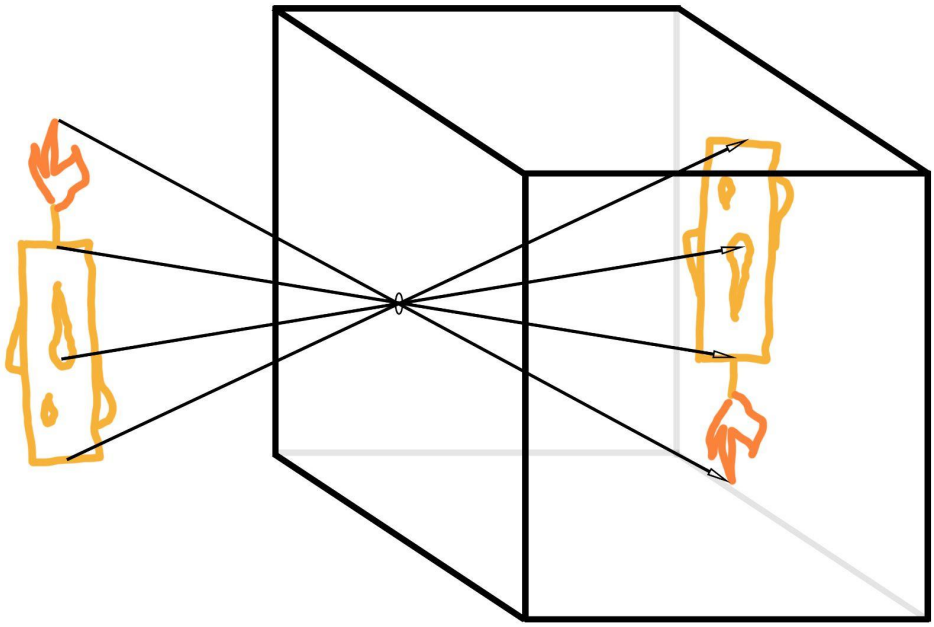
Perceptual Optics — 1. What Does a Pinhole Do?

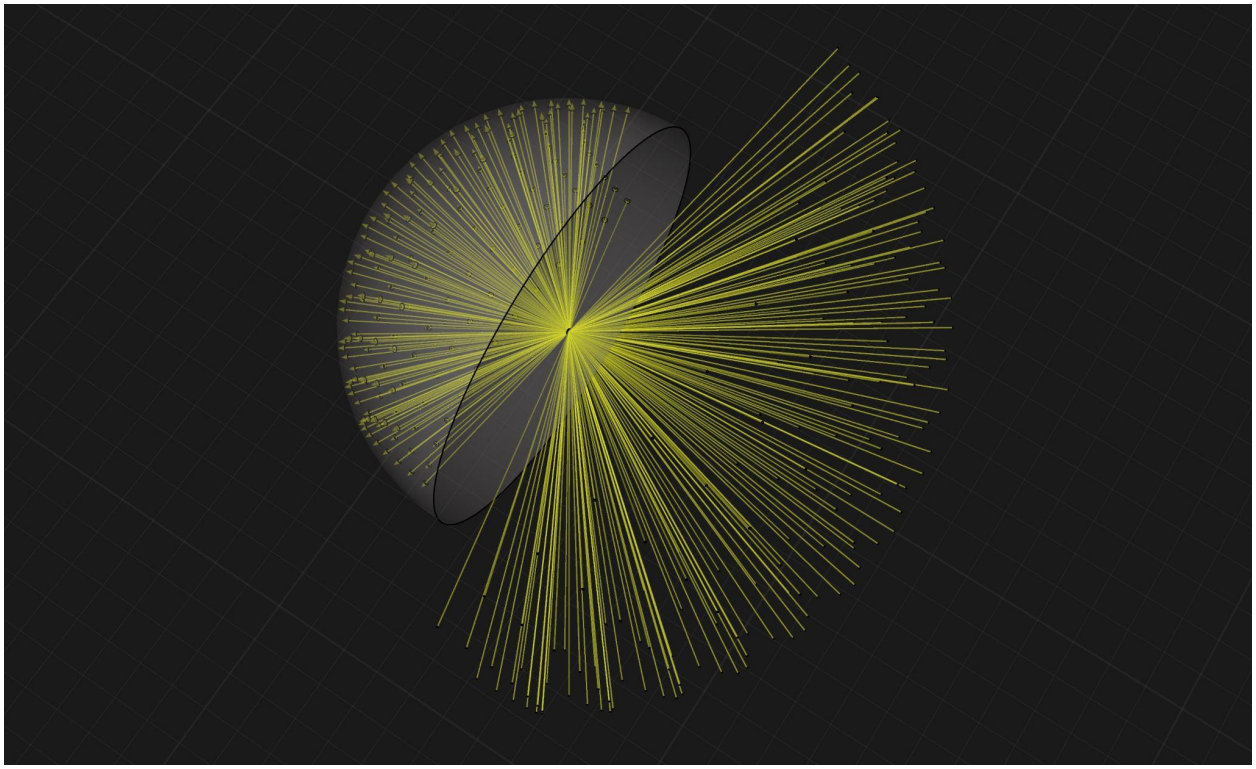
Brad Caldwell

The **pinhole** is the starting point for understanding cameras, eyes, and visual perception. It forces one to exclude the oft-used, but terribly confusing, diagram of “parallel rays,” as a pinhole is capable of creating an image with no rays parallel. In essence, a pinhole enables the spreading out of an image over a larger surface so that each pixel can be separated, without letting any of the “mud” of confusing rays onto the landing area. Pinhole cameras *exploit the fact that images already exist at every point in space, and then enlarge them to a reasonable size*.

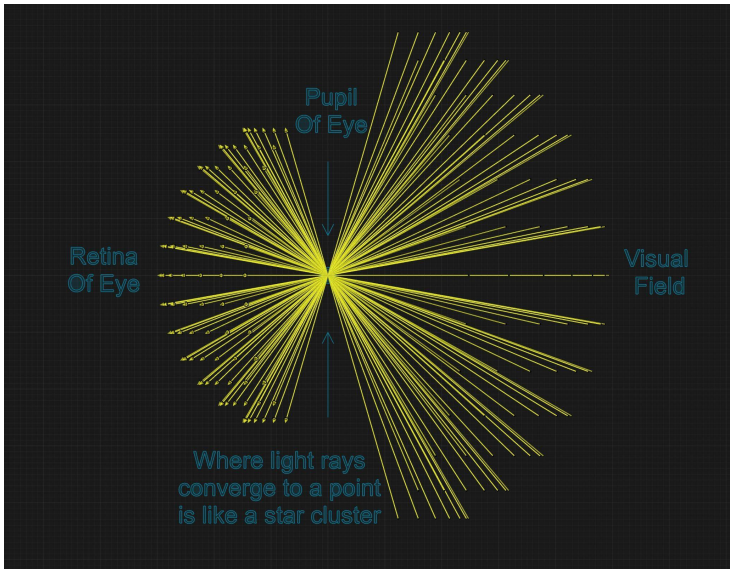


There exists a 360º (really, 41,253 square degree) picture at every tiny point in space; albeit to get a specific image, you have to take it from a specific place. *The pinhole camera doesn’t do anything magic — it just allows half of the image in any infinitesimal point to get magnified to a degree that is useful to us* (the other half of the sphere is occluded by the camera box!). The image is inverted because you’re catching the aftermath after all the rays have converged to a point and then kept going. But the 3D vector orientation of each ray is exactly the same while “right side up” and after getting inverted, much like clockwise actually is counterclockwise when viewed from the opposite side. If the eye and brain use the 3D vector rather than the pixels for creating perception, there is no need to “flip it back over.” Note that the inversion point of the rays, at the pinhole, is ground zero of where one is sampling the universe for an image. The back of the camera is useful, but its location is not relevant. The rays sketched here are two-dimensional, but you can imagine the rays from the third dimension. Note that everything from every degree of distance is “in focus” by definition.

The box design works out okay, as we are used to flat (planar) 2D images, but if we wanted every sensor on the back of the camera to already be at the right attitude to catch the rays of light (and represent the attitude of light captured), we would use a hemisphere camera:



The eyeball goes halfway, and uses an entire sphere, which means that the rod/cone sensors have to bend a little on the sides (like trees growing towards light) to be at the right attitude to catch the incoming light, which they do.`



The pinhole, despite its simplicity, has some drawbacks:

(1) Not much light is allowed in.

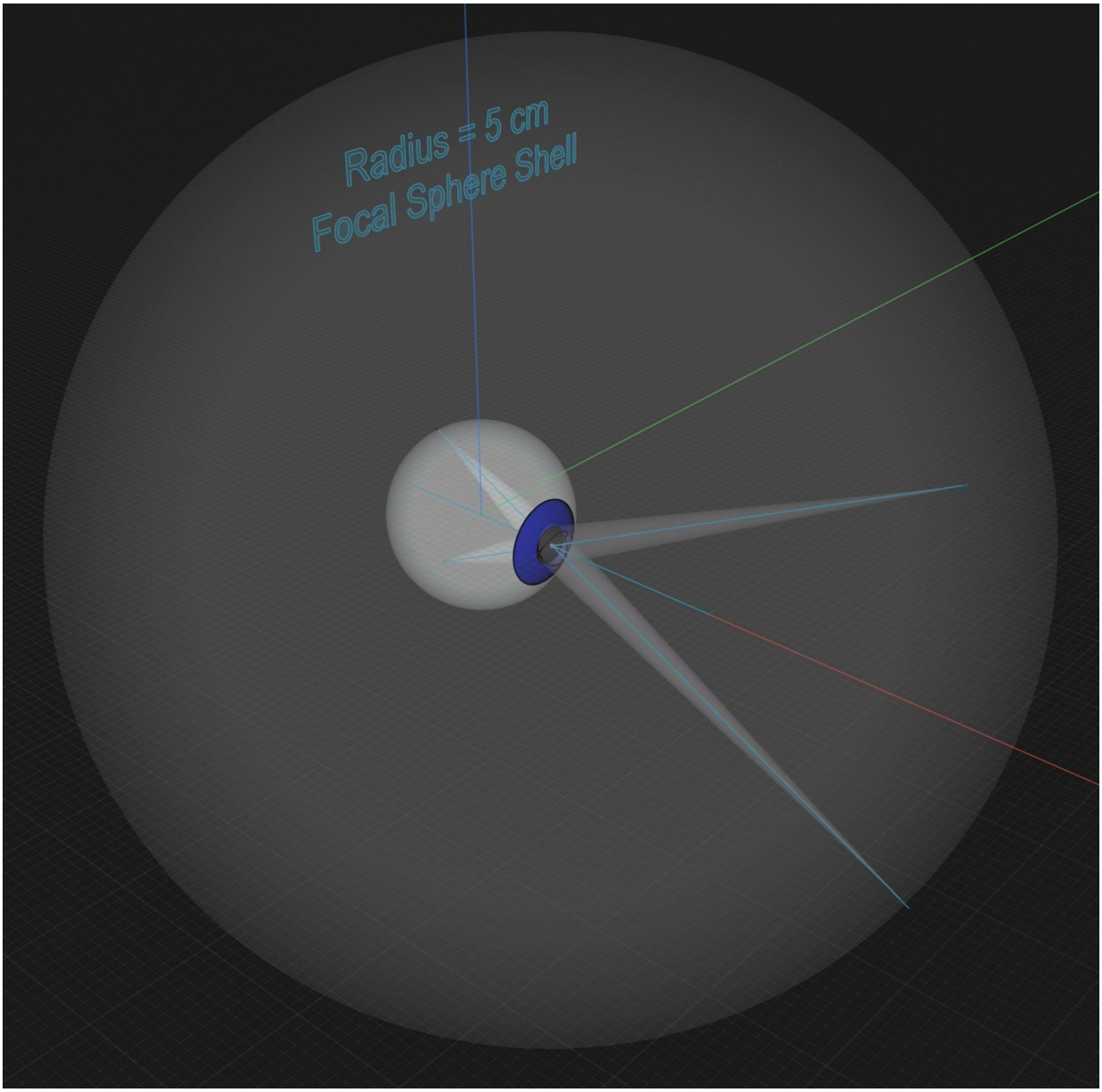
(2) Even with a fairly small opening (1 mm), it still allows light from a single point in the world to spread slightly by the time it reaches the back of the camera.

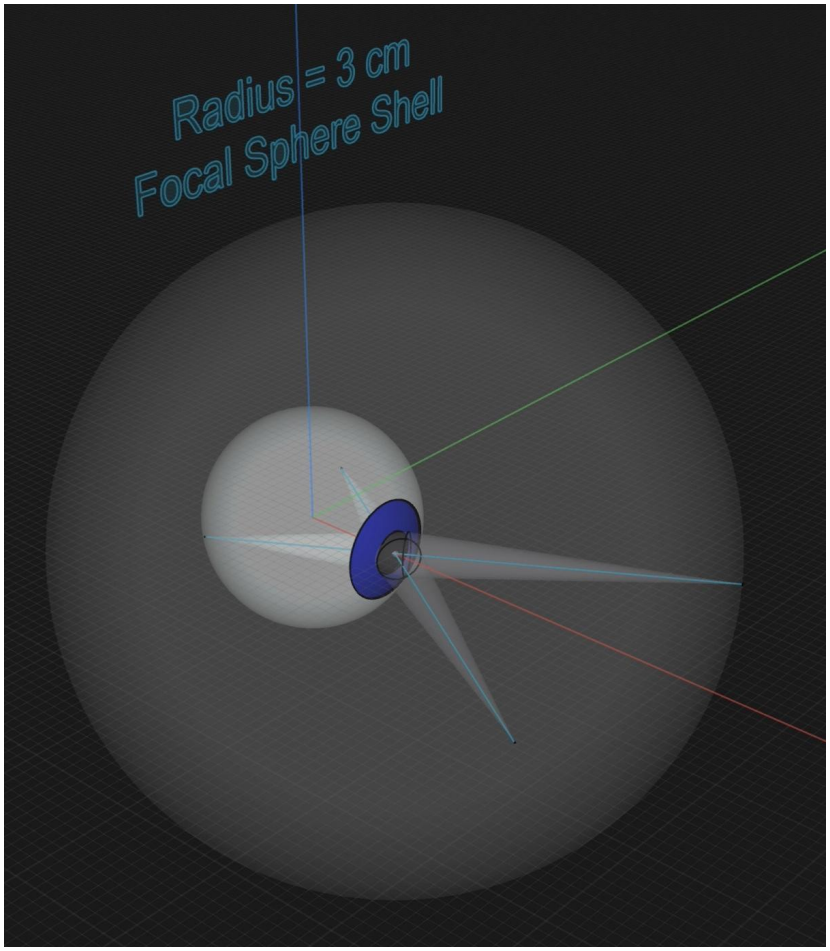
(3) With an extremely small opening (0.1 mm), the edges of the pinhole begin to diffract the light rays and cause some blurring.

2. What Does a Lens Do?

Brad Caldwell

In lesson one, we saw that the pinhole camera was pretty good, but could improve in letting more light in and in increasing image clarity. To let more light in, we make the hole a good bit larger. In doing so, we must now add a **lens** to *bend* the now present *cone of light* from each point in the world back to a point upon reaching the sensor. In the eyeball below, each light ray is able to spread into a cone the width of the pupil upon reaching the pupil, and must be bent back to a point at just the right distance. *But this solution introduces another problem! Now we can only have a single "plane" in focus, because light rays from differing distances make different "solid angle" cones that need differing degrees of focusing power to correct them!*





The human eye uses the cornea to provide most of the focusing power, while allowing the lens to be variable in its degree of power, allowing us to focus on near or far things.

In reality, focusing on near things (as image at left) requires the most focusing power, and the human uses a triad — eye focus, vergence (turning eyes medially), and constriction of the pupil (to make it more like a pinhole!). It seems the eye is already well on its way to starting off with voxels rather than just pixels (understanding the depth of where the light came from), at least for the one in the foveal center-view (the most important one).

As a point of explanation, a lens is able to bend light based on the shape and the index of refraction. But remember that the whole point is, as in the image above, to bend a muddy cone of light back into a single point at a certain distance. In so doing, the lens also solves the other two problems of pinholes — too much “circle of confusion” from size of hole and too much “circle of confusion” from diffraction effects — the lens can, at least for a single focal “plane,” artificially create an extremely high level of resolution and clarity!

As a final point concerning lenses, consider that each point that is equidistant from the pupil or lens will create a cone of light that shares the same solid angle and therefore needs the same power of focus to resolve to a point on the sensor. What shape is equidistant from the center of the lens? A hemisphere shell! Surprisingly, there is no such thing as a flat “focal plane,” but rather a “focal sphere shell.” Of course, the back side of the sphere would be occluded, so it’s more like a hemisphere shell.

There’s no point trying to understand anything about cameras or eyes or visual perception until you have a firm grasp of the concepts in this and the pinhole paper, and until you realize why they are the fundamentals of optics. The contents of these two papers are rare, but true. You must spend the time here to appreciate these concepts because it will get “curiouser and curiouser” as we work up to visual perception and consciousness itself!

3. What Is Focus?

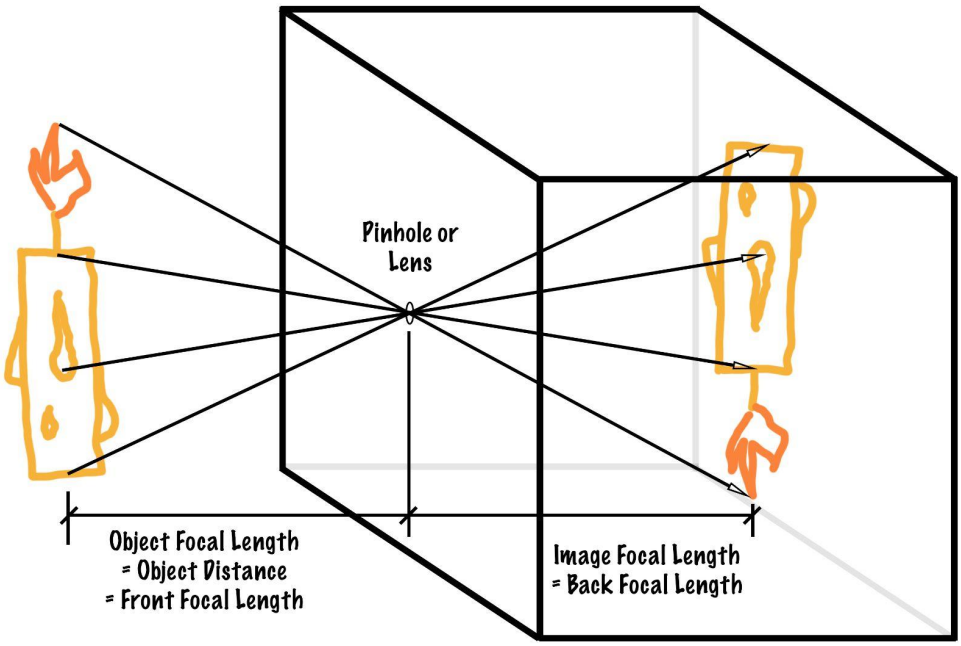
Brad Caldwell

You already kind of learned what **focus** is in lesson two; nevertheless we cover it again here in more depth. Logically, the term **“focus”** should only ever refer to *the bending of a light cone from a pixel in the world back into a point upon reaching the sensor*. But as is often the case, this actual definition is hardly ever encountered, and everything else that *isn’t* focus gets mislabeled as being focus! So let’s familiarize ourselves with a couple other concepts so as to know what they are referring to.

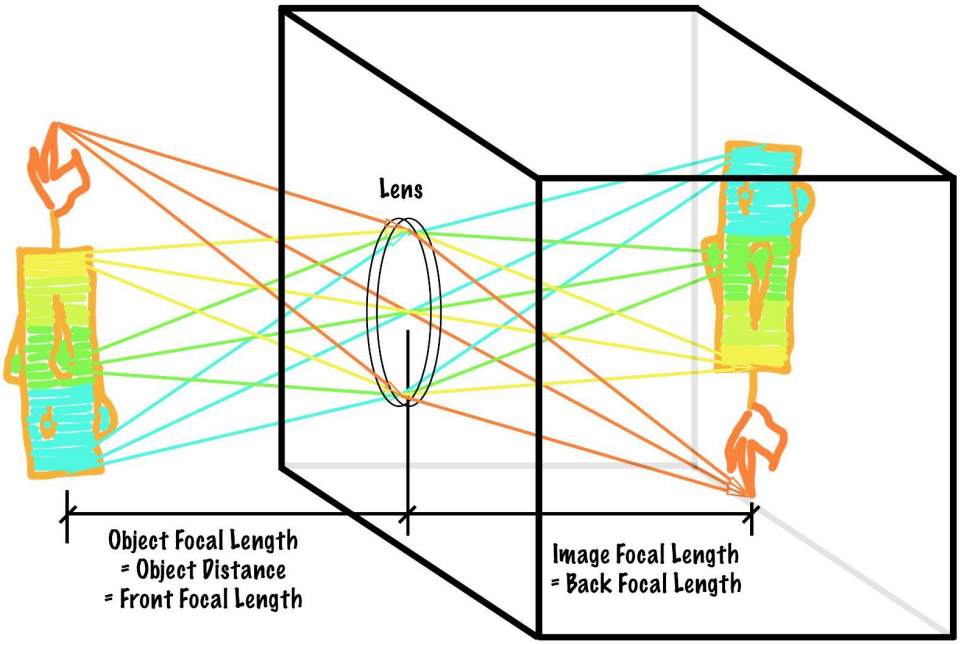
“Image focal length” or **“back focal length”** is the term *conventionally used to describe the distance from a lens to a sensor*. *The effect of varying this is to change the percentage of the sphere of perception (sphere of the world around you) included within the image on the sensor*. Increasing “focal length” means the image will be bigger upon landing on the sensor, and thus will crop the image, meaning that a smaller solid angle portion of the hemisphere is on the image. It’s a means of zoom. But if you don’t change the focal power of your lens along with it (“refocus”), it would result in the focal object being out of focus.

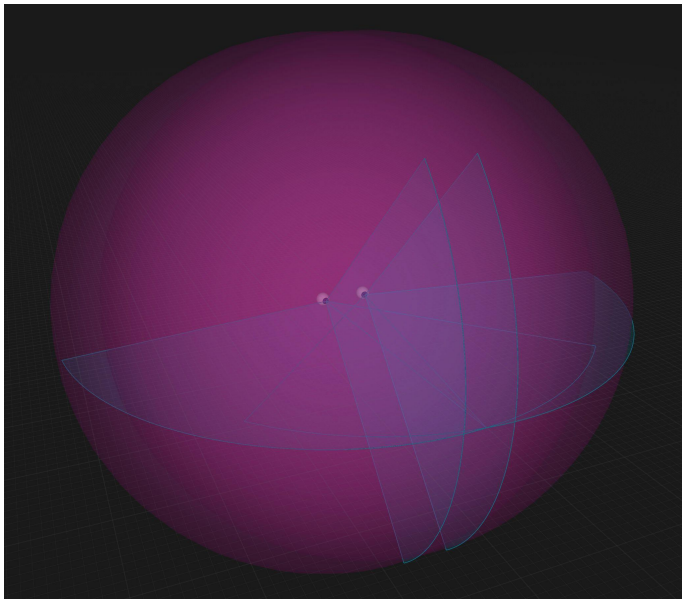
“Object focal length,” “object distance,” or **“front focal length”** *describes the distance from a lens to the object you want to be in focus*. Focusing on far objects requires less focusing power; focusing on near objects requires a lot of focusing power. *If you need to focus on something close, you can push the lens away from the sensor so that there is more distance to work with in bending the cone back to a point.*

DSLR lens-makers use lots of lens elements, but tend to define optics very confusingly.



The below visual gives a good idea of both what **focus** is and what a **lens** does. The graphic is in 2D, but if you use your imagination, you can envision cones in 3D. As you can see, every point in the world sends off light in every direction. This makes an expanding sphere. But at the point of crossing the lens, you're only sampling a very small area of the sphere shell, so the shape appears as a cone. The goal of the lens is to bend that cone back into a single clear point right as it reaches the sensor.



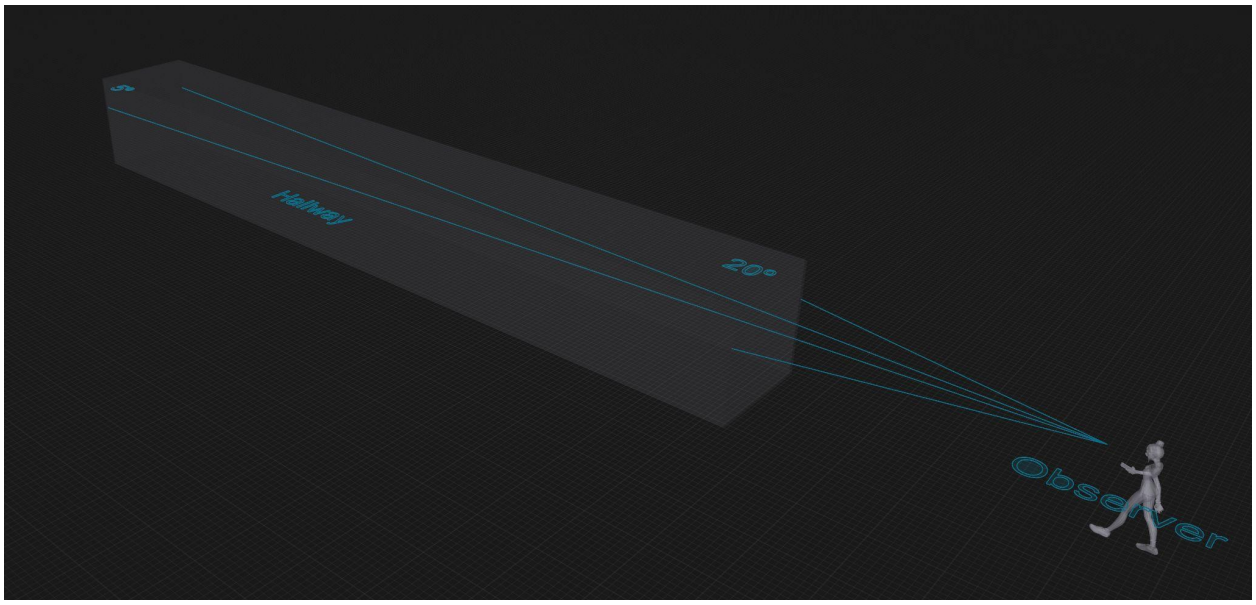


The human can see roughly 180º horizontal by roughly 100º vertical, but the fovea can discriminate 1/60th of a single degree. That which is in foveal view also has clues as to depth because it knows what focus the lens is currently using, the degree of vergence, and the degree of pupillary constriction dictated to help see close up focal objects. Images lose depth and this depth component must be rebuilt by the brain prior to visual perception. The brain may keep tabs on **the eye's current focal power** alongside incoming vision to *help specify depth in converting foveal pixels back to voxels*.

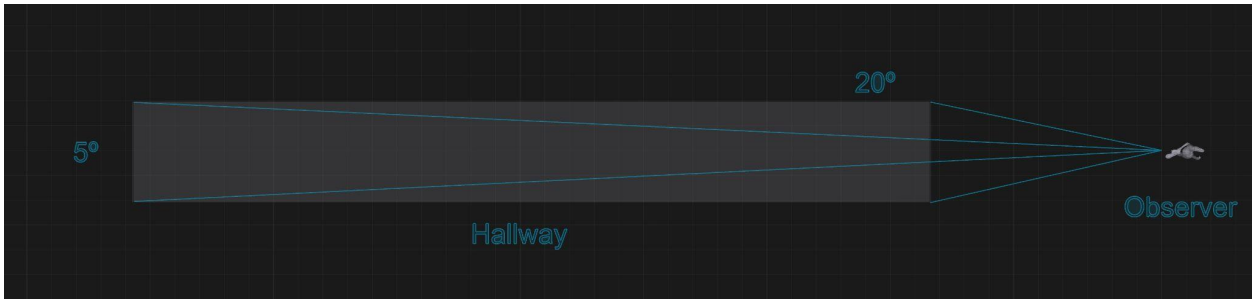
4. Perspective Vs. Orthographic Images.

Brad Caldwell

Perspective images *are the ones we see everyday*. They are the images our eyes see. They are the images our cameras take. They warp 3D space a certain specific way to make it fit onto a flat planar, or hemispheric-shell, 2D surface. The way they do this is to shrink more distant objects. If you imagine looking down a long hallway, at the far end of the hallway its width takes up less “theta” angular space than the width at its close end, even though you know it is the same width at both locations. Taking an image by geometrical fact requires this compression of the size of farther objects.

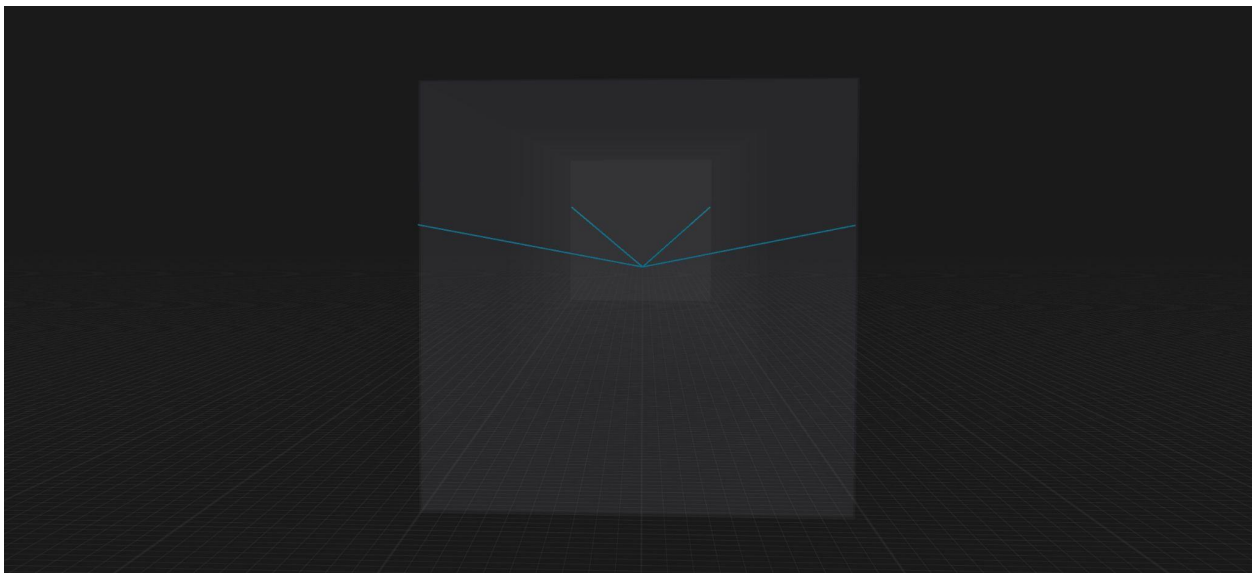


Let’s examine this idea with the use of an orthographic image. **Orthographic images** *are like you could place a 1D eye that just looks straight ahead at every pixel in the image*. You get an image in which all the distances are true to life, but you can’t see depth well. In the image below, an orthographic (“pure 2D”) image is taken of an observer looking down a long hallway. As you can see, to her, the front of the hallway looks wider (20º of theta) than the back (5º of theta!).

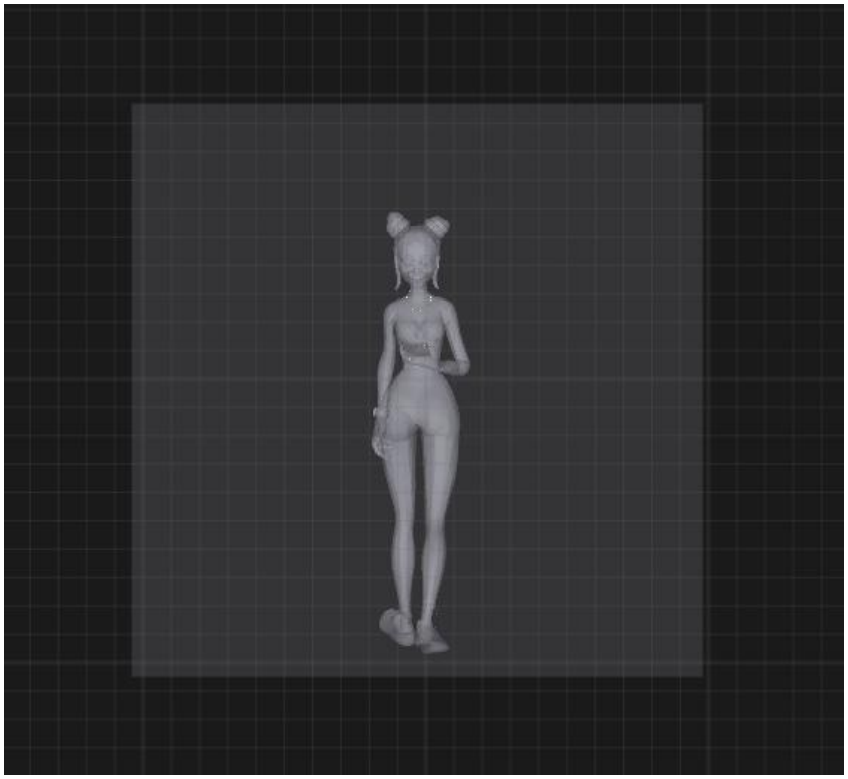


In the image below, you have a “point of view” perspective image of what the woman looking down the hall would see. As you notice, for visual perception and consciousness, *we never*

leave the perspective image! The back end of the hall looks narrower, at least in terms of the angular theta width that it takes up in our perceptual space.



If we were able to make consciousness go into orthographic mode, then we'd see the hallway like below (note that the front and back of the hallway cannot be made out orthographically).



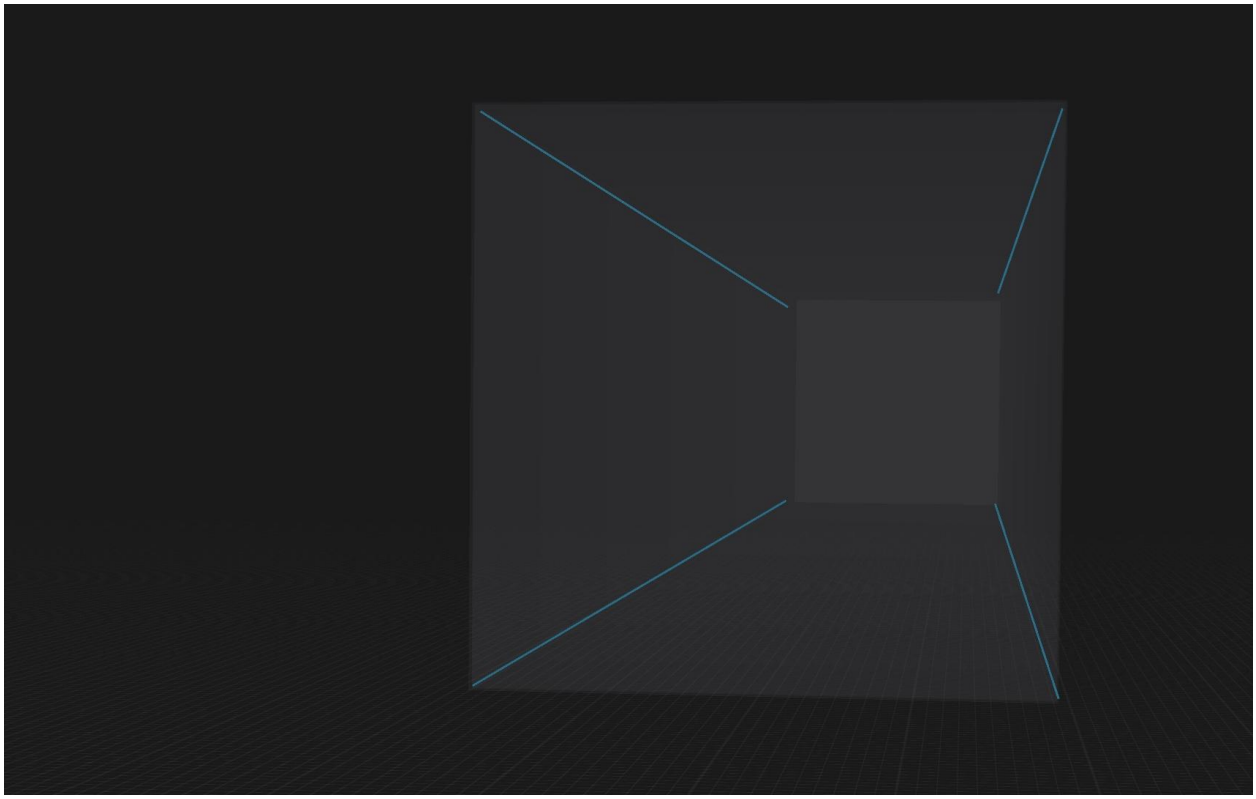
A set of architectural drawings usually includes both several orthographic drawings (elevation view, bird's eye view) as well as a perspective drawing on the front cover to show a real-life 3D angled perspective that shows depth as if you were there in real life. The orthographic drawings are good though, but are confined to a single 2D plane. Perspective images can reveal things at differing depths and give a sense of 3D. Nevertheless, a perspective image is still fundamentally a 2D image.

5. “Casting” Perspective Images to 3D.

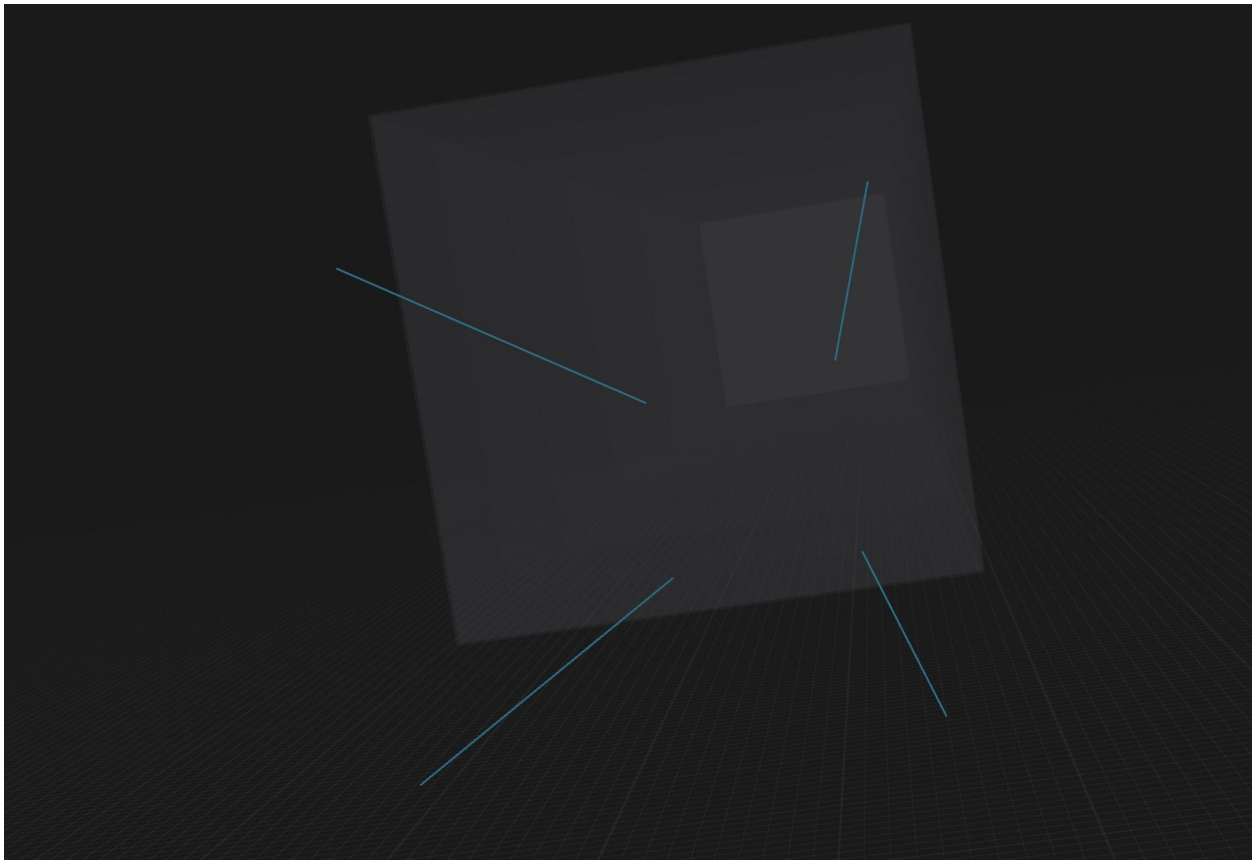
Brad Caldwell

We reach a point now where we enter speculative territory. The study of consciousness requires us to go places where no one has any certainty yet. We will never understand visual perception if we do not admit and look into the unknown areas. When we discussed perspective images in the last lesson, we noted that consciousness uses perspective, not orthographic, images. Both of these are 2D images, although perspective images let you deduce more depth details. But fundamentally, the perspective image is still just 2D. If a computer were to look at the perspective image, it would see no depth at all, and would see the wall-ceiling lines going down the hallway as completely within an x-y plane, not, as they are in reality, angled partly into the depth of space.

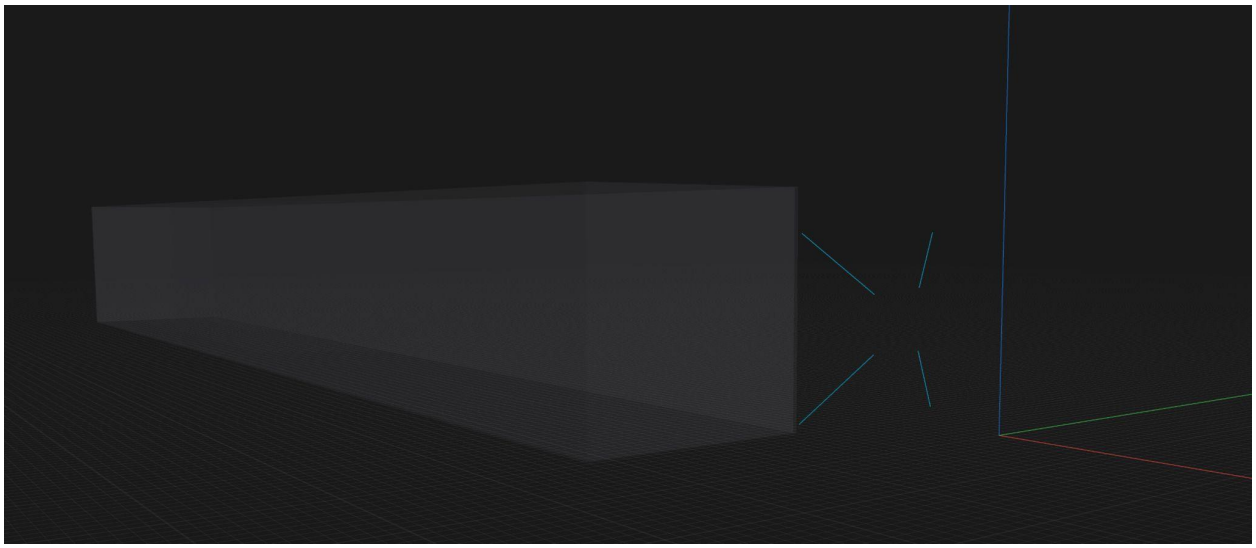
In the first image, it appears that maybe the computer understands the 3D vector orientation of the blue wall-ceiling and wall-floor lines of the hallway.



But as I rotate around, I see the computer did not understand at all, because it is in fact just a 2D perspective image, so how would it have understood depth?



Finally, as I continue to rotate, I see exactly how 2D every line on a seemingly 3D perspective is!



So, this just proves that even though our retinas capture perspective images that we think look 3D enough, if there weren't depth info for every pixel, we would not have the 3D experience of consciousness that we obviously do have (unlike computers).

So, the question is, how does the brain “cast” pixels to certain depths so as to create an understanding of 3D from the warped perspective images that are all we know?

At a bare minimum, there must be understanding of the proper depth for every pixel. But to really make it pop, it may have to utilize a ring order of which pixel when to show the depth relationship between pixels.

Just as an old cathode ray style TV had to shoot electrons across in a row, then down a row, down a row, etc., until the whole screen was complete; so also consciousness may have to refresh the 3D screen by painting in various orderly routines. One such proposed routine of refreshing is the ring which is able to 3D rotate in perceptual space, and which occurs roughly 2-12 Hz. Much of vision obviously appears to be refreshed faster than this, and there may be a faster 40+ Hz ring or some other method of order for going around to each pixel and updating, but following an order that helps to reveal the held understanding of what depth each pixel is at, so as to reveal you are refreshing a 3D scene.

Another method may be refreshing by 3D “skins” of the boundaries between skin and air, or table and air, or ground and air. The pixels of those skins would be refreshed in a sudden burst to draw attention to the form. Pixels of the translucent backs of items could also be fired at the same time to show knowledge of the back side form of items.

Another seemingly outlandish proposition is that the rings could be all there is (perception/consciousness wise), occurring at 2-12 Hz, and at 4-6 stops along the ring, a “star cluster” is placed over a short duration of time, the patterning contained therein sufficient to “understand” a little backwards and forwards in time the short segment of visual 3D movie intended from that star cluster stop to the next star cluster.

Another possibility is that 3D “zigs and zags” in the ring encode the details for the colors and locations of voxels of consciousness.

At the end of the day, everything around you is the perceptual world. The grass, the roads, other people, your bed, your body — it is all a fabrication of your brain. You’ve never seen the physical world, although the perceptual model is extremely accurate, so you have a good idea of what it is like in form. The real world probably isn’t even at the same location as the perceptual world, for the perceptual world is a “meaning world” that is understood from the codes of neurons, in distinction to the physical world that has actual space and actual matter.

If you take an anesthetic, you will see that everything (your body included) dissolves away. That’s because those things are all a story of meaning told by the brain to guide it. You do have a physical body, it’s just not your perceptual body that your brain paints all the time. There is a real physical world out there, just not the one that you’re always looking at.

6. What Is Stereoscopic Vision?

Brad Caldwell

As a recap of perceptual optics, we showed:

(1) Sampling an image from the physical universe involves the fact that images already exist in point form at every location, then using a pinhole and black container to let the inertia of the light invert and spread out into an image. This loses the depth (radius) information.

(2) Lenses allow more light in, and also focus each muddy “cone” of light back to a point by the time it reaches the back of the camera (or eyeball). This can only be done precisely for a single focal surface, which we showed to be a spherical shell in shape; other radius focal sphere shells will have more or less “circle of confusion” since their associated cone of light had a greater or lesser solid angle and needed more or less focusing power to be applied than the cones of light from the focal sphere shell.

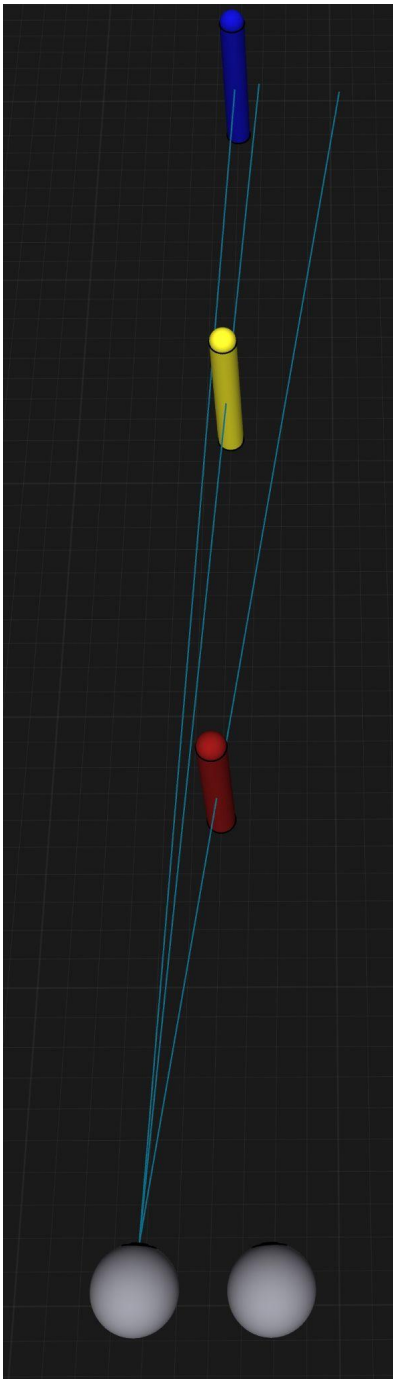
(3) Focus means bending a cone of light from a specific radius sphere shell (or hemisphere shell, or cone shell, depending on level of zoom) back to a point at the precise distance of where the sensor is. In theory, perhaps a hemisphere sensor would appear to be a good idea, as the individual tiny sensors would be oriented parallel to each one’s incoming light, and 3D vector orientation could be associated with each beam. The eyeball uses something close to this, although spherical, meaning each cone and rod has to bend away from the plane of its planting a little to align with the incoming light (like trees growing towards light). Whether the eye uses the pixel layout of the cones, or the 3D vector associated with each cone, is (to me) as-yet unknown, although I do think we should remain open to the idea that it could use the 3D vectors rather than the pixel layout, as the layout of neurons in the eye or brain is never spatially precise enough to have meaning that would correspond to the precision of the conscious experience.

(4) What a single eyeball sees (in the perceptual realm) is definitely a perspective image, even though we understand depth. The *width* of the far end of a hallway always subtends less theta angular perceptual spread (say, 5º) than the width of the close end (say 20º), even though they are the same width (4’); and the *height* of the far end of a hallway always subtends less phi angular spread (say, 10º) than the height of the close end (say, 40º). This is all just a fact of sampling the 3D universe for an image, and occurs with both cameras and eyeballs, as the light rays from farther widths subtend smaller angles than the light rays from closer widths of the same value (4’). Notice that a translucent box in perceptual space would appear as a perspective translucent cube shell and that the increments of focal surface would appear as translucent sphere shells.

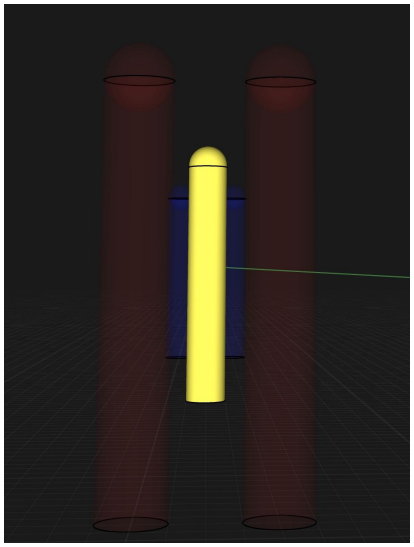
(5) A 2D perspective image, while looking fairly “depthy,” is fundamentally just 2D. To make it 3D, each point must be understood at its proper radial depth. In other words, the pupil of the eye (or the lens of a camera) is the sampling point in the physical universe, and each 3D vector light ray must place its causing point in the physical universe “cast” to the right depth in the image, to reconstruct an understanding of a 3D world from a 2D perspective image. The pupil (or lens) is the centroid of a “sphereset” of translucent sphere-shells (and if we add another aspect for ease in displaying commonly encountered walls and tables, etc.), an overlaid “cubeset” of translucent cube-shells.

Anyway, the point here was that the brain somehow “casts” each pixel into a “voxel” (3D pixel) prior to experience (at least for the majority who have learned to understand 3D depth from vision). There is this whole 3D scene that each eye contributes.

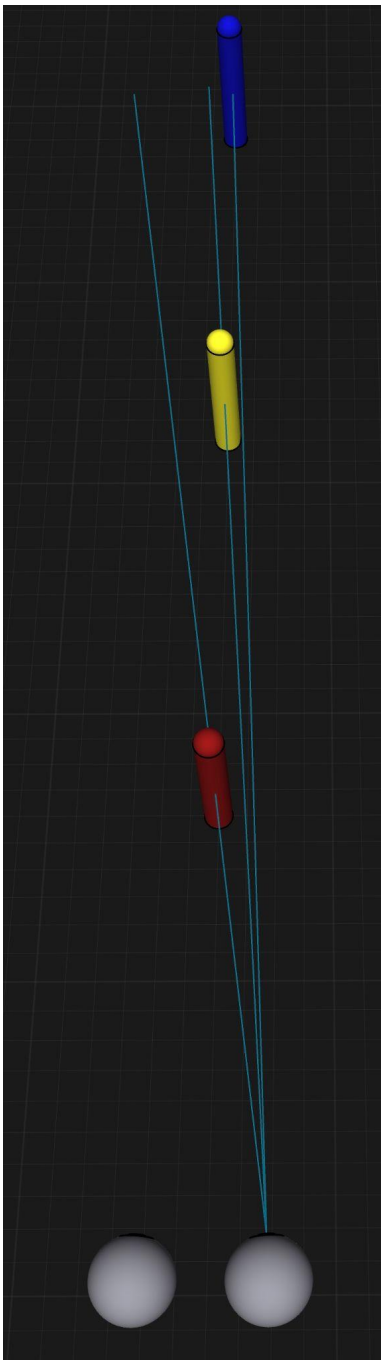
(6) But, once we introduce the fact that we have two eyes, it may become useful to consider the *foveal focal voxel* to be the center of rebuilding 3D rather than the eyes, for it is the fixed centerpoint that marries the two overlaid 3D scenes. Or, keep the centroid of each individual eye’s contribution, but build a “third” “bank” (overlay of sphere-shells/cube-shells) with its centroid at the *foveal focal voxel* (the thing visually focused upon). Let’s first look at how the two eye’s contributions fit together looking at a medium focal depth object (crayon), and what it does to a closer and farther crayon.



So, the left eye sees the closer red crayon to the right of the focal yellow crayon, and the farther blue crayon to the left of the focal yellow crayon. But, the right eye sees the opposite of this! It sees the closer red crayon to the left of the yellow, and the farther blue crayon to the right of the yellow crayon.

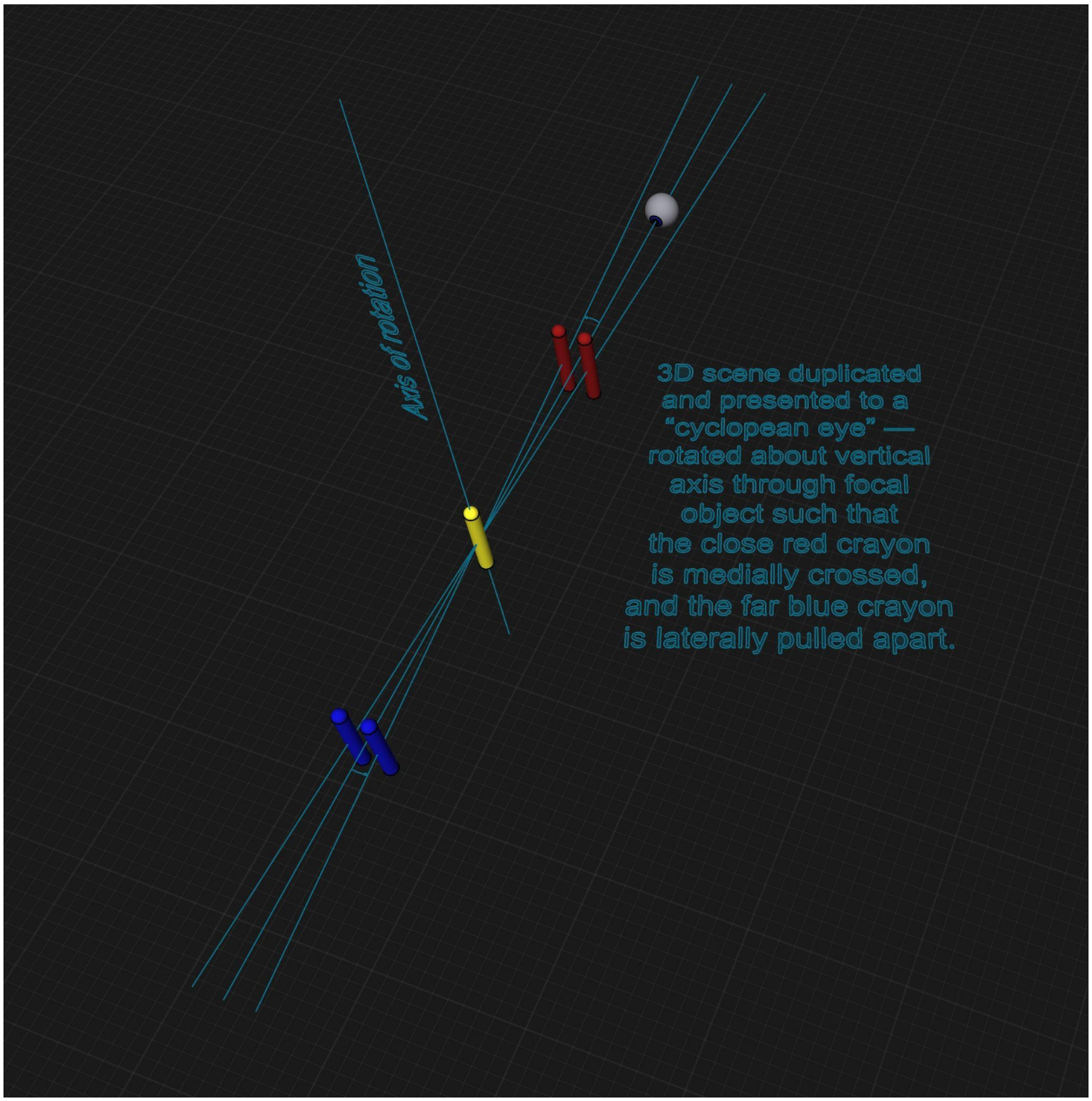


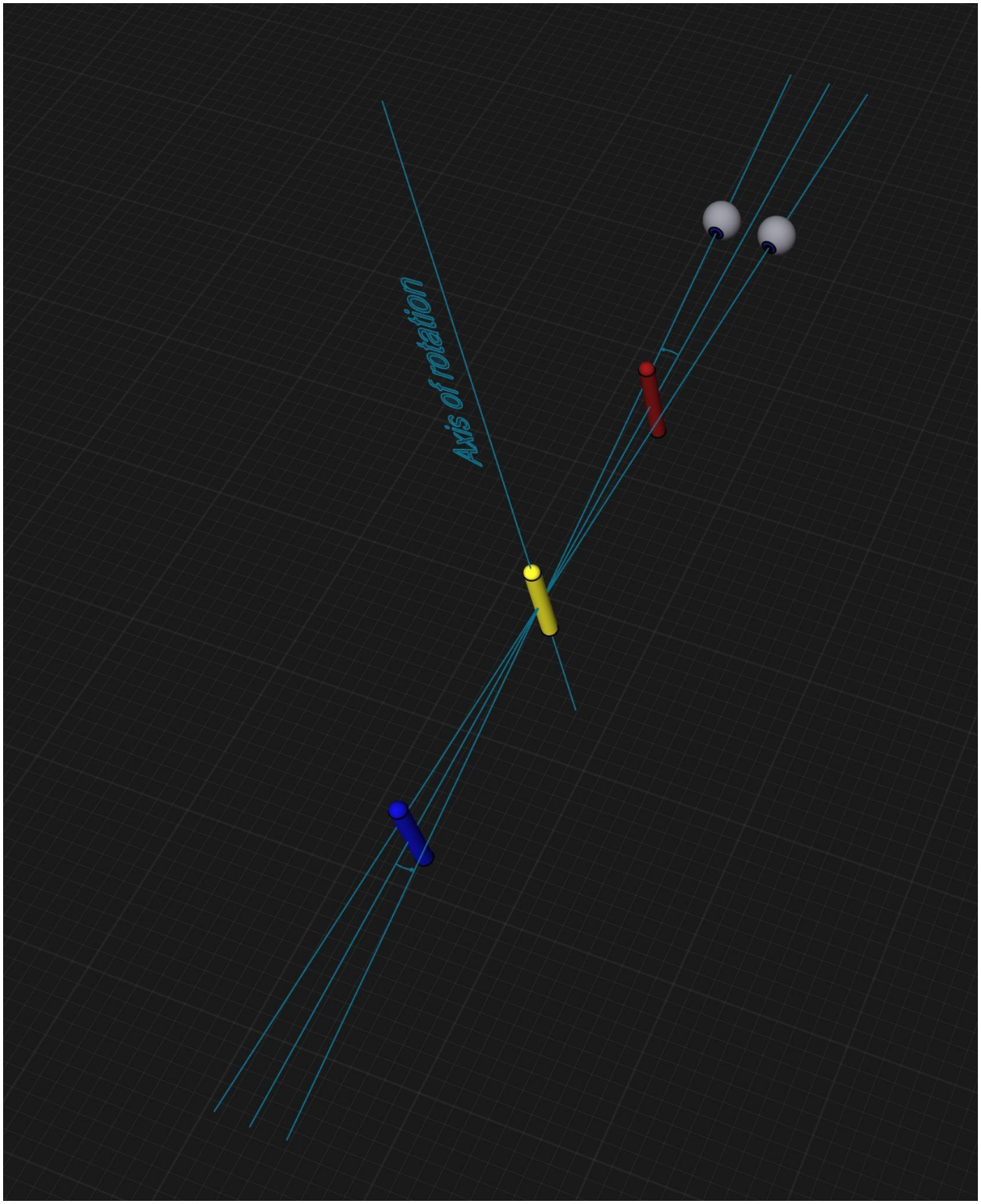
As a result, the red crayon duplicates and crosses medially (“crossed disparity”) and the blue crayon duplicates and pulls apart laterally (“uncrossed disparity”).



In reality, the duplication is of the entire scene (since each eye sees and entire 3D scene at the level of brain perceptual vision), with the centering axis of aligning the two being a vertical axis through the focal yellow crayon. The red and blue crayons become “ghosts” since they each only contribute 50% to the scene but the yellow crayon is seen as fully solid as it is drawn by both eye’s contributions.

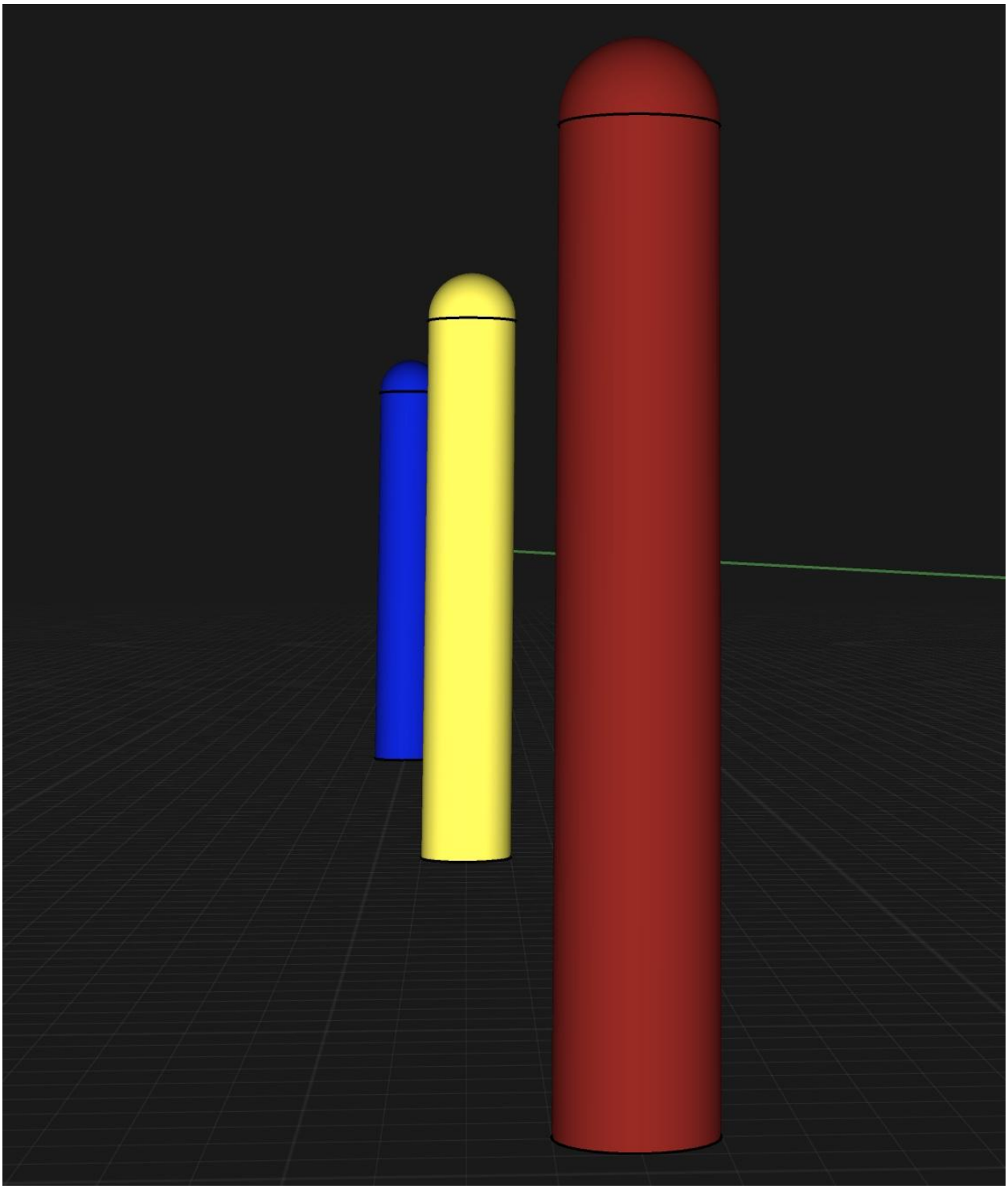
It seems nearly equivalent to consider ourselves as having a cyclopean (single, center) perceptual eyeball, with the two 3D copies of the perceptual world rotated about a vertical axis going through the focal object the amount that would offset the cyclopean eye into two copies that match the distance between our eyes (roughly 2”).



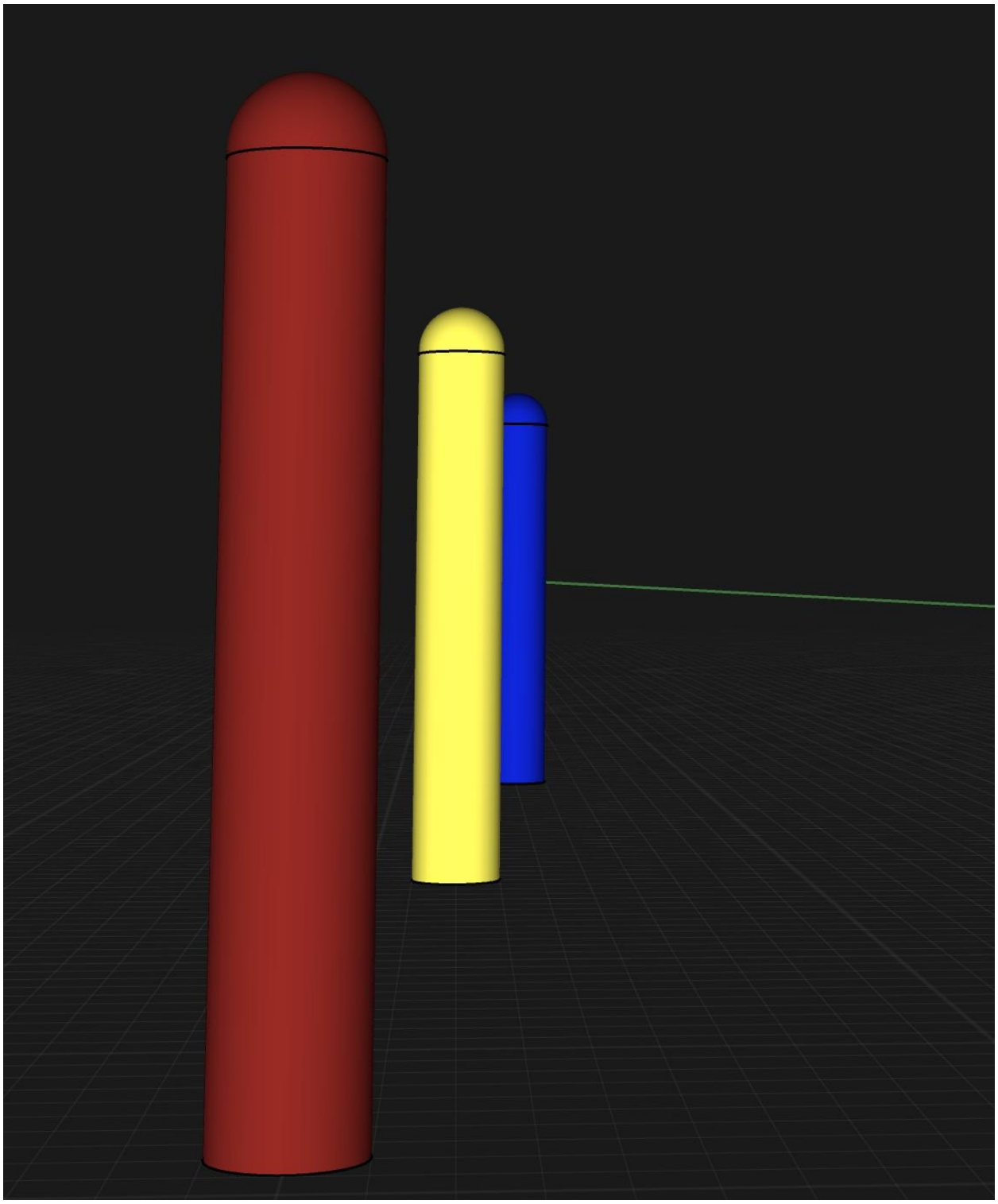


Here, let’s take another look at it. Below are what the left and right eyes see individually. Now it should make sense why together they create the “stereoscopic vision” of the ghost image a few pages previously shown.

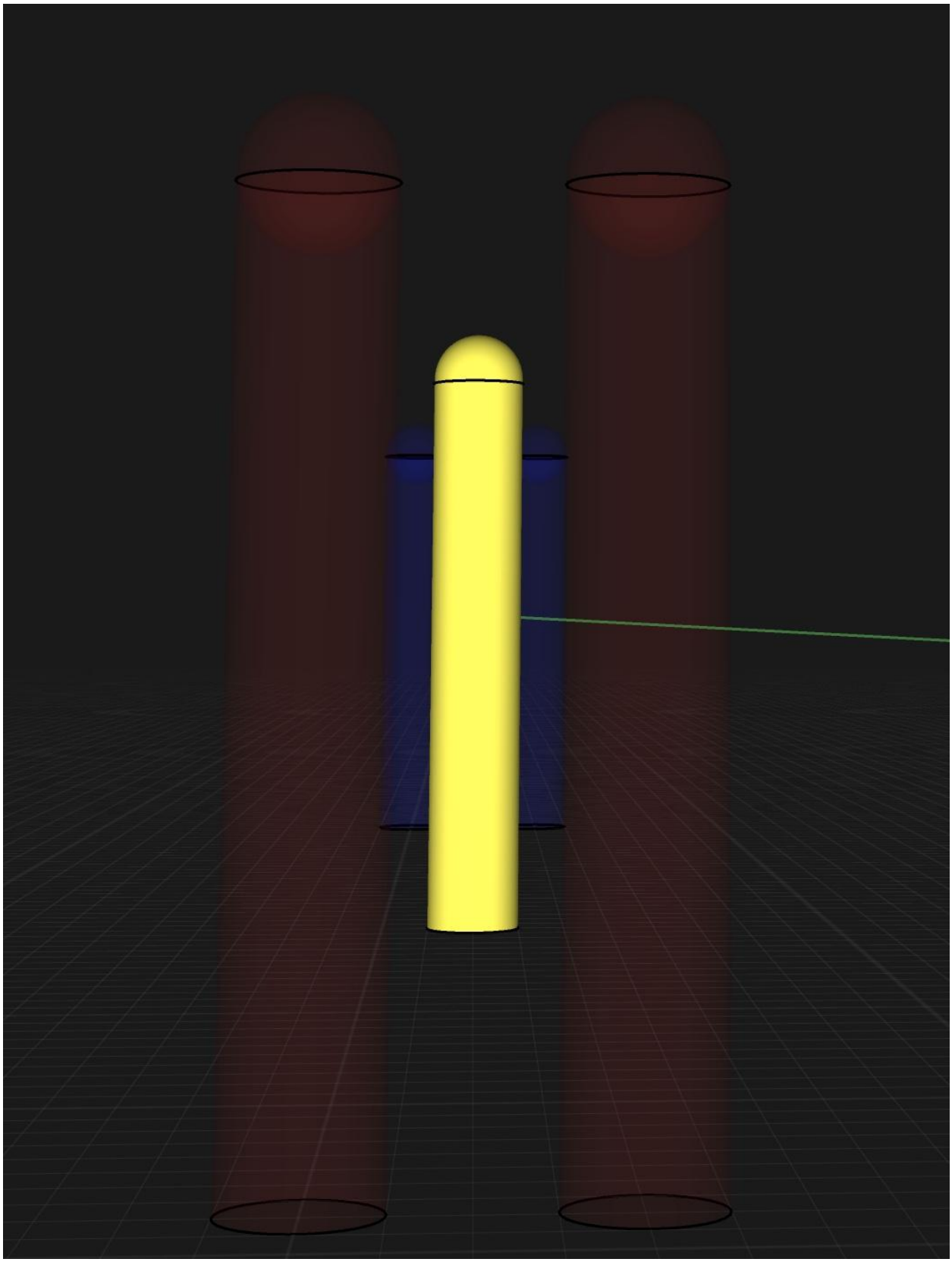
Left eye sees:



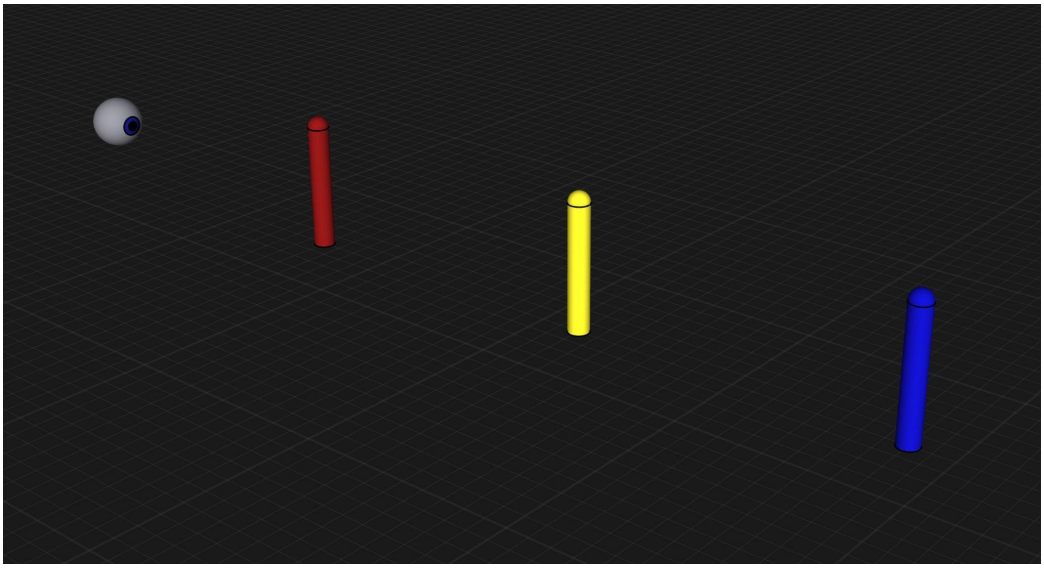
Right eye sees:

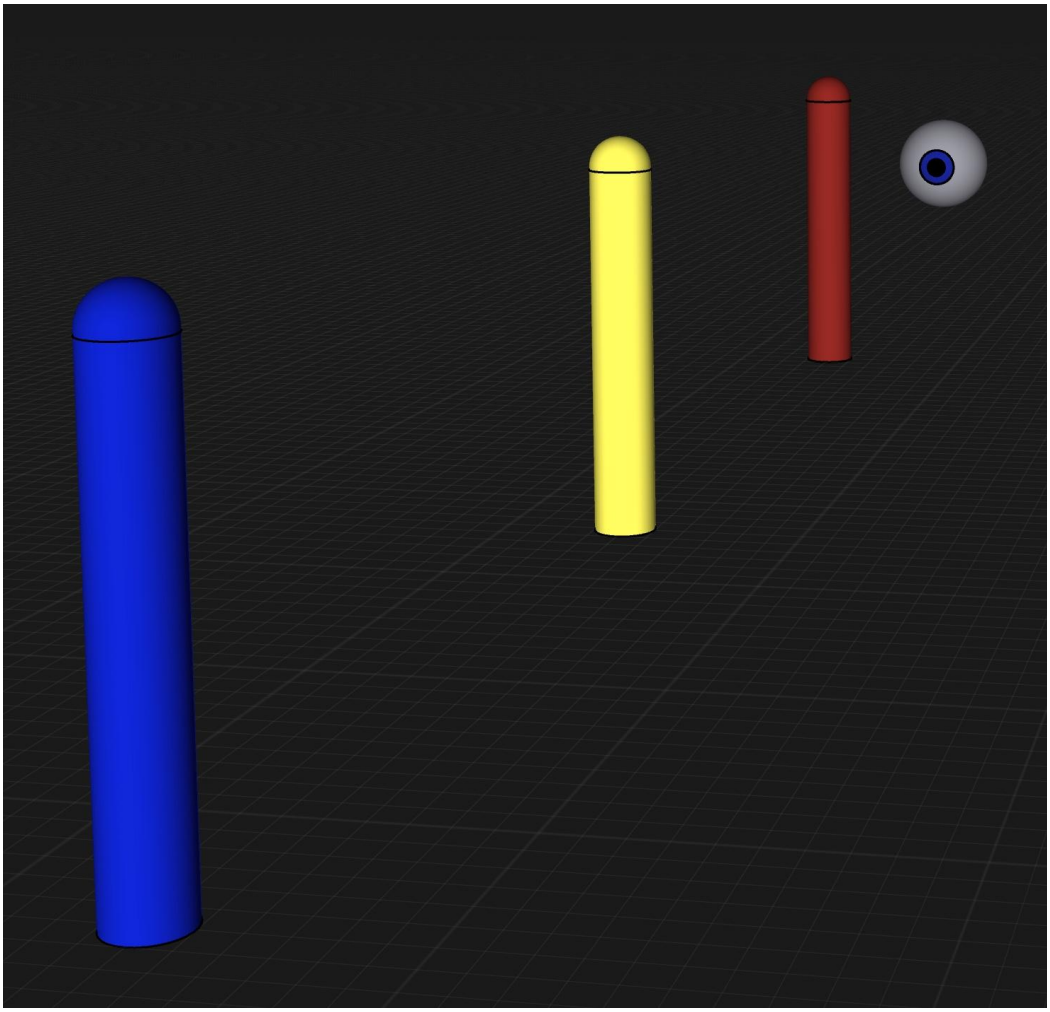


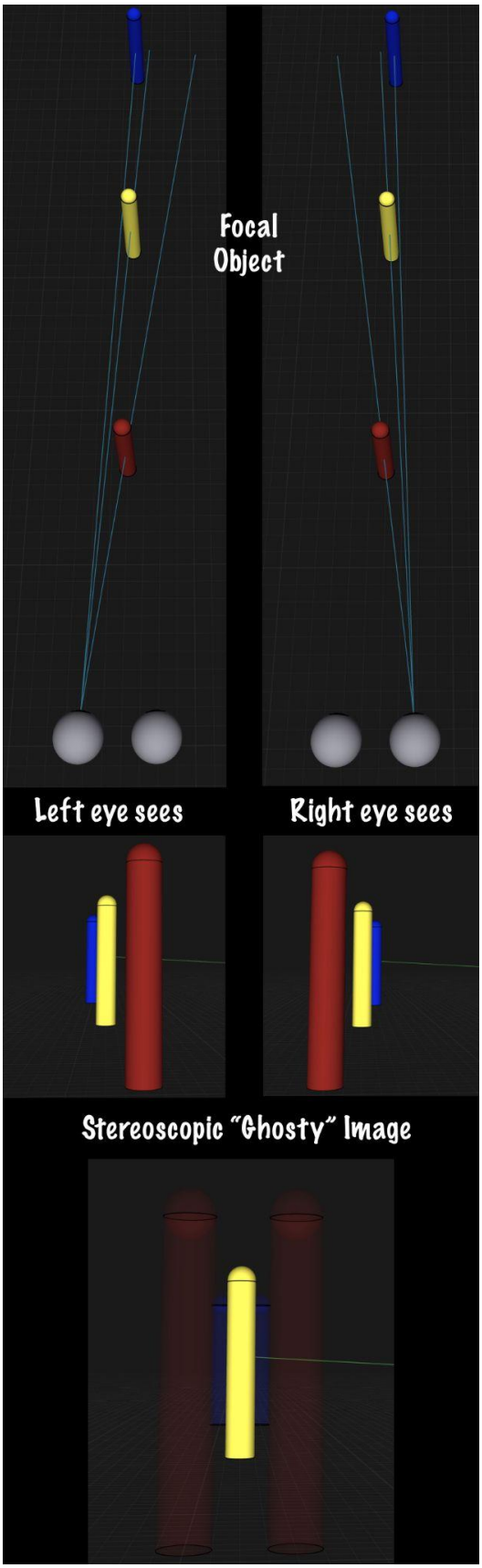
“Ghost” stereoscopic vision image, provided again:



If our physical eyes were one eyeball rather than two (and centered), we would only be able to see the red crayon, as it would block the view of the yellow and blue crayons.







In summary, for perceptual stereoscopic vision, two perspective images that had already been "cast" to 3D are married together (overlaid) with the focal object serving as the line through which a vertical "rotation" axis goes. There needn't be any actual rotation — rather, the rotation was used to illustrate how the two eyes double objects that are closer or farther than the focal object. It is *as is* the whole 3D world had been duplicated and swiveled about that vertical focal axis by the number of degrees needed (at that depth) to pivot a perceptual cyclopean eye out to the location of the two physical (technically, perceptual) eyeballs.

There seems to be an equivalence between a cyclopean eye looking at a doubled and properly crossed (rotated) world, and two eyeballs being rotated out from center along a 1" arc each, arcing about the foveal vertical axis, and their two scenes being overlaid at their angles. Either way, it seems that the same perceptual phenomenon would result — the "ghost" stereoscopic 3D-cast double perspective image.

As a final note, the axis is vertical only if your head is vertical. If you tilted it to the side, the axis would need to tilt too. This should be straightforward.

In the image to the left, this lesson has been condensed to show how the left eye and right eye each contribute a perspective image that has been "cast" to 3D, and these two 3D realms are overlaid and (since they differ) agree to center upon the focal object, letting the crossed and uncrossed disparities displace into "ghosty" 3D objects of less opacity than the focal object. This also serves to draw attention to the focal object and move attention away from the things closer and farther than the focal distance.

Since two eyes are looking at the yellow crayon, perception allows you to "look around" the sides of the crayon slightly (not depicted in this image, and not really capable of being depicted in a single 2D image).

7. Ten Trillion Voxels in Perceptual Space?

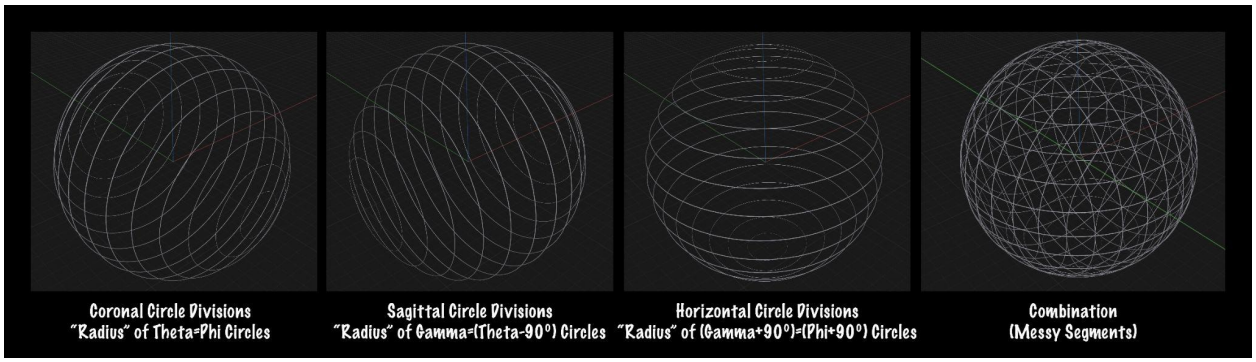
Brad Caldwell

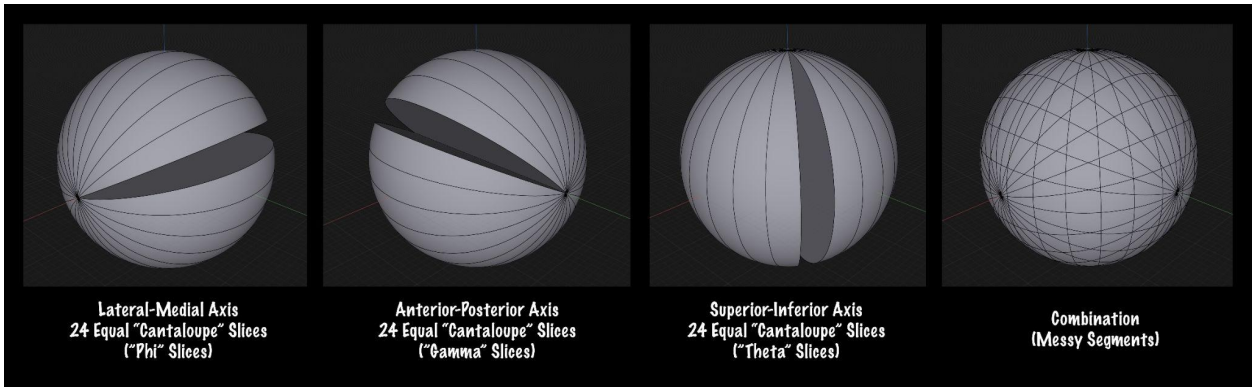
A main takeaway from lesson six is that it may be that the brain builds 3D perceptual space from the foveal object outwards, rather than starting with painting the body or eyeball. Or, it may go further and simply tether your foveal object voxel or the back of your body to the centroid of perceptual space. This would allow some movement, some give, some “hysteresis.” I am not sure which is true, or if something slightly different is true.

Nevertheless, as with lesson five, if we are ever going to learn how consciousness (and visual perception) works, we must go forth into uncertain territory boldly.

At a minimum, geometry dictates that there is a sphere of environment around us, perceptually. That is, there are 360º of theta (yaw), 360º of phi (pitch), and 360º of gamma (roll). Or, more concisely, there are 41,253 square degrees which the solid angle of the sphere of perception can be divided into, like a bunch of tiny cones forming a Christmas star cluster ornament.

The surface of a sphere is a funny thing. It does not allow a large number of points to be evenly distributed over its surface. Rather, one must place points at hexagon intervals while throwing in perhaps twelve pentagons. So any way you divide the space up, it won’t be nice and neat.





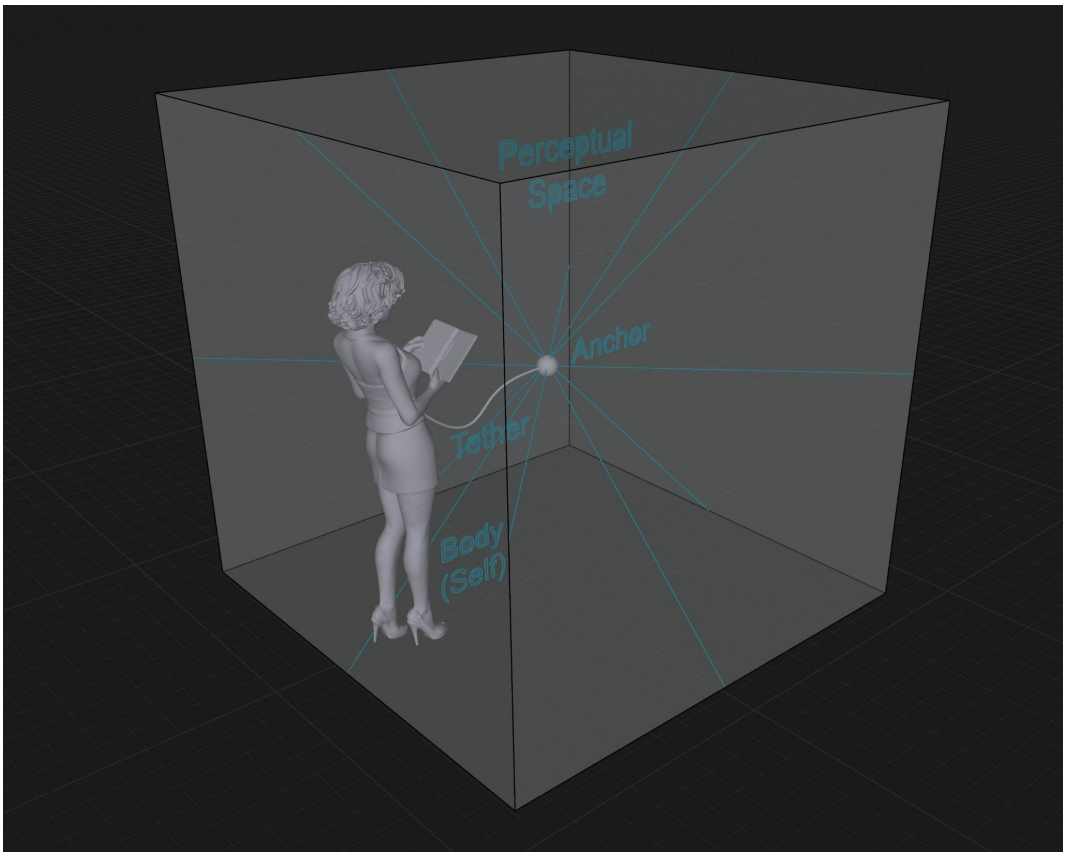
I think it is true that perceptual space is fixed. Or, stated another way, the “TV screen” that shows the “story” of consciousness is fixed and never moves.

What I am unsure of is whether your foveal fixation point stays fixed to the center of this 3D “TV screen” your entire life. If it did, then when you move your physical eyes, the brain would paint a changed scene that “kicks out” the older stuff by the distance between the old voxel and the new voxel. Perhaps a separate “bank” is given (by the greater hippocampal region) to the (perceptual) “world” to ensure it is always “held together” as an immutable 3D raised relief topo skin.

But it also seems possible to me that where you are looking (the foveal fixation voxel) is allowed to have a little slack (like a steering wheel that must be turned a certain amount left before you can recover from it having been turned right) with a different centroid point of the most fundamental perception bank.

In favor of the foveal fixation point always being by definition the centroid of the 3D “TV screen” of consciousness, the same neurons in the visual cortex are always called upon to represent foveal stimulus.

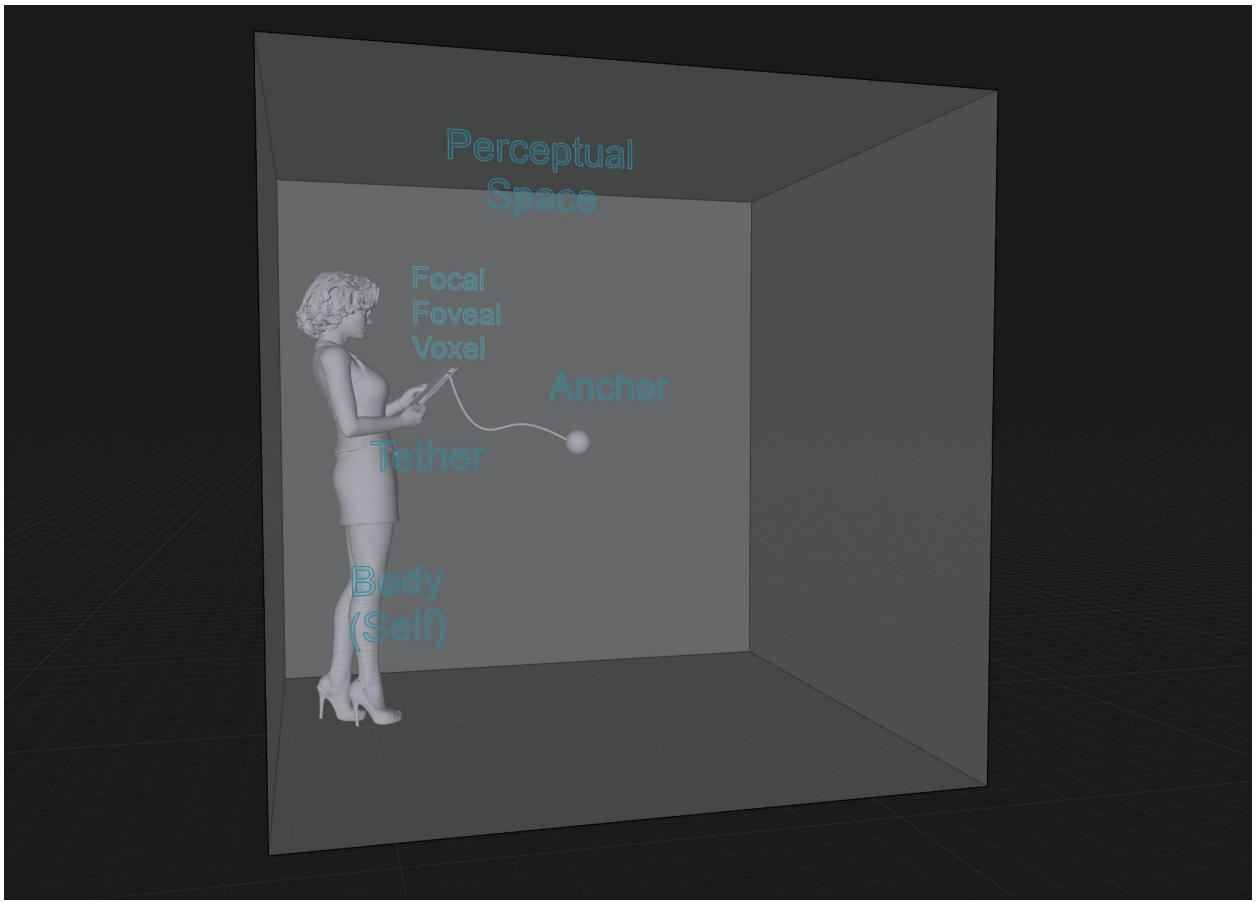
In favor of having a “random” centroid that never moves, and simply tethering your foveal fixation point or the back of your body to this centroid, it would allow for perhaps more sense of stability in the world as you dart your fovea (eyes) back and forth.



In the image to the left, the “tether” of the perceptual body (the back in particular) to the perceptual 3D “screen” is depicted. Some hysteresis in this model is allowed, although movements greater than three feet trigger a repainting of

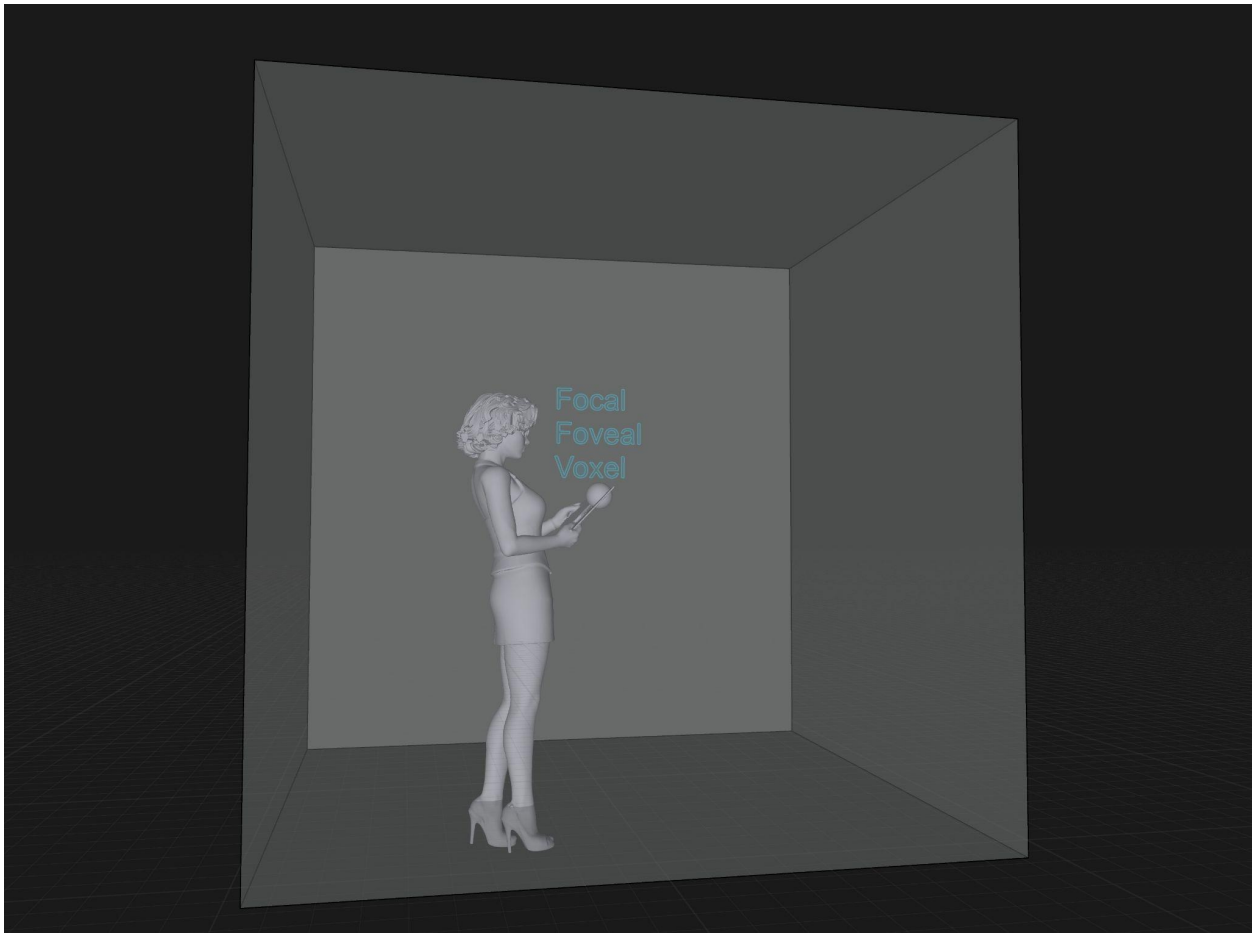
the perceptual world and a recentering of that perceptual world relative to the “anchor.”

In the next image, the tether is attached to the focal foveal voxel. Of course, this “tether” is a functional, not a spatial or actual, one. It is a “rule” of “keep focal foveal voxel close to “centroid anchor” of 3D “consciousness TV screen.” Or a rule of “keep back of perceptual human tethered close to the centroid of the 3D screen.” Anyway, below is the tether attached to focal foveal voxel.

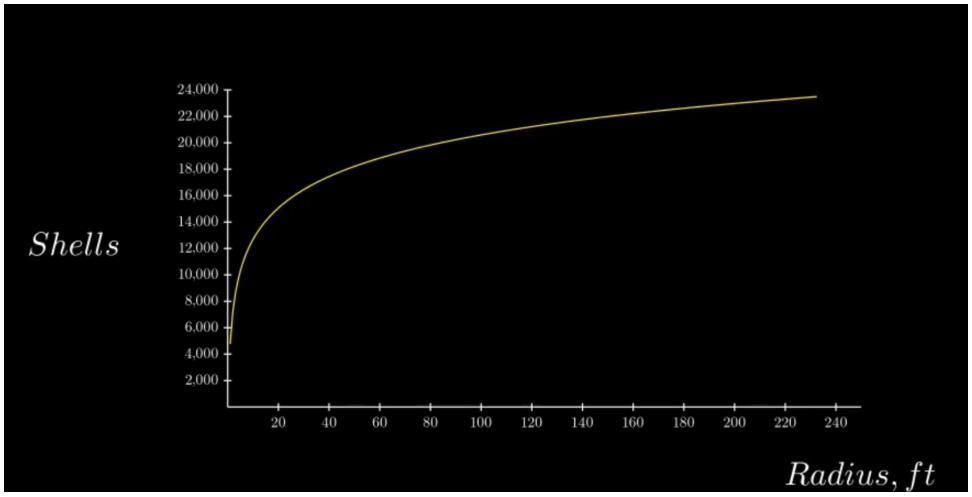


Finally, I will show the anchor being eidetic with the focal foveal voxel, a third exploratory possibility (see image next page).

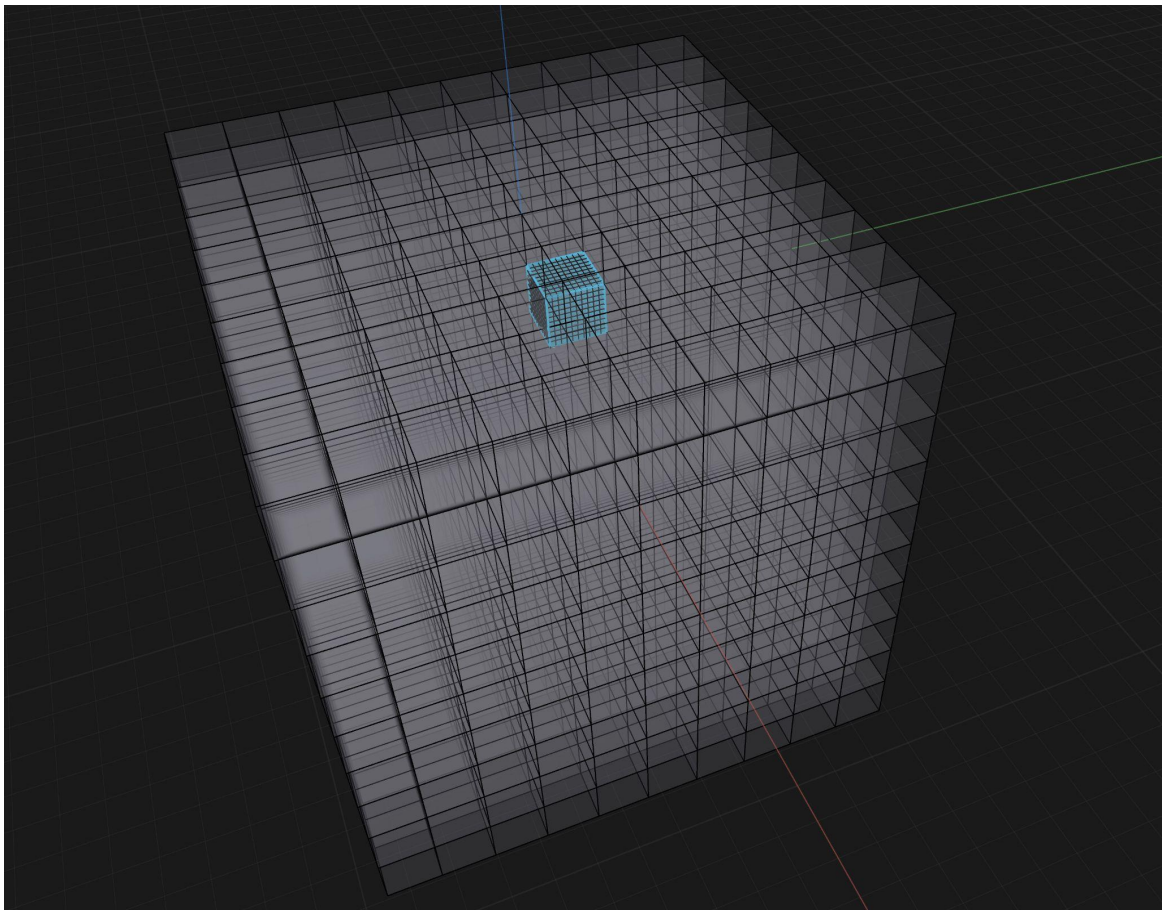
If the focal foveal voxel is indeed the center of the 3D perceptual “screen,” then the (perceptual) world gets “man-handled” and spun around like a spider and stretched in all sorts of ways, just by the geometrical fact that farther items are compressed spatially when made into a perspective image. If this is true, the brain probably devotes another “copy” of the “bank” to the perceptual world so as to keep it adhered together as a single unified whole. The “bank” seems similar to, or eidetic with, the 3D perceptual “screen.”



In any event, we know the fovea has incredible resolution, being able to discriminate a single arc-minute (one-sixtieth of a degree of theta/phi/gamma). That means it can discriminate 50 million pixels in the human's typical field of view, if allowed to move around. This is roughly ¼ of perceptual space or ¼ of the entire sphere solid angle surrounding you. If we assume similar capabilities of depth resolution (at 2', we could discern well under a millimeter of width [theta] or height [phi]), we get a distribution of "sphere shells" as depicted in the figure on the next page. Assuming about 100 feet of depth is in view, that would be roughly 20,000 steps of depth discrimination, or 1 trillion voxels. For there, it is each enough to allow the head to rotate around and up and down so as to allow the fovea to look in any direction outward from the center of a sphere. If we allow a larger scene (say, 10,000 feet of depth), we could theoretically discriminate 36,420 steps of depth, yielding nearly 7.5 trillion voxels without ever moving (translating)! I'd say that's pretty close to ten trillion!



That gets us to another question. Surely the brain is smart and saves resources. If you only need to precisely know the locations of voxels of one area, it could use a bank with 10x10x10 subdivisions of a 10' cube, then select the right cube, and subdivide it into 10x10x10 voxels, then find the right voxel there, and subdivide it into 10x10x10 smaller voxels. The "basic location" of objects could be given first, then "honed in" to greater and greater resolution until you reach that maximal single arc-minute of theta/phi (/radial depth?) resolution.



The image on the previous page shows *two* orders of magnitude of 10x10x10 division of a cube of perceptual space. The image below shows *three* orders of magnitude of 10x10x10 division. Theoretically, this could go on as much as desired. This is theorized to be the 3D “screen” on which consciousness is painted (includes perceptual world and perceptual body).