for optimal focus.
4. For pupil constriction, the pretectal nucleus communicates with the Edinger-Westphal nucleus to control the sphincter pupillae muscles, resulting in pupillary constriction under bright light conditions.
5. This pathway ensures that we can maintain clear vision and protect our eyes from excessive light exposure by dynamically adjusting the lens and pupil size.

3. Superior Colliculus Pathway: Saccadic Eye Movements and Head Direction

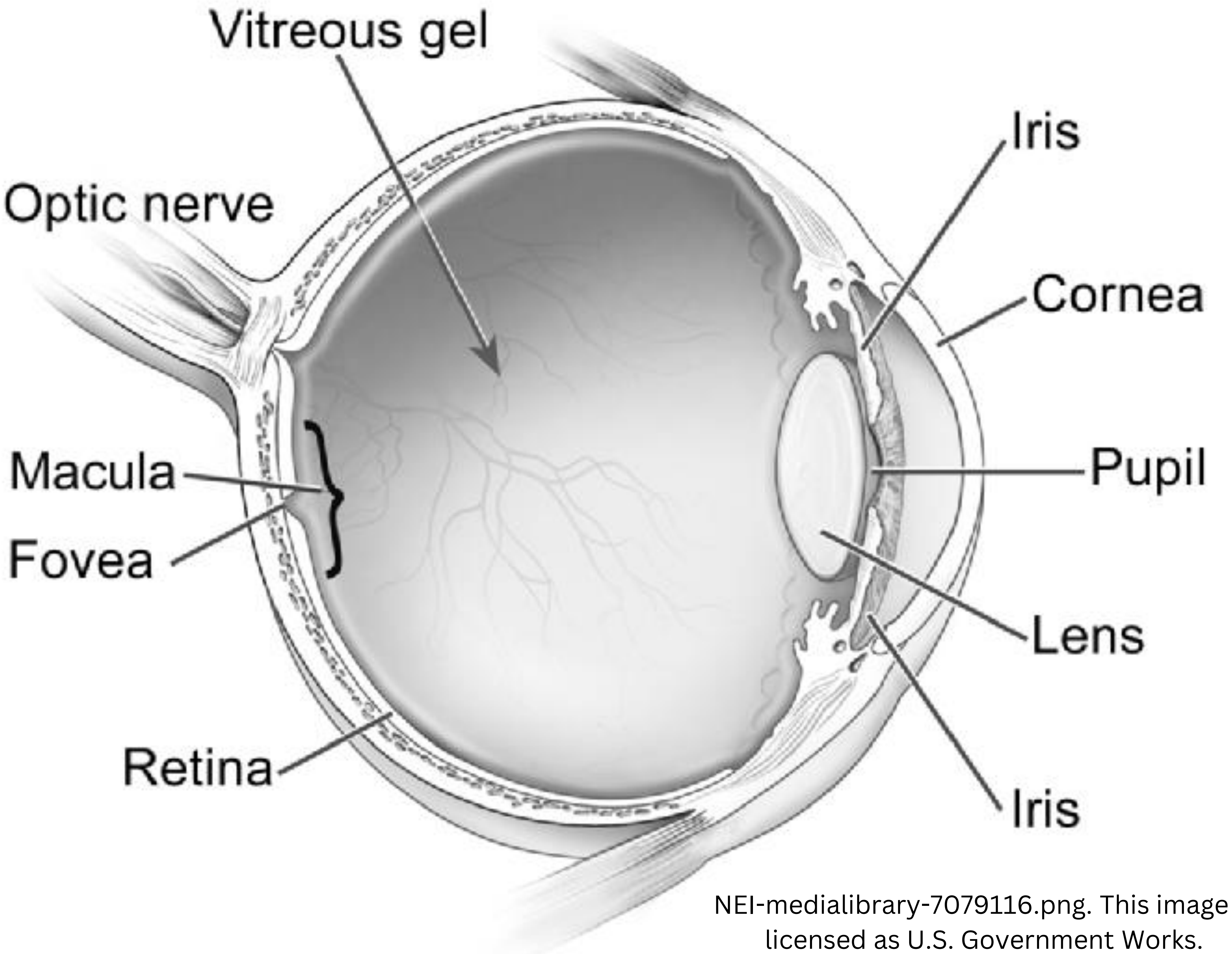
1. The superior colliculus pathway coordinates saccadic eye movements and head direction adjustments to quickly change our focus between different points in the visual field.
2. The retina captures visual information and sends signals to the superior colliculus in the midbrain, which plays a crucial role in coordinating these movements.
3. The superior colliculus integrates visual, auditory, and somatosensory information to determine the target, direction, and magnitude of saccadic eye movements and head adjustments.
4. It then sends signals to other brainstem nuclei, such as the paramedian pontine reticular formation (PPRF) and the rostral interstitial nucleus of the medial longitudinal fasciculus (riMLF), as well as the cerebellum to control eye and neck muscles.
5. This pathway ensures rapid, precise adjustments in eye position and head orientation, enabling us to efficiently process and react to the visual environment.

4. Suprachiasmatic Nucleus Path (Circadian Rhythms)

The pathway from the eye to the brain region that regulates sleep at lower light levels involves specialized photosensitive retinal ganglion cells (pRGCs) containing the photopigment melanopsin. These cells are responsible for non-image forming visual functions, such as circadian rhythm regulation and the pupillary light reflex.

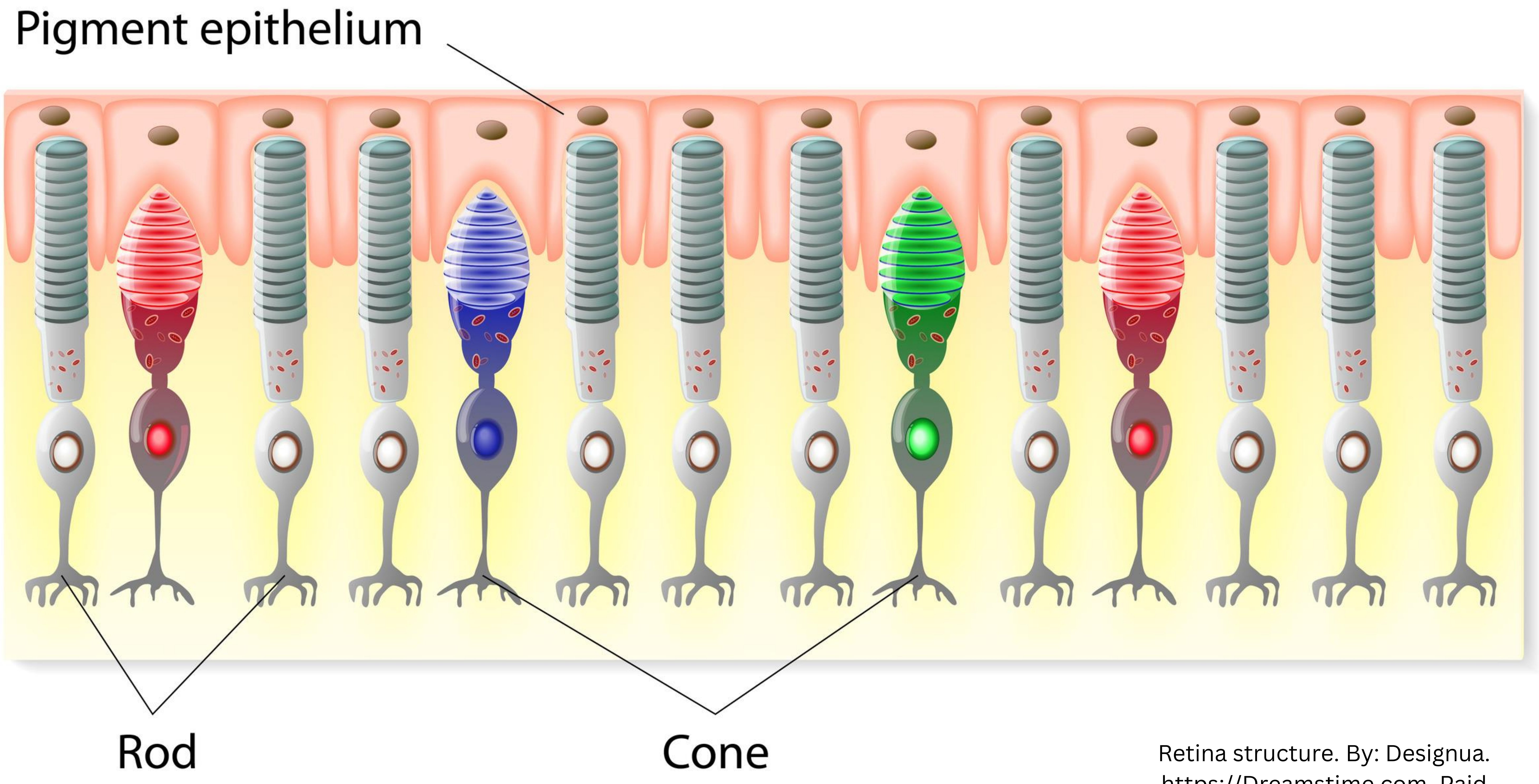
1. Light enters the eye and is detected by the melanopsin-containing pRGCs in the retina.
2. The pRGCs transmit light information via their axons in the retinohypothalamic tract (RHT).
3. The RHT projects to the suprachiasmatic nucleus (SCN) in the hypothalamus, which is the primary circadian pacemaker in mammals.
4. The SCN processes the light information and sends signals to other regions of the brain, such as the pineal gland.
5. The pineal gland, in response to these signals, regulates the production and release of the hormone melatonin, which plays a critical role in sleep regulation.

At lower light levels, the melanopsin-containing pRGCs are less active, leading to reduced signaling to the SCN. As a result, melatonin production increases, promoting sleepiness and helping to regulate the sleep-wake cycle.



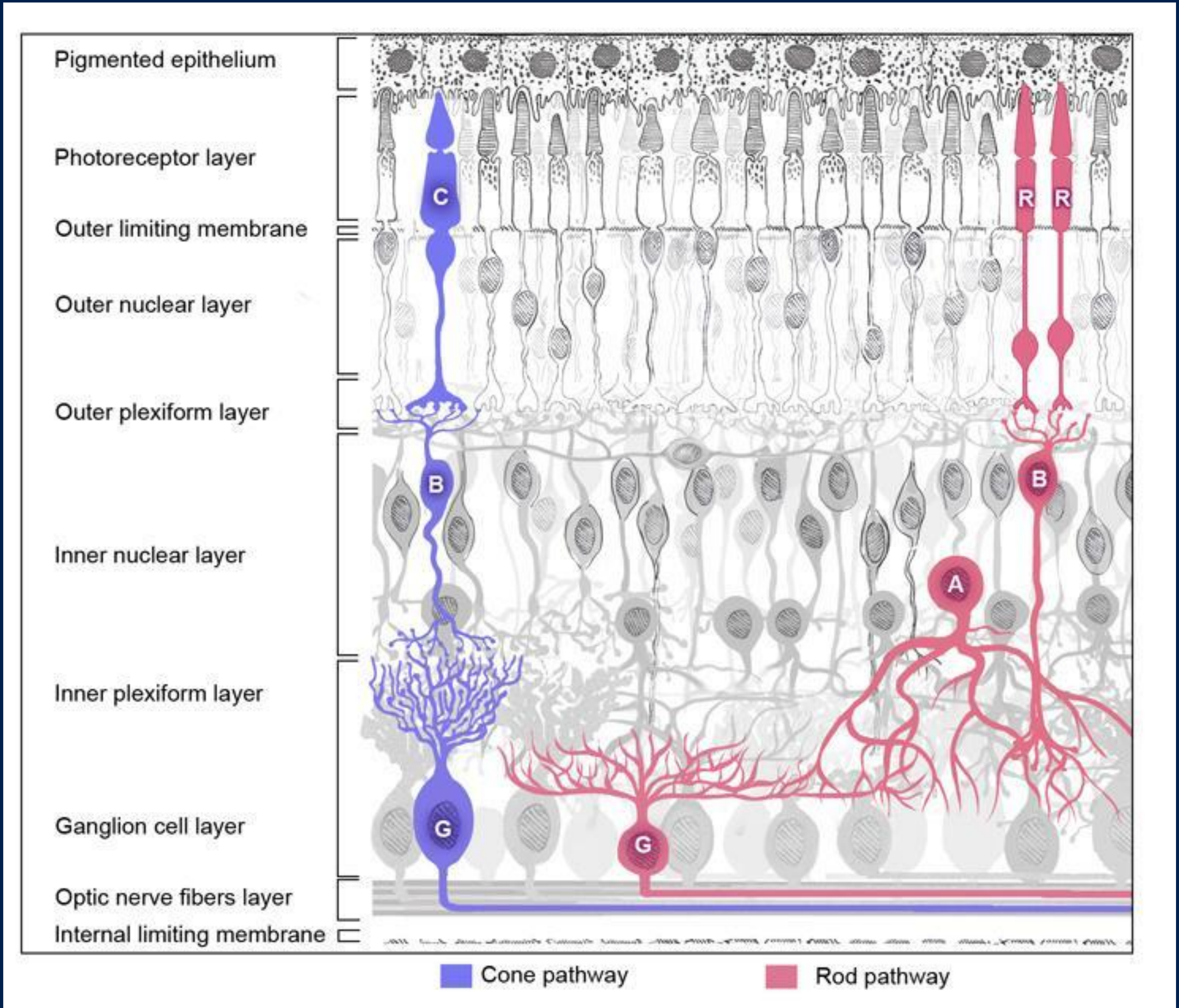
NEI-medialibrary-7079116.png. This image is licensed as U.S. Government Works. https://medialibrary.nei.nih.gov/search?keywords=eye%20anatomy#/media/3544

STRUCTURE OF THE RETINA



Retina structure. By: Designua. https://Dreamstime.com. Paid use.

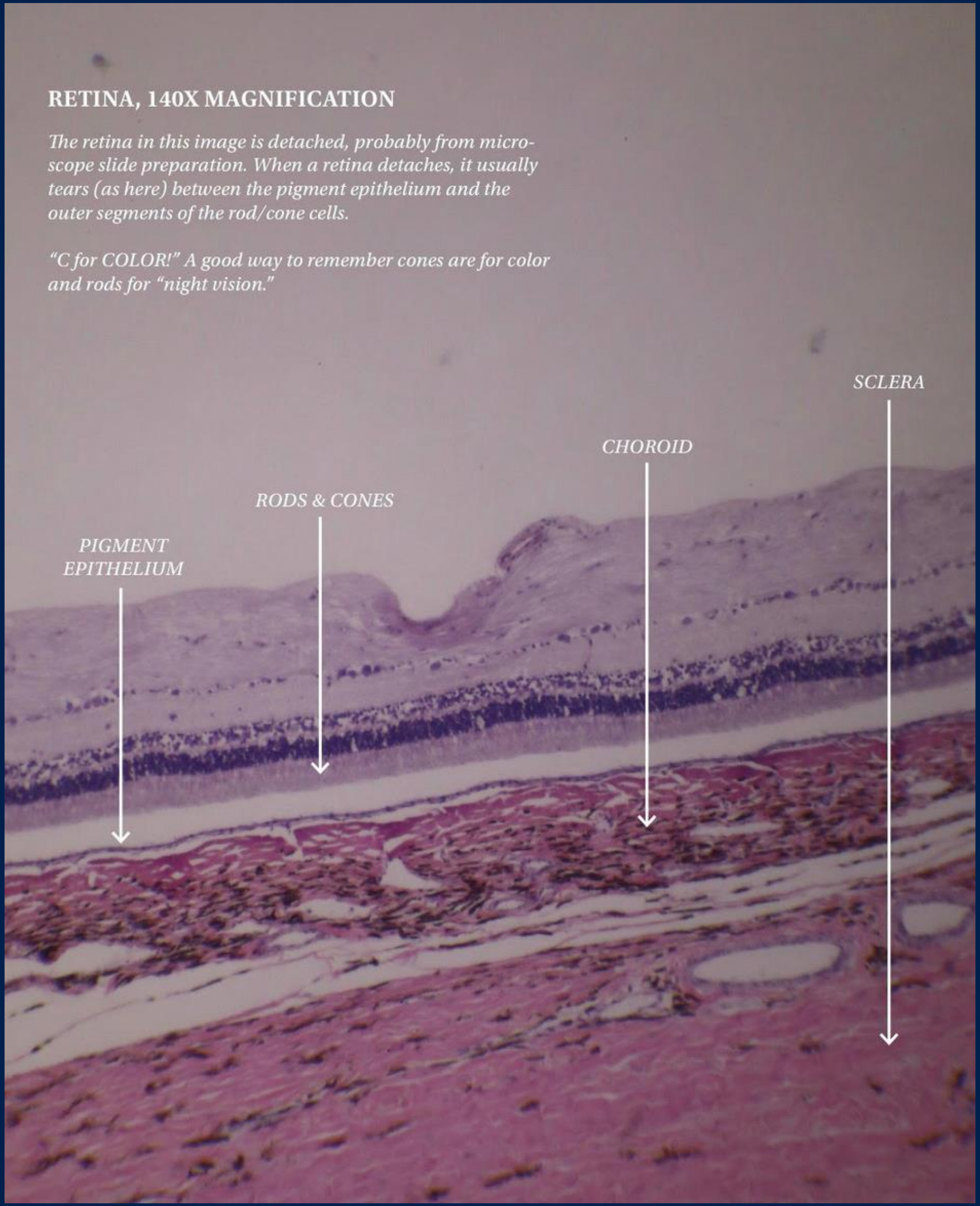
First Stages of Visual Processing

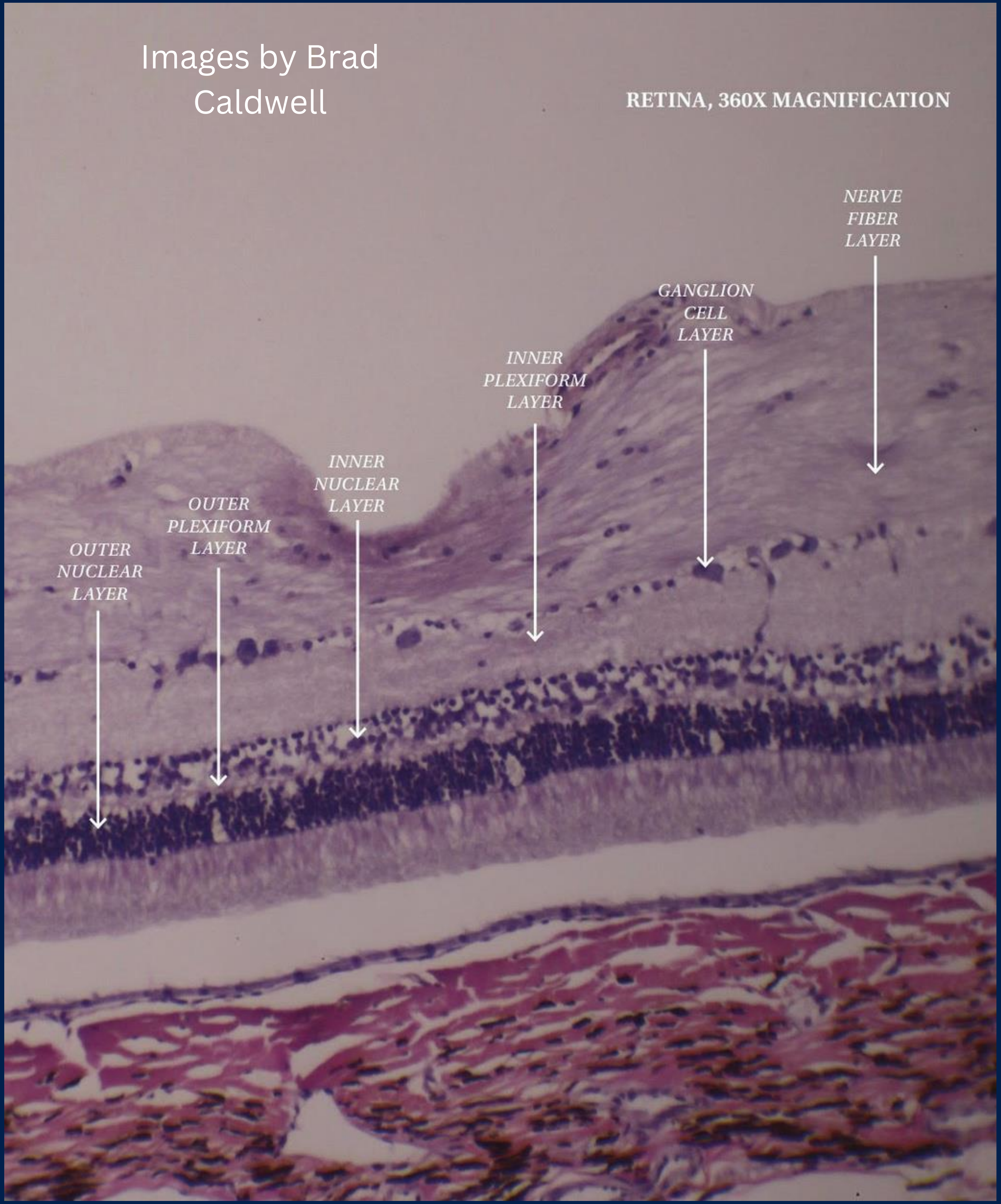


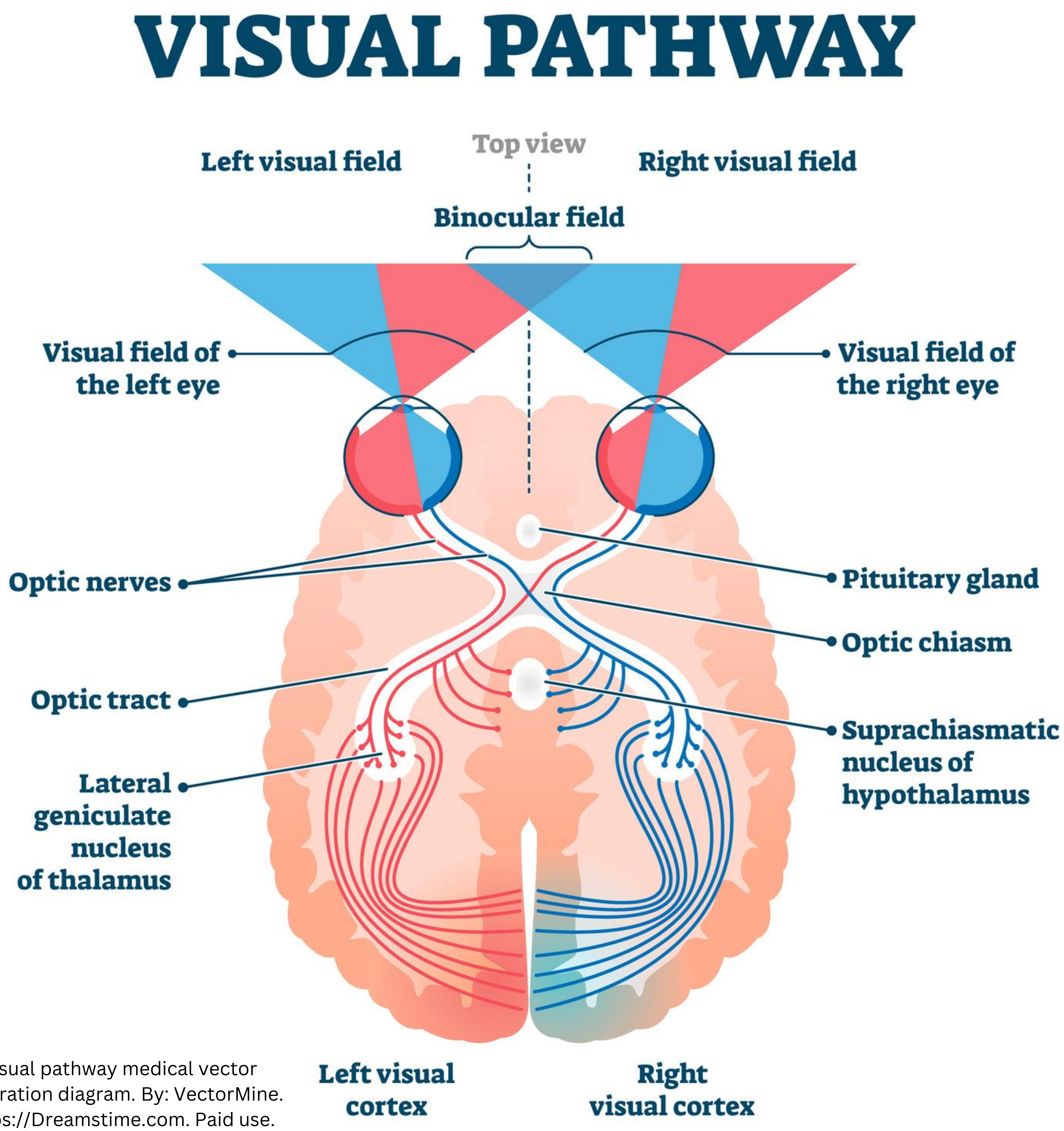
The process of visual processing begins when light enters the eye and is focused onto the retina by the cornea and lens. The rods and cones in the retina then respond to the light by producing electrical signals, which are transmitted to the other cells in the retina.

The next stage of processing involves bipolar cells, which get input from the rods and cones and pass this information on to the ganglion cells. The ganglion cells then send their output through the optic nerve, which carries the electrical signals to the brain for processing.

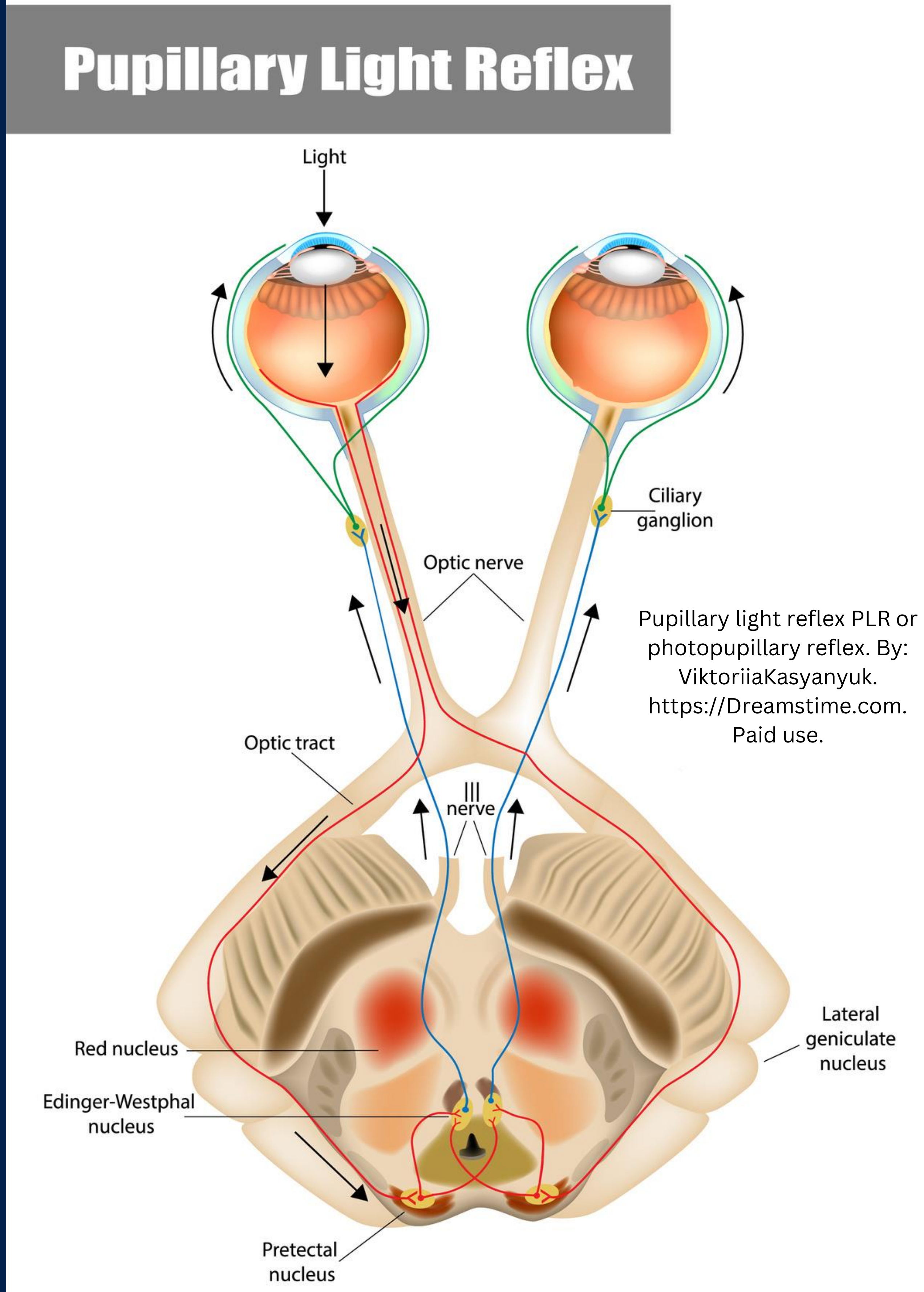
In addition, the retina also contains several other types of cells that play important roles in visual processing, including horizontal cells and amacrine cells. These cells help to refine and modulate the electrical signals generated by the photoreceptor cells, and enhance the ability of the visual system to detect and process different aspects of the visual world.


RETINA, 140X MAGNIFICATION
The retina in this image is detached, probably from microscope slide preparation. When a retina detaches, it usually tears (as here) between the pigment epithelium and the outer segments of the rod/cone cells.
"C for COLOR!" A good way to remember cones are for color and rods for "night vision."
SCLERA
CHOROID
RODS & CONES
PIGMENT EPITHELIUM


Images by Brad Caldwell
RETINA, 360X MAGNIFICATION
NERVE FIBER LAYER
GANGLION CELL LAYER
INNER PLEXIFORM LAYER
INNER NUCLEAR LAYER
OUTER PLEXIFORM LAYER
OUTER NUCLEAR LAYER



Visual pathway medical vector illustration diagram. By: VectorMine. https://Dreamstime.com. Paid use.



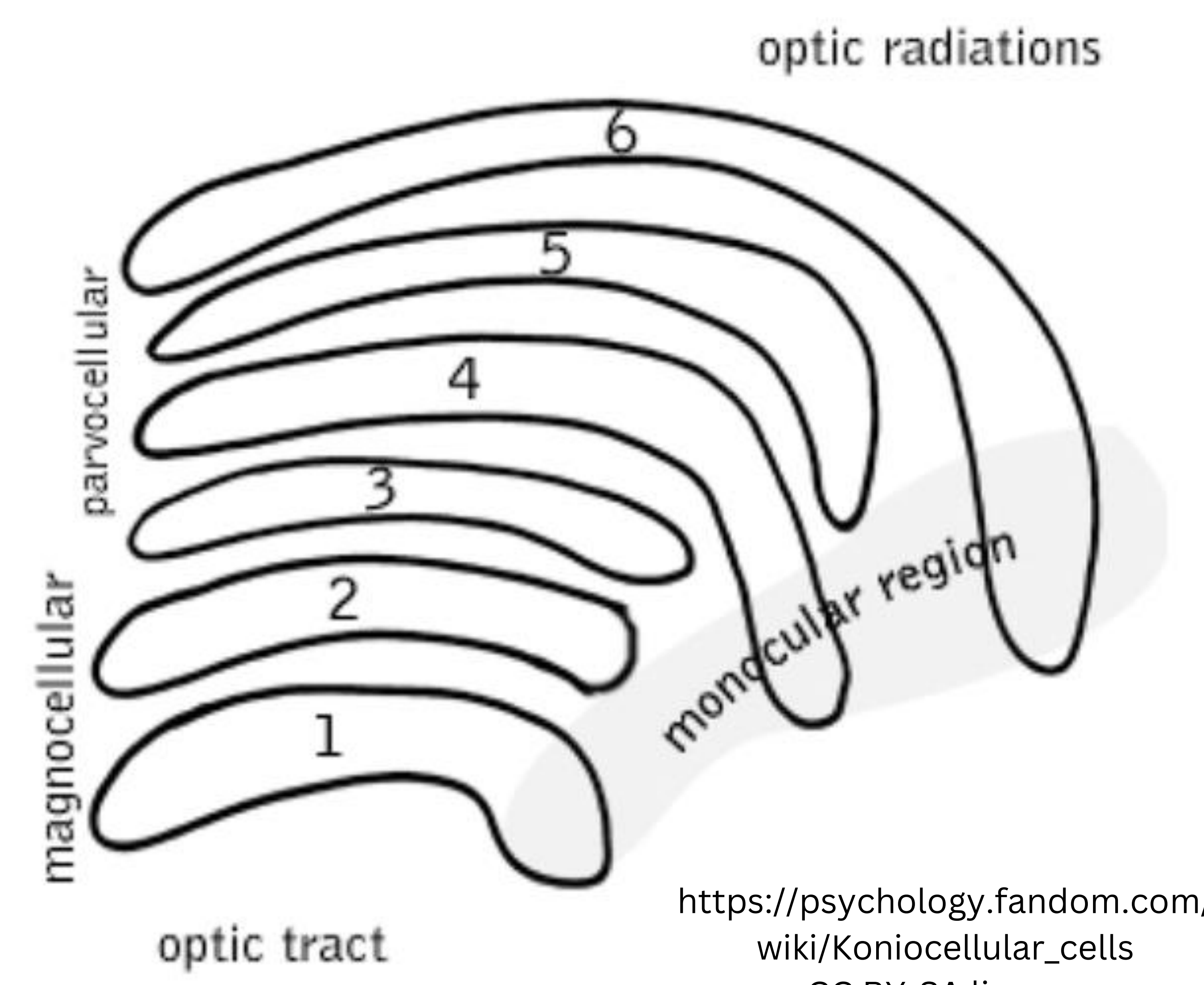
Pupillary light reflex PLR or photopupillary reflex. By: ViktoriiaKasyanyuk. https://Dreamstime.com. Paid use.

Lateral Geniculate Nucleus Layers

- The **parvocellular** layer is primarily responsible for processing information about color, fine detail, and spatial frequency. This layer receives input from the "midget" retinal ganglion cells, which have small receptive fields and are specialized for processing high-resolution visual information.

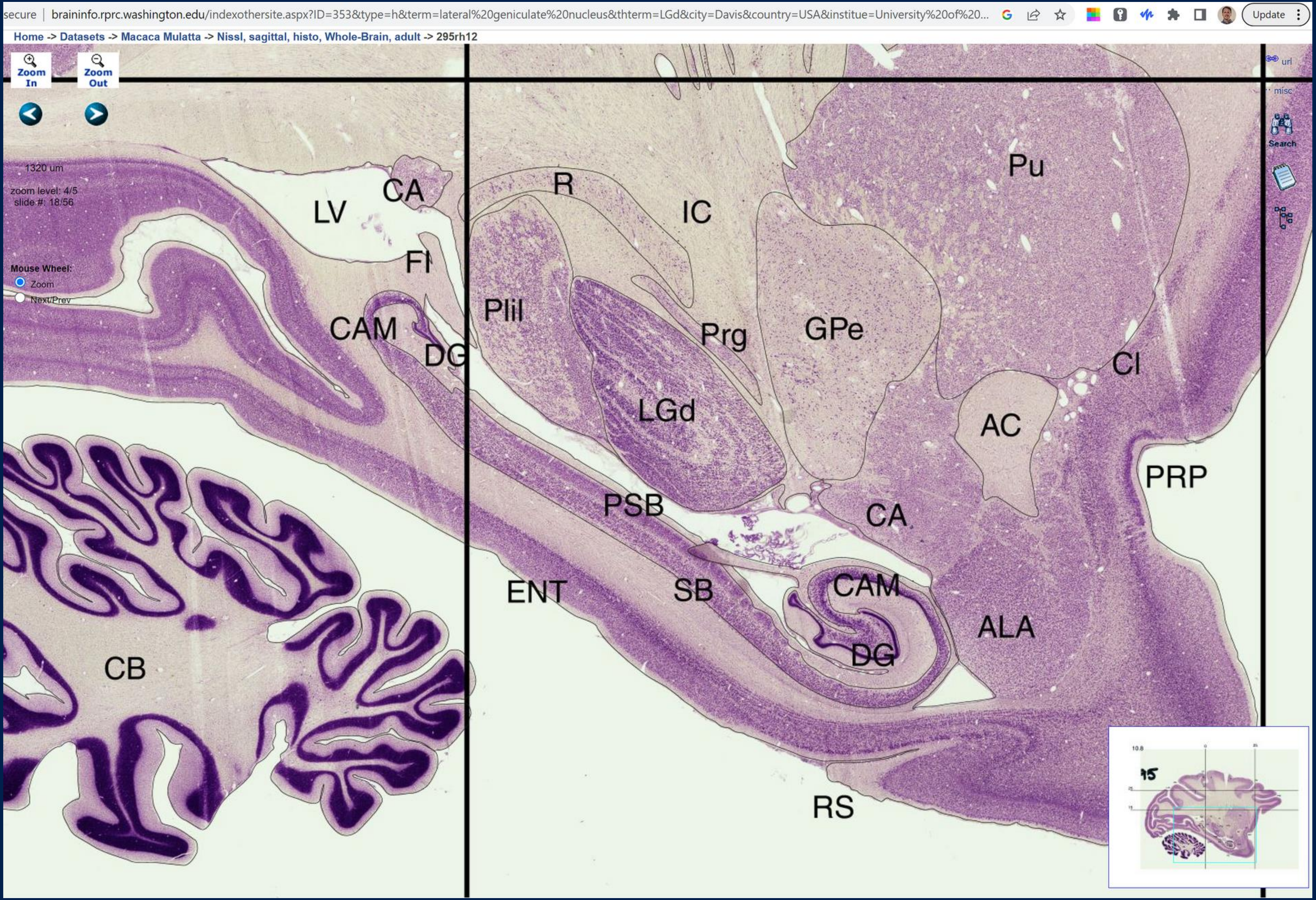
- The **magnocellular** layer is primarily responsible for processing information about motion, spatial orientation, and depth. This layer receives input from the "parasol" retinal ganglion cells, which have larger receptive fields and are specialized for processing low-resolution visual information.

- The **koniocellular** layer (in between) is the smallest and least well understood of the three layers. It is thought to play a role in processing color and other visual features, but its precise functions are still not fully understood.



https://psychology.fandom.com/wiki/Koniocellular_cells

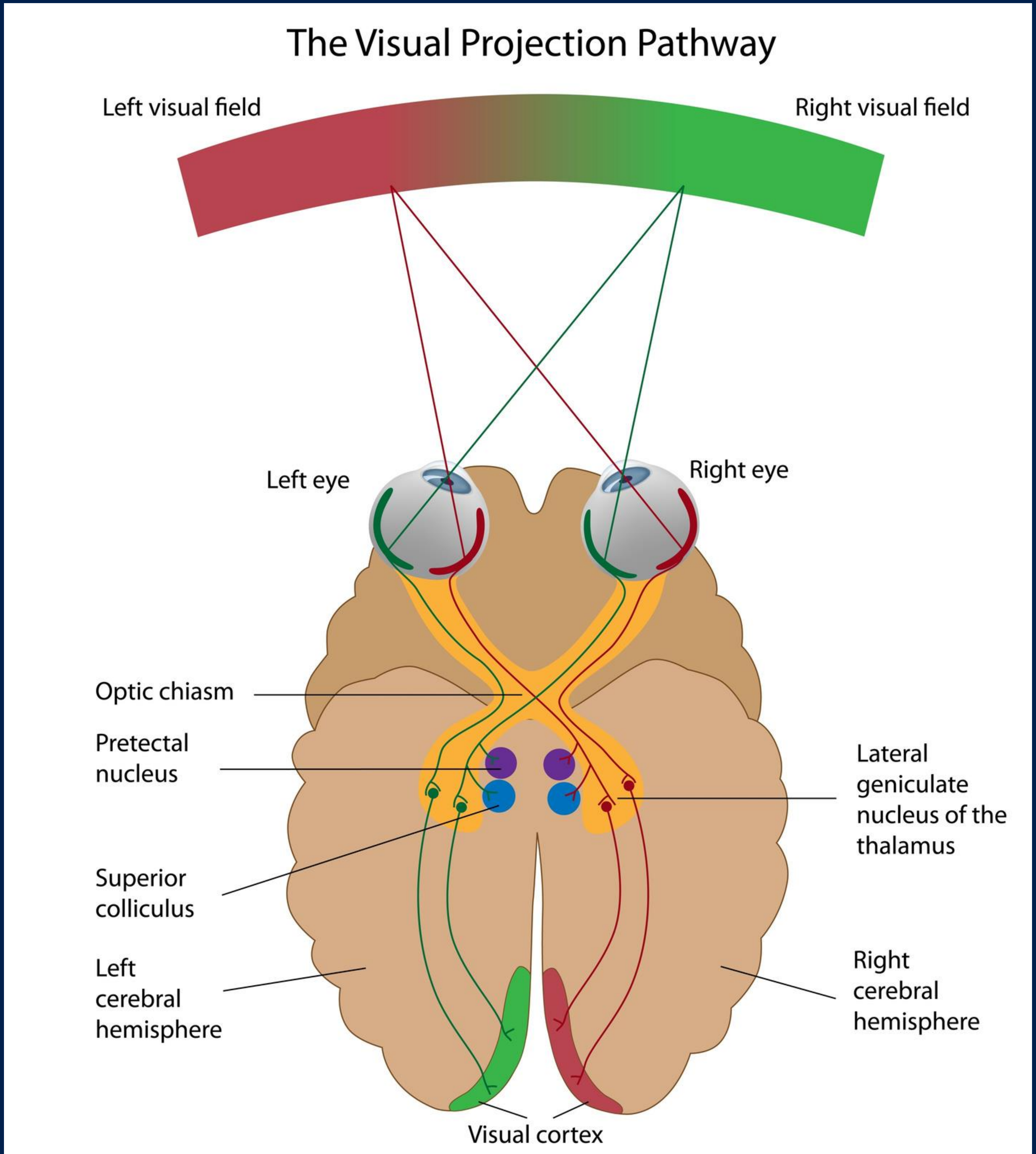
Lateral Geniculate Nucleus



Left visual field goes to right hemisphere and vice versa.

Functions of the four paths from the eyes:

- Pretectal path: Pupillary constriction (to bright light); lens accommodation (near focus). *Involuntary and unconscious*.
- Superior colliculus path: Reflexively turn head to (dangerous) visual stimuli. *Involuntary and unconscious*.
- Suprachiasmatic nucleus path: Circadian rhythms.
- Visual cortex path: Eventually leads to *conscious perception* of visual field. FEF for voluntary movement of eyes. SEF to maintain gaze when turning head.



The visual projection pathway.
By: Alila07. https://Dreamstime.com. Paid use.

Basic Path from Visual Stimulus to Response (Primary Visual Path)

"Light turns green" ↗ Visual input 0ms

→ Retinal ganglion cells (RGCs) 20ms

↓

Lateral geniculate nucleus (LGN) 50ms

↓

Visual striate cortex (V1) 70ms

"What" ↓ ↓ "Where"

Fusiform gyrus 110ms | Post. hot zone (PHZ) 110ms

↓ ↓

140ms Prefrontal cortex (PFC) → 170ms Premotor cortex → 200ms Motor cortex

↑

250ms Motor output

↑

"You start running!"

Requirements from Experience

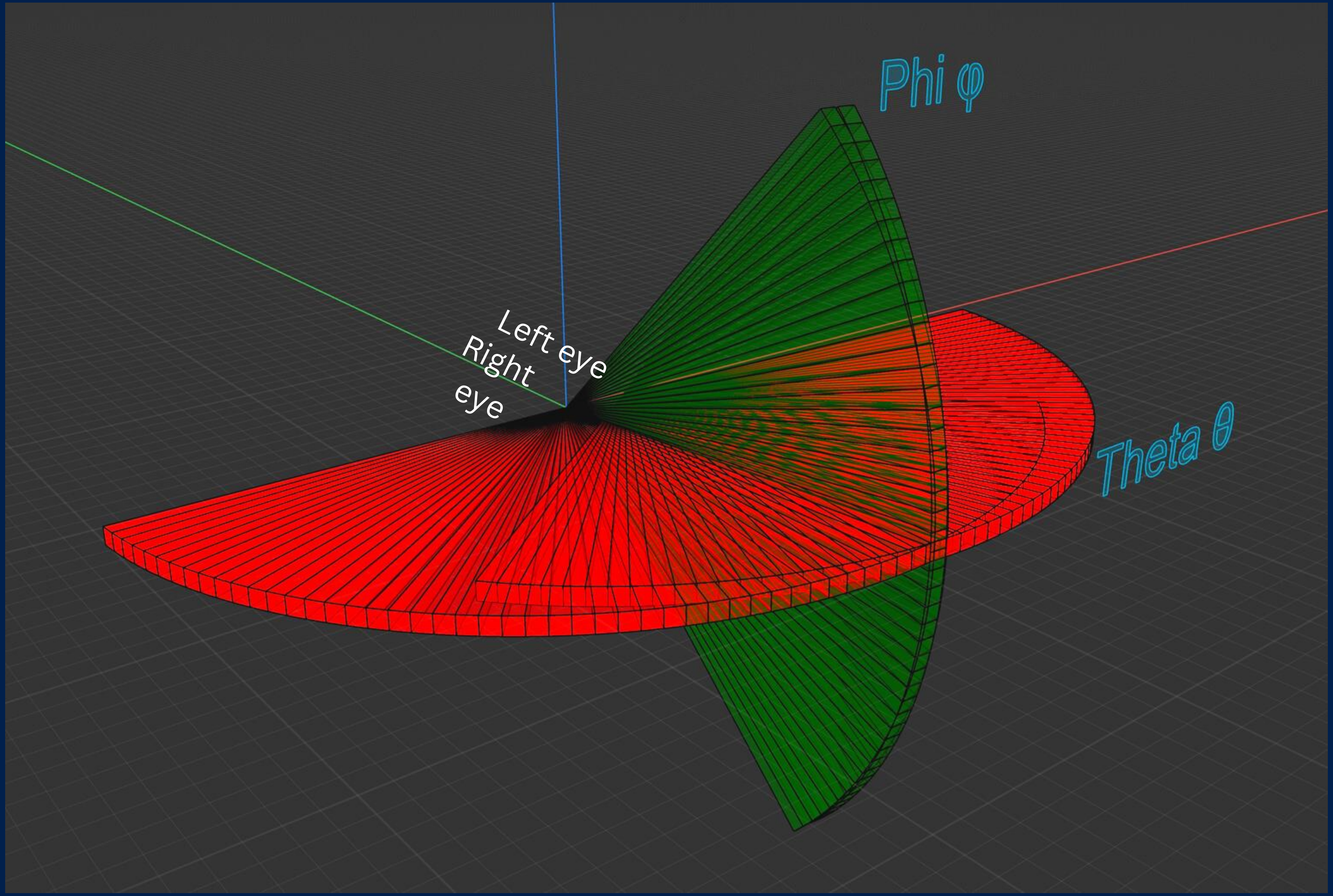
- Seamless merger of left and right visual fields, for left eye contribution, and for right eye contribution.

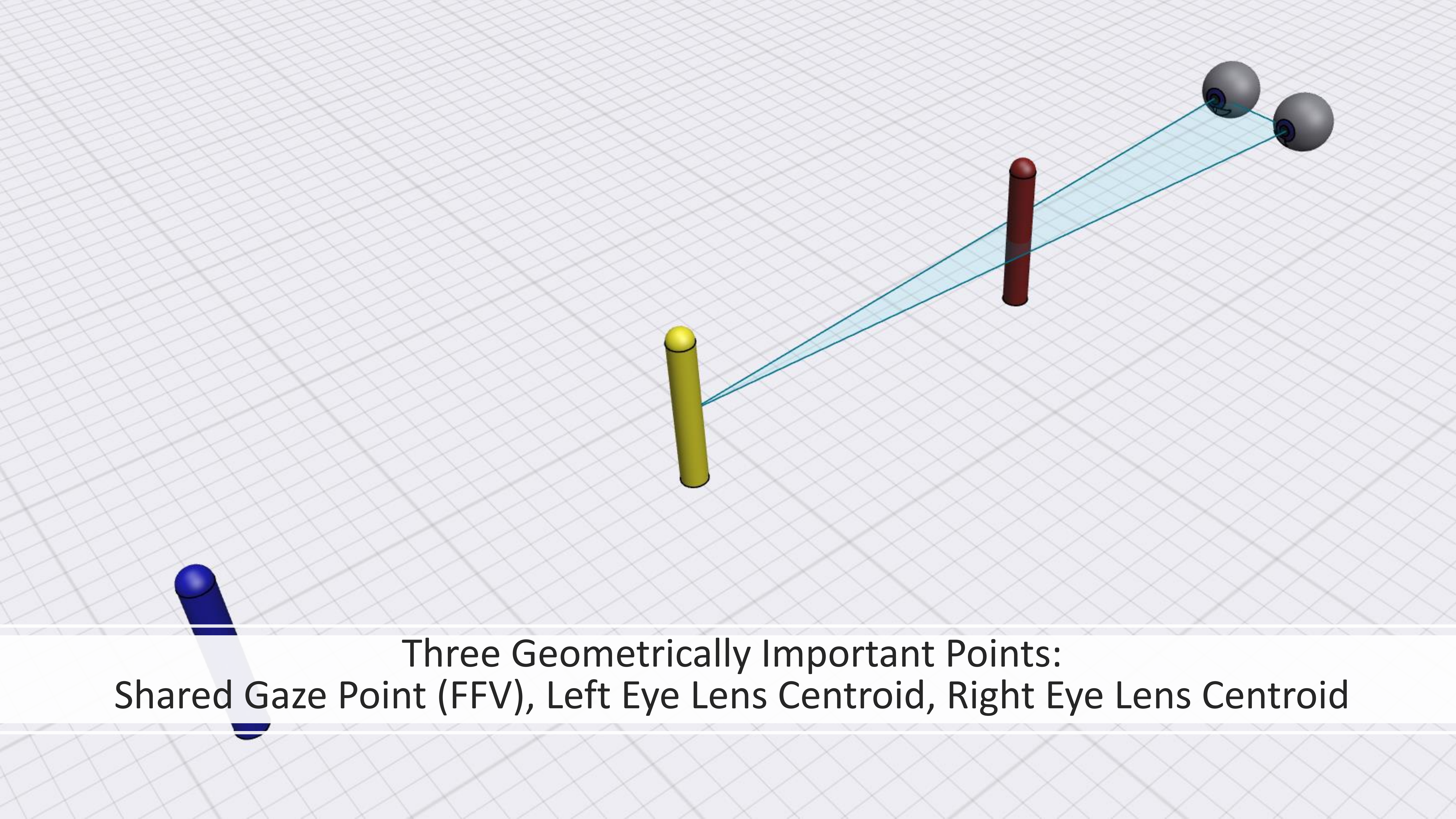
- A single eye sees in 3D - a "voxel theater."

- Left eye voxel theater laid over right eye voxel theater, focused-item (point) is most aligned point, increasing crossed or uncrossed disparities outwards. The shared point fixated by both eyes is dubbed the "focal (central most) foveal voxel" (FFV).

- An estimate for the number of voxels distinguishable by humans is around 7 trillion++.

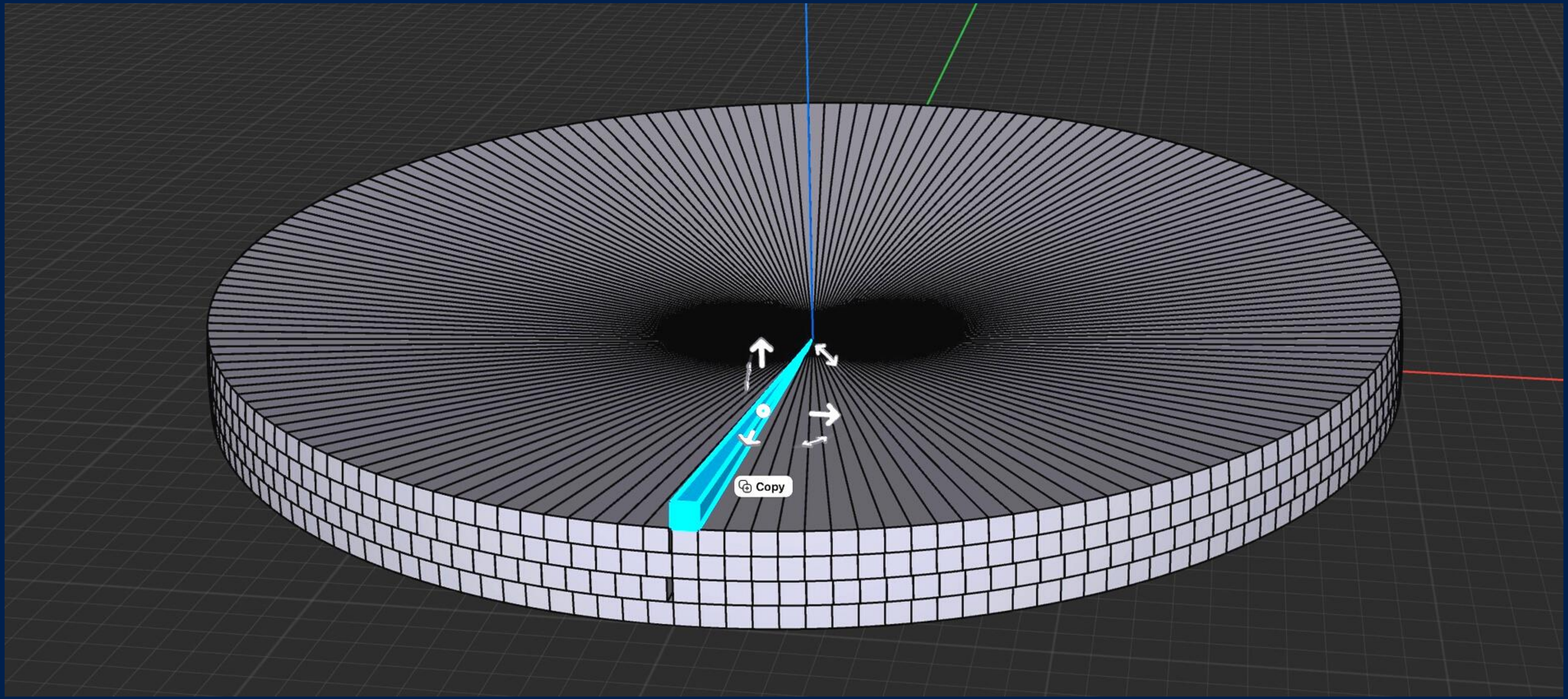
Only Focal Point Is Aligned





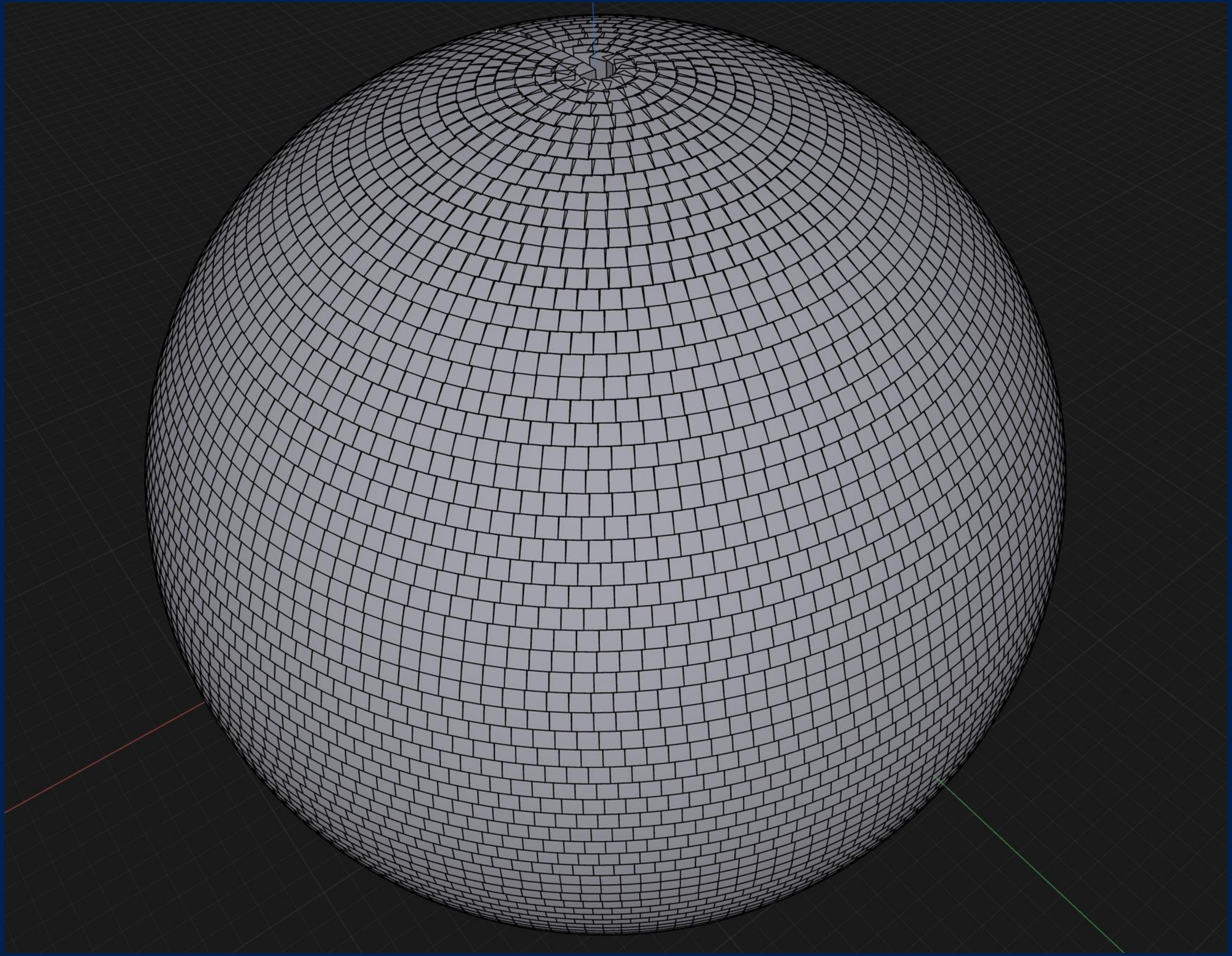
Three Geometrically Important Points:
Shared Gaze Point (FFV), Left Eye Lens Centroid, Right Eye Lens Centroid

Building a Voxel Theater

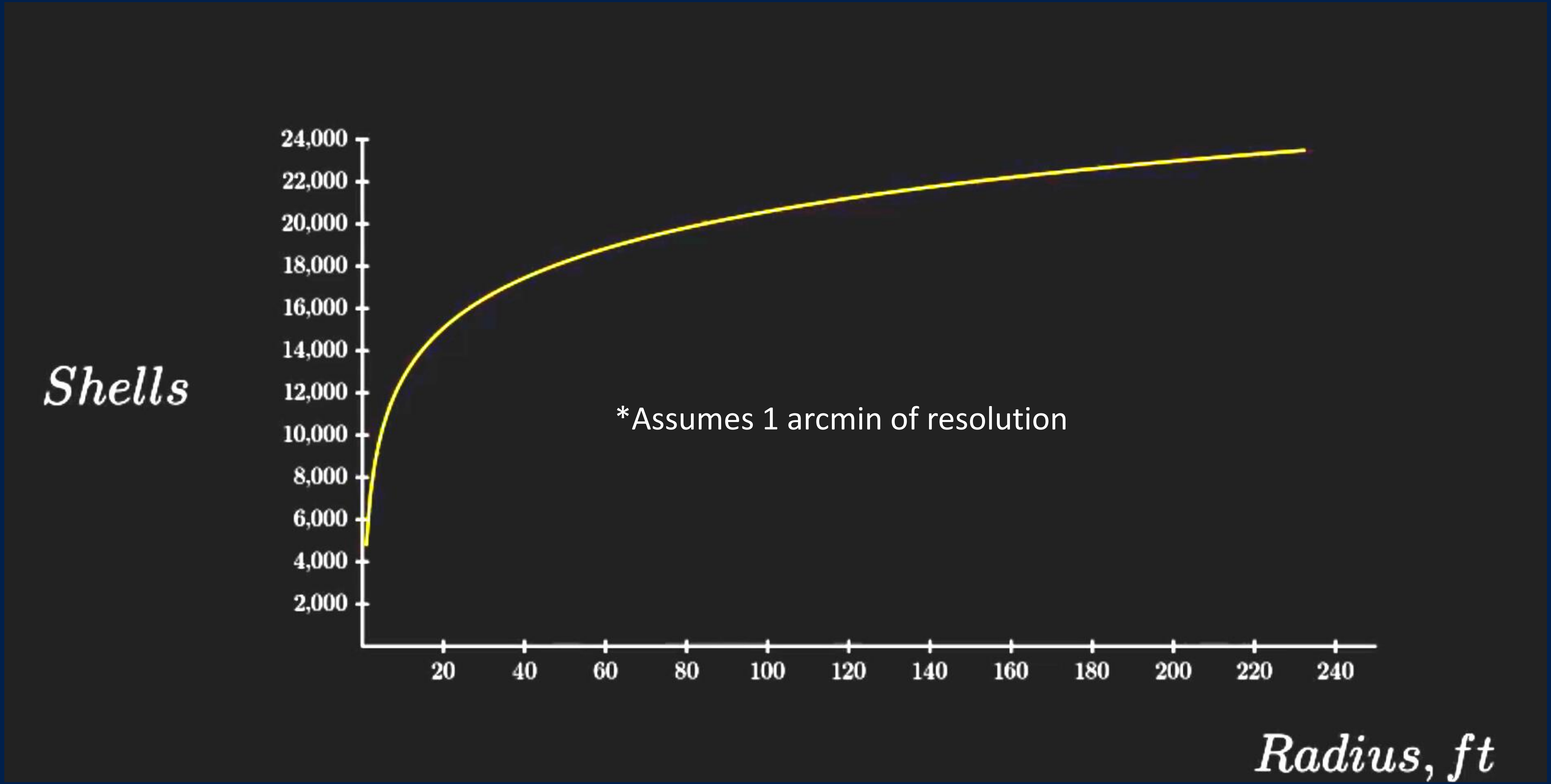


1-Million-Voxel Theater Example

This sphere depicts 10,000 angles of division. If you provide 100 concentric layers of depth (radius), it would be one million voxels. Human vision can see at least a million times more resolution.



Decreasing Depth Resolution at Increasing Radial Distance



7 Trillion Perceptual Voxels?

The calculations behind arriving at 7 trillion voxels discriminable in human perception from 50 million pixels in visual field (moving fovea around) or 200 million pixels in whole sphere of perception (moving head around), assuming roughly 35,000 steps of depth discrimination.

The calculations involve taking the number of pixels in the visual field or sphere of perception, and then factoring in the ability to distinguish depth. With roughly 35,000 steps of depth discrimination, it is estimated that the human eye can discern around 7 trillion voxels in the visual field and sphere of perception.

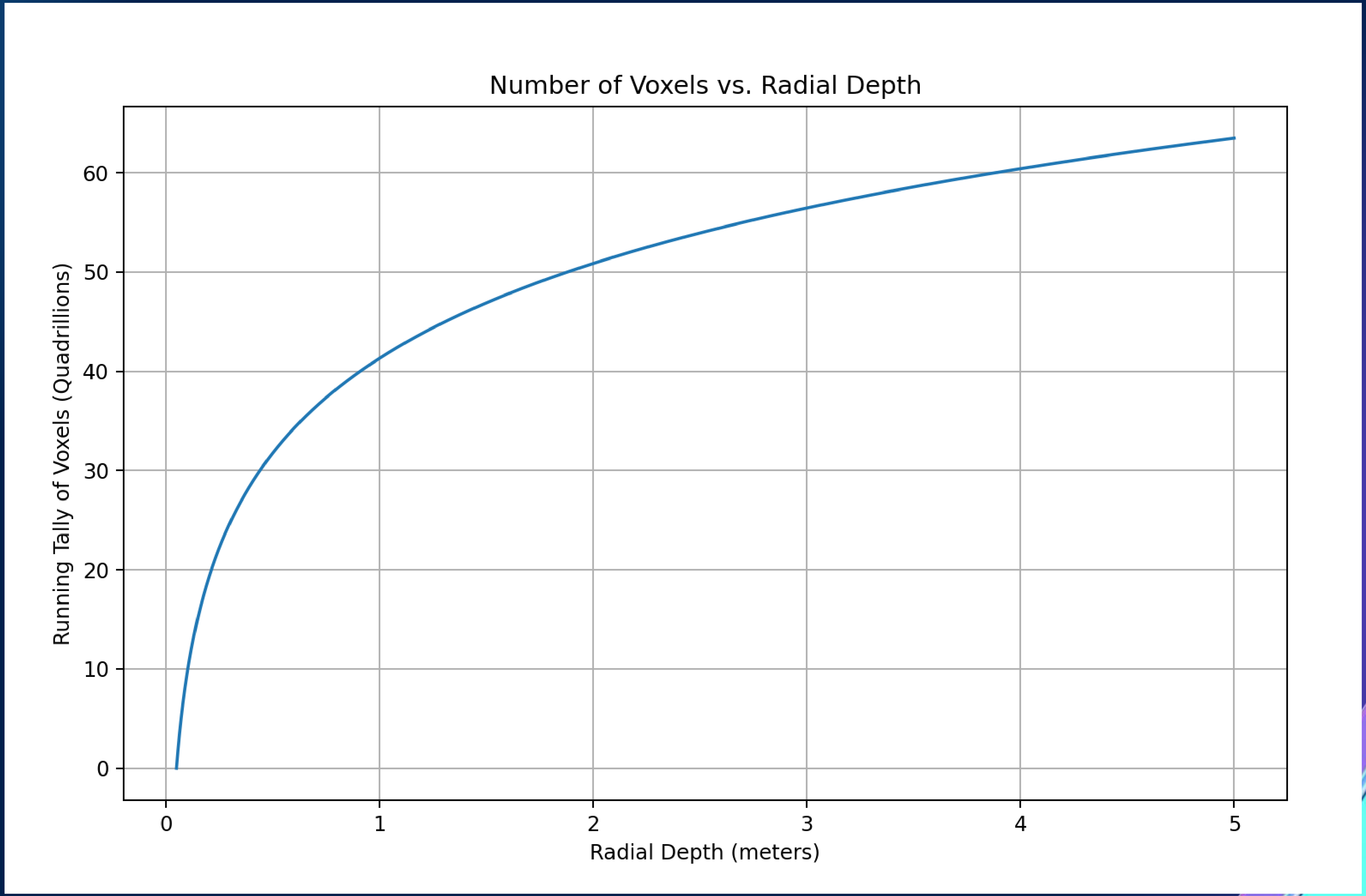
In reality, while the eyes can probably discern well over a trillion voxel locations, their contribution of locational voxels are often added to "skins" that are saved in various places like the parahippocampal place area, fusiform face area, and fusiform gyrus in general. These skins are the 3D boundaries of objects, which we will discuss in the next class on "Bank Necessities," exploring the brain's storage of invariant shapes (skins).

Voxel Resolution of Perceptual Space Under Different Assumptions

Theta Discrimination	Pixels per Degree	Pixels per Square Degree	Square Degrees per Sphere	Total Pixels per Sphere	Depth Range	Number of Shells	Number of Voxels
1 arcmin	60	3,600	41,253	149 million	5 cm - 3.2 km	38,050	5.67 trillion
6 arcsec	600	360,000	41,253	15 billion	5 cm - 3.2 km	380,449	5.67 quadrillion
2 arcsec	1,800	3,240,000	41,253	134 billion	5 cm - 3.2 km	1,141,335	153 quadrillion

Cumulative Depth Increments Per Distance (with 1 arcmin, 6 arcsec, and 2 arcsec visual resolution assumptions)

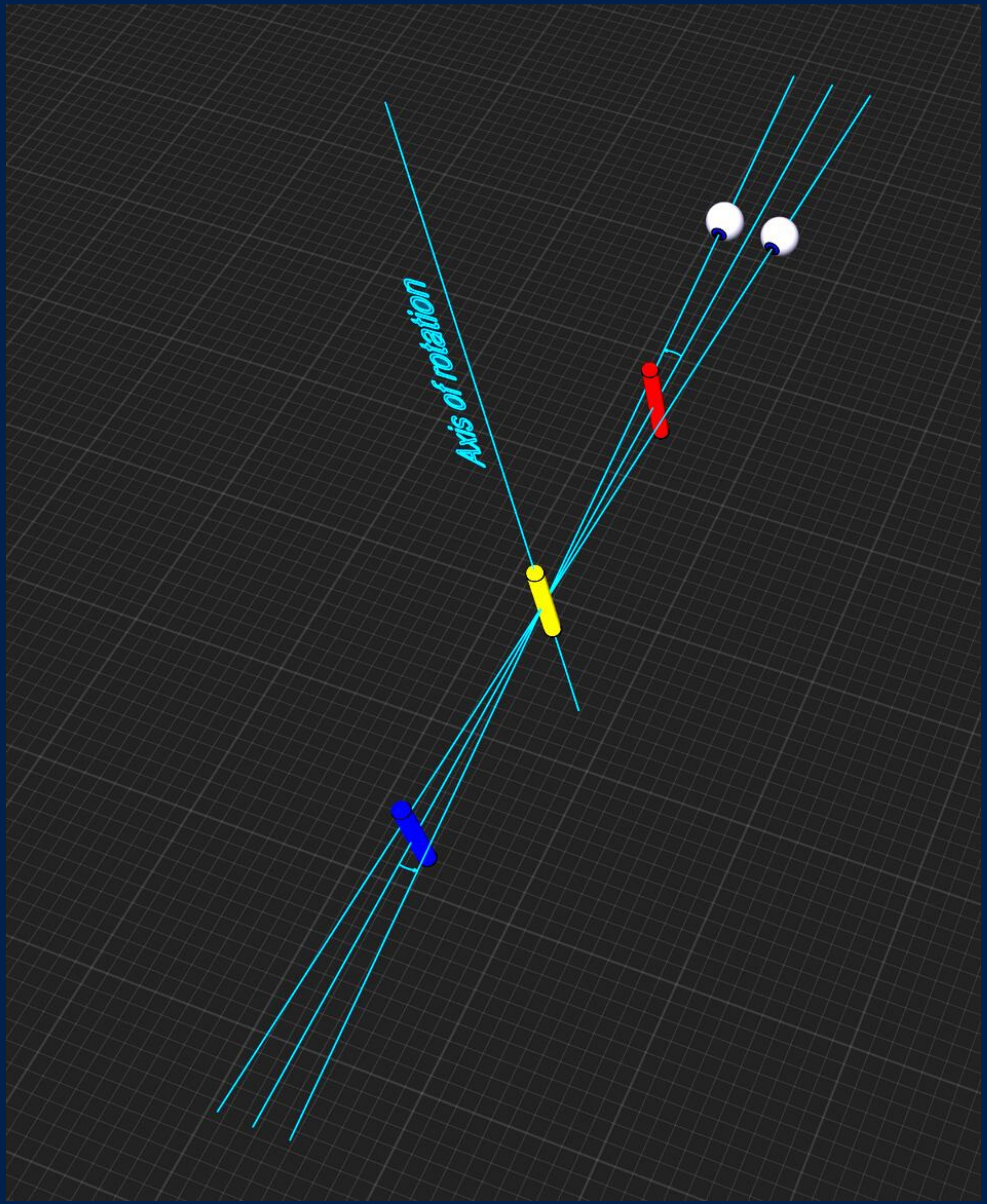
Cumulative Voxels per Depth (2 arcsec resolution)

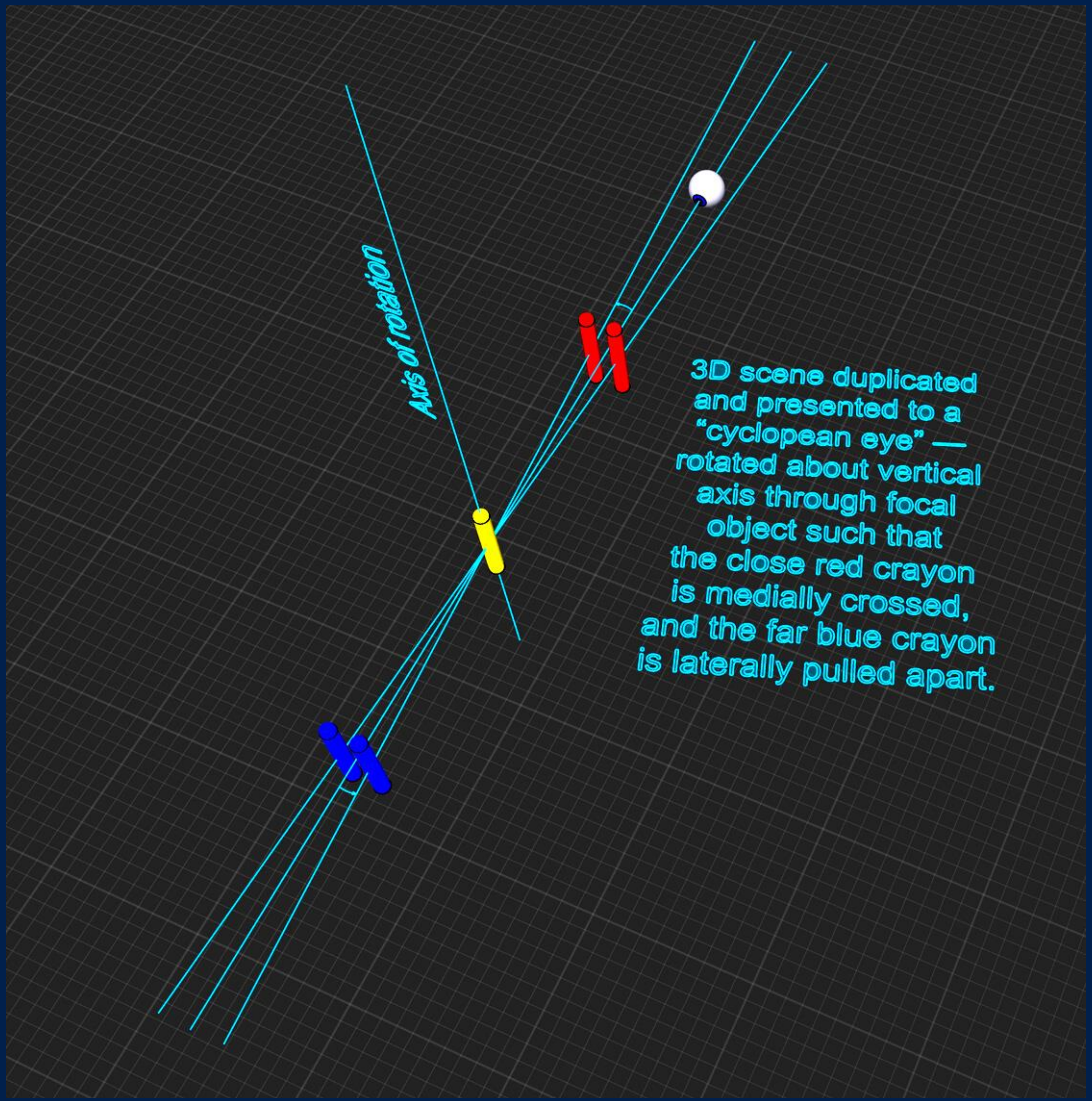


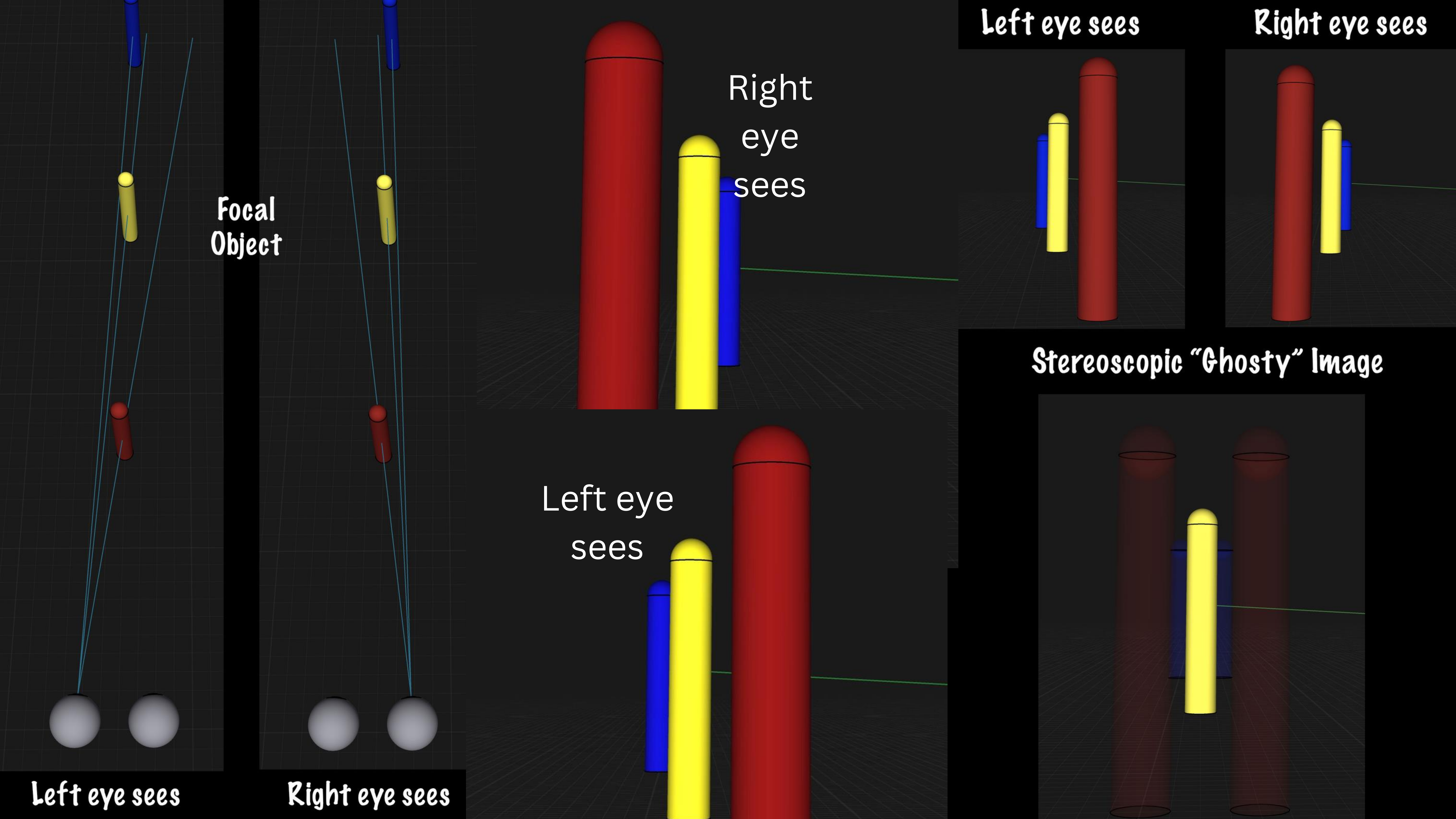
Stereoscopic Vision

1.Stereoscopic vision is often misunderstood as a simple "fusion" of left and right eye scenes, but the actual process is more complex.
2.The brain overlays the two 3D scenes in the same 3D space, using the central foveal voxel (the focus point in space) as the spatial alignment point.
3.Objects in front of the focal point experience increasing crossed disparity, while objects beyond the focal point experience increasing uncrossed disparity.
4.The two 3D realms, constructed from left and right eye information, align at the focal point, creating a vertical axis with the realms appearing counter-rotated slightly from the isometric 3D reality.
5.This sophisticated system accounts for the roughly 2-inch distance between our eyes, enabling accurate depth and distance perception in the 3D world.

Two eyes 2" apart see two different 3D scenes. This seems to be equivalent to a single "cyclopean" perceptual eye seeing one 3D scene doubled (overlaid) and counter-rotated about axis of rotation.






Focal Object
Left eye sees
Right eye sees
Right eye sees
Left eye sees
Left eye sees
Right eye sees
Stereoscopic "Ghosty" Image

Depth from Disparity

It is believed that depth may be calculated at times from disparities. To do this requires foreknowledge of the depth of the focal point, and then the horizontal retinal disparity of a shared feature (edge of an object) between the images from left/right eye can be used to gauge distance in front of or behind focal point. For example, if you hold your index finger in front of your nose while focusing on a far wall, then move your finger away from you, you will notice the two “ghost copies” getting closer. This is horizontal disparity. Vertical disparity can be seen by holding a pencil vertically in front of your right eye and fixating on it, then alternating eyes shows lengthened/contracted angular (perceived) height of the pencil.

Other Clues for Depth

Monocular depth cues:

- Linear perspective (expected radial distance for a given observed size [theta-phi]).
- Occlusion (occluded object is farther, also called "interposition").
- Texture gradient (~spatial frequency of texture pattern denser close up).
- Focal power ("accommodation") of lens (the more spherical the lens, the closer the object).
- Degree of vergence of the two eyes (more vergence for closer items).
- Motion parallax (observer moves laterally, farther objects change slowly). *2-6 arcsecond resolution!*
- Proprioception information integration from somatosensory/vestibular.

Binocular depth cues:

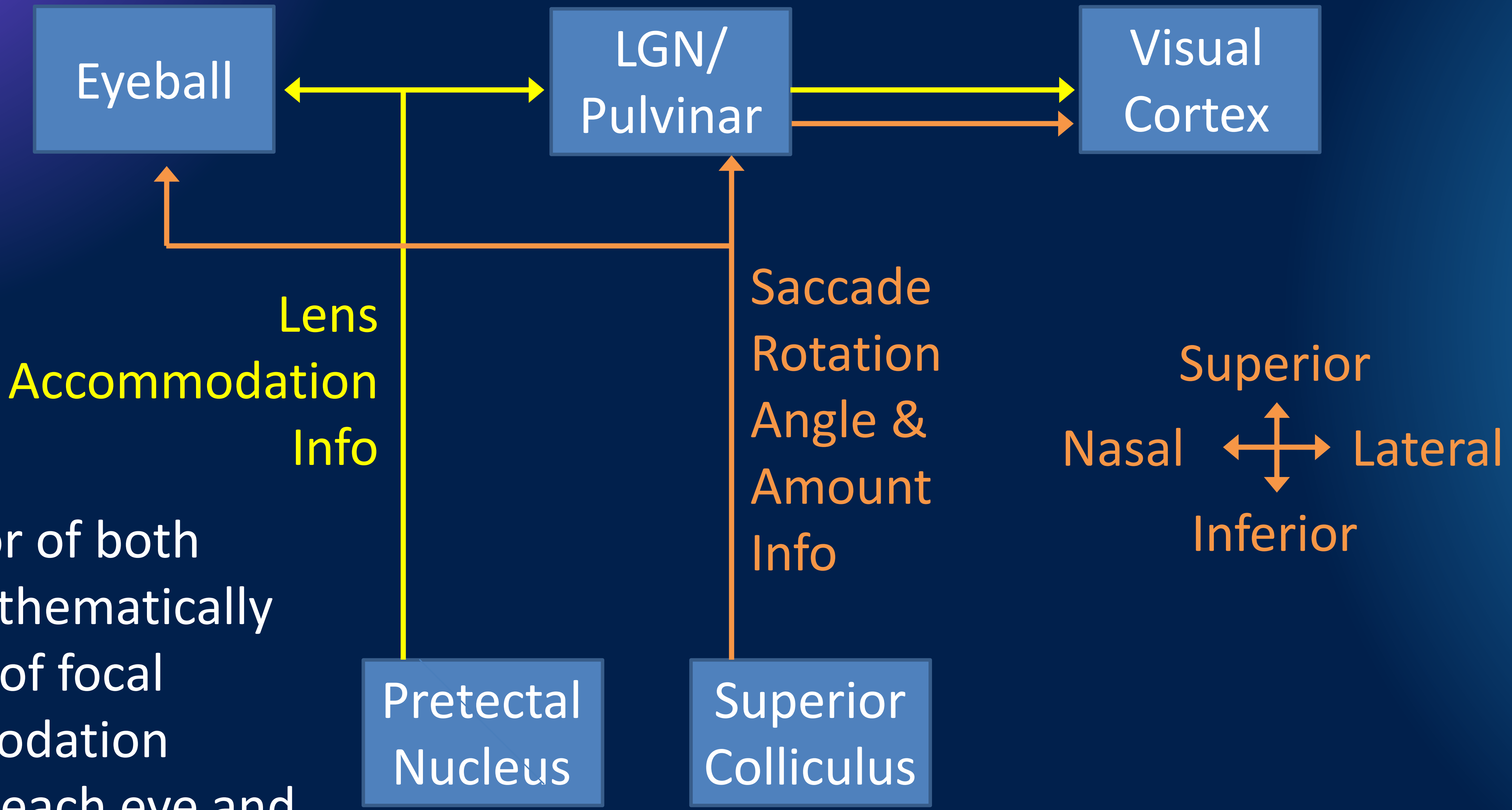
- Stereo disparity. *2-6 arcsecond resolution!*
- Saccadic disparities between eyes.
- Gaze direction vectors, disparities of accommodation.

Howard, I. P. (2012). Perceiving in Depth: Volume 1: Basic Mechanisms. Oxford University Press.

Howard, I. P. (2012). Perceiving in Depth: Volume 2: Stereoscopic Vision. Oxford University Press.

Howard, I. P. (2012). Perceiving in Depth: Volume 3: Other Mechanisms of Depth Perception. Oxford University Press.

Accommodation/Saccade Info



Gaze direction vector of both eyes is sufficient mathematically to determine depth of focal point. Lens accommodation (amount of focus of each eye and the difference) is another tool.

Perceptual Shape Theories

- **Volumetric (Voxel-based) Representation**: The brain represents 3D shapes using small cubic units called voxels, similar to pixels in a 2D image. Each cortical column may learn a unique ensemble of its neurons to map to each voxel.

- **Object-centered Representation**: The brain constructs 3D shape representation using a series of 2D views taken from different angles. This allows recognition of 3D objects from various perspectives, supported by evidence from neurophysiological studies.

- **Structural Description Representation**: The brain uses simpler geometric primitives (e.g., lines, curves, surface patches) and their spatial relationships to represent 3D shapes hierarchically. This approach is inspired by computer vision algorithms that recognize 3D objects based on structural descriptions.

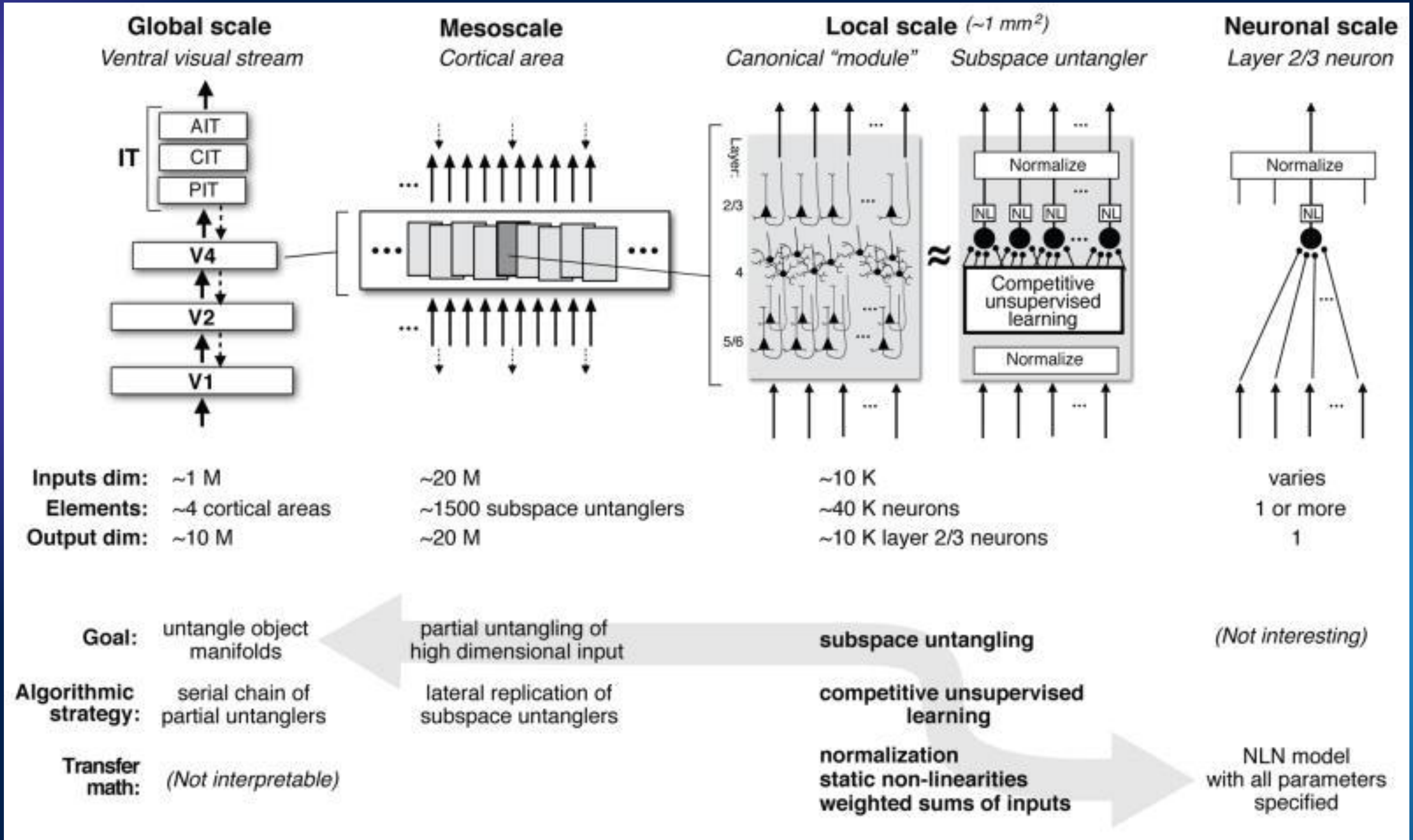
Refined "Bank" Theory

- **Transformable Volumetric Representation (TVR)**: A sophisticated 3D shape representation approach that combines the advantages of volumetric, object-centered, and structural description representations. TVR employs a Euclidean reference system to store isometric 3D shapes that are invariant to perspective, grouping similar shapes within designated regions (e.g., balls, cubes/planes, cylinders/wheels/language/tools). The entire reference system can undergo rotation, scaling, and translation to adapt to real-world situations and align a shape with any orientation it is encountered in, similar to object-centered representation. This flexibility also enables subtle activation of related shapes for familiarity and mental brainstorming.

Multiple Instantiation

- This 'transformable volumetric representation' is hypothesized to have multiple simultaneous instances – one 'copy' dedicated to representing the real-world schema, another for the imagination schema, and a third for the fixed memory store of 2D and 3D shapes (the bank). The fixed memory store is specifically used for recording new shapes to their designated regions, aligning learned shapes with newly encountered orientations, and triggering memories or familiarity based on shared shapes or other connections. All three schemas are overlaid in 3D space.

Encoding Depth



Here we highlight four potential abstraction layers (organized by anatomical spatial scale), and the approximate number of inputs, outputs, and elemental sub-units at each level of abstraction (M=million, K= thousand). We suggest possible computational goals (what is the "job" of each level of abstraction?), algorithmic strategies (how might it carry out that job?), and transfer function elements (mathematical forms to implement the algorithm). We raise the possibility (gray arrow) that local cortical networks termed "subspace untanglers" are a useful level of abstraction to connect math that captures the transfer functions emulated by cortical circuits (right most panel), to the most elemental type of population transformation needed to build good object representation (see Fig. 2C), and ultimately to full untangling of object identity manifolds (as hypothesized here).Paid Use.
Figure 5. iCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? Neuron, 73(3), 415-434.

Vision: Large Field and Focus Point Combined

1. Vision is often perceived as a broad, continuous experience, but it also involves a focused point where the central rays of both eyes converge.
2. This focal point, or "3D cursor," determines the eyes' vergence and orientation, playing a crucial role in depth perception and spatial awareness.
3. As we track moving objects, like a housefly, our eyes make several saccadic movements, adjusting their positions and depths independently to maintain focus.
4. Within a single saccade, the object may change position and depth, requiring each eye to adapt its movement accordingly.
5. This dual nature of vision, combining a broad field with a precise focus point, allows us to efficiently process and navigate the complex visual world around us.

Experiment: Monocular Movement of FFV

If I have one eye open and focus on the tip of my right index finger, I can focus on a single pixel. But if I focus slightly above the index finger, and change my lens accommodation, I can focus on back wall. It is essentially the same pixel, but two very distant voxels. Thus, I can move my focal gaze point (FFV) to any point in 3D space with only a single eye open.

Citation for "3D Cursor" Aspect of Vision

"Most eye movements in the real-world redirect the foveae to objects at a new depth and thus require the co-ordination of monocular saccade amplitudes and binocular vergence eye movements. Additionally to maintain the accuracy of these oculomotor control processes across the lifespan, ongoing calibration is required to compensate for errors in foveal landing positions. Such oculomotor plasticity has generally been studied under conditions in which both eyes receive a common error signal, which cannot resolve the long-standing debate regarding whether both eyes are innervated by a common cortical signal or by a separate signal for each eye. Here we examine oculomotor plasticity when error signals are independently manipulated in each eye, which can occur naturally owing to aging changes in each eye's orbit and extra-ocular muscles, or in oculomotor dysfunctions. We find that both rapid saccades and slow vergence eye movements are continuously recalibrated independently of one another and corrections can occur in opposite directions in each eye. Whereas *existing models assume a single cortical representation of space* employed for the control of both eyes, our findings provide evidence for ***independent monoculomotor and binoculomotor plasticities*** and ***dissociable spatial mapping for each eye***."

- Maiello, G. et al. Monocular and Binocular Contributions to Oculomotor Plasticity. Sci. Rep. 6, 31861; doi: 10.1038/srep31861 (2016). Emphases mine.

Citation for FFV/"3D Cursor"/Gaze Point

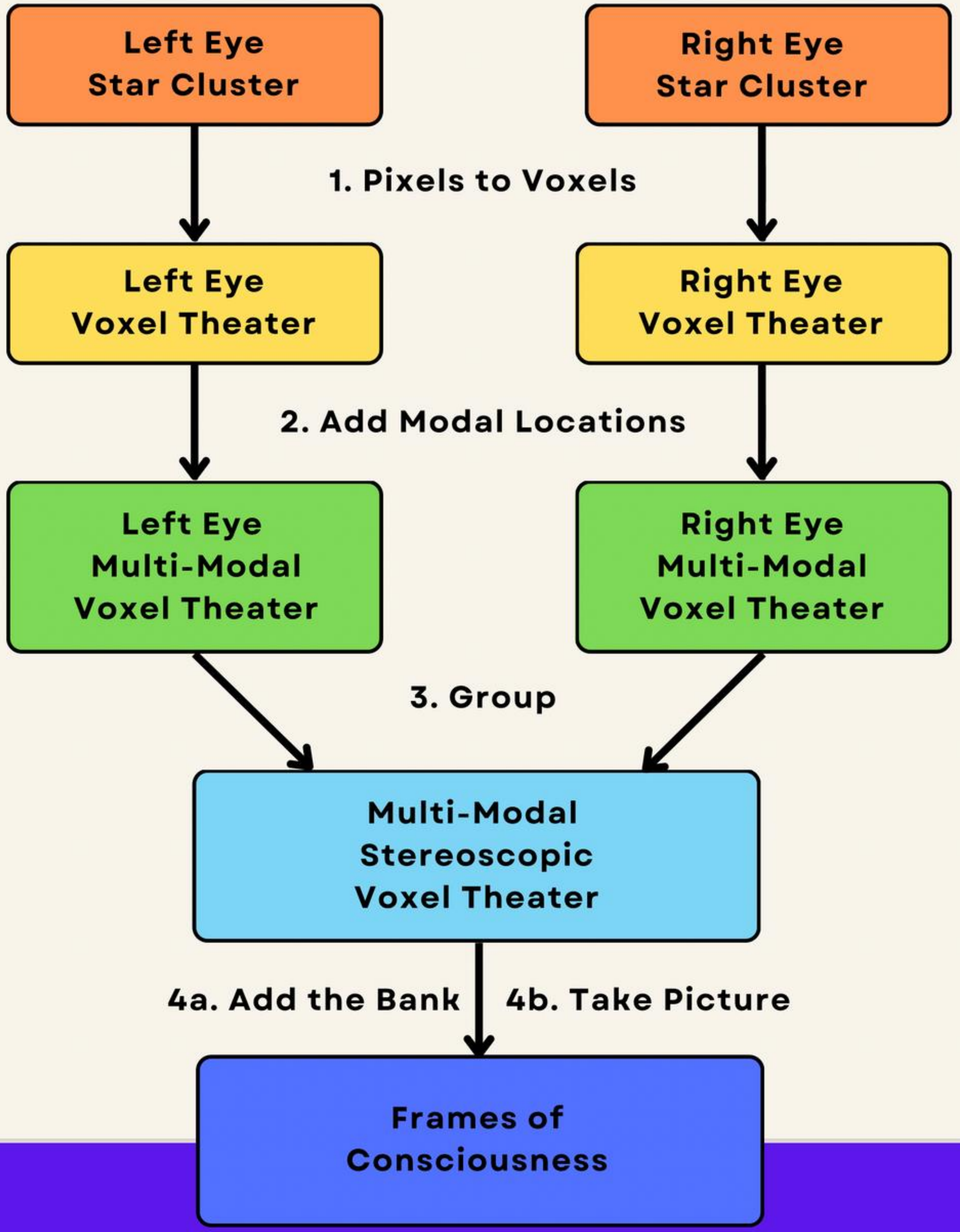
"Humans and many animals with overlapping binocular visual fields ***continuously shift their gaze point in three-dimensional (3D) space*** in order to bring targets of interest onto the high-resolution fovea or area centralis of the two eyes. These gaze shifts require highly precise and coordinated fast saccadic and slow vergence eye movements. To achieve binocular coordination in a 3D environment, Hering proposed that both eyes are innervated by conjugate command signals. Alternatively, Helmholtz argued that binocularly coordinated visual systems evolved from independently moving lateral eyes and that each eye is controlled independently (for a review see). The finding that some saccadic eye movements in 3D space result in each eye moving by a different amount has generally been seen as a challenge to Hering's law of equal innervation. However, it has been argued these data may yet be explained by a vergence command superimposed on a conjugate saccadic command."

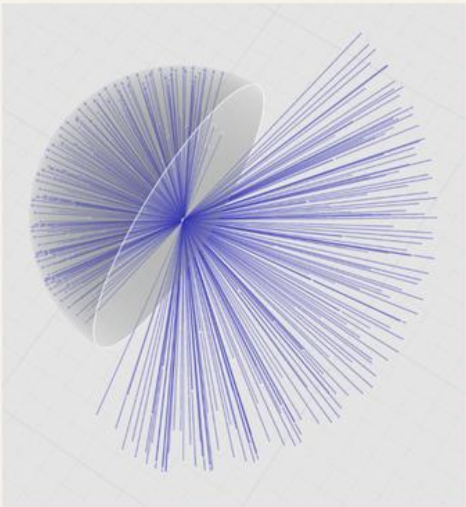
- Maiello, G. et al. Monocular and Binocular Contributions to Oculomotor Plasticity. Sci. Rep. 6, 31861; doi: 10.1038/srep31861 (2016). Emphasis mine.

Consider Housefly as Illustrating a Moving FFV/"3D Cursor"

As a point of geometrical requirement, the "focal (center most) foveal voxel" of either eye will be pointing in some specific direction (some pixel), and there will be some depth of lens accommodation (some radius, therefore some voxel). The world would look muddy if, when two eyes are open, they did not agree upon a single point in 3D space to use as shared FFV. Even with this shared FFV, there is the gradient of increasing crossed disparity in front of the FFV and gradient of increasing uncrossed disparity behind the FFV. There will be slight disparities of size and even of orientation of the pixels of each eye going away from the FFV.

Visual Perception

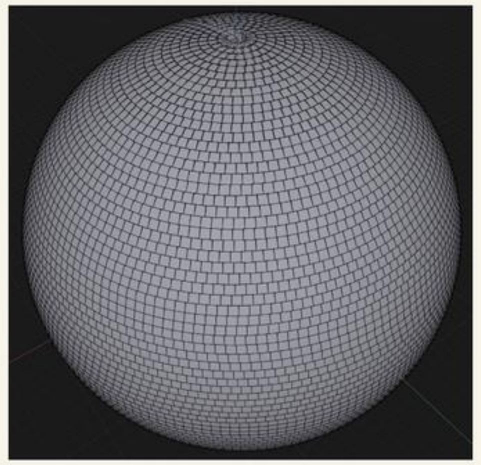




Star Cluster

The information collected by the retina is essentially theta-phi information. As such, it can be said that each "pixel" represents a particular vector in a composite "star cluster" of vectors, the sum of which make up the 2D info of the image.

1. To convert pixels to voxels, provide a radial depth (provide a "scalar" for each "vector").

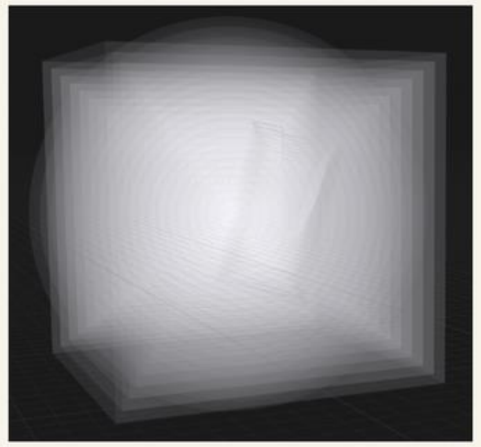


Voxel Theater

Each eye's "voxel theater" has its centroid at the avatar's pupil – i.e., your (perceptual) pupil. The eye can discern an estimated 1-10 trillion voxels around your perceptual space.

2. Add the locations of avatar, sounds, touches, coloring.

3, 4a, b. Add/rotate/scale the bank and use to take pics.



The Bank

The bank is the brain's invariant storage of 2D and 3D shapes. Your avatar's skin shapes are saved here, as well as everyday objects, faces, scenes, written language.