Perceptual Optics — 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.

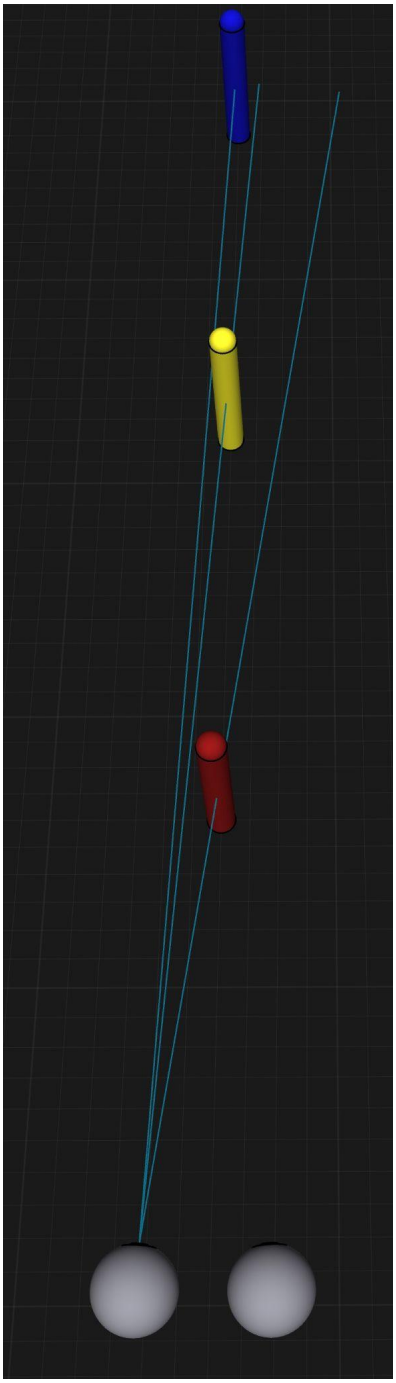
(3) Focus means bending a cone of light from a specific radius sphere shell (or hemisphere shell, or cone shell, depending on level of zoom) back to a point at the precise distance of where the sensor is. In theory, perhaps a hemisphere sensor would appear to be a good idea, as the individual tiny sensors 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.

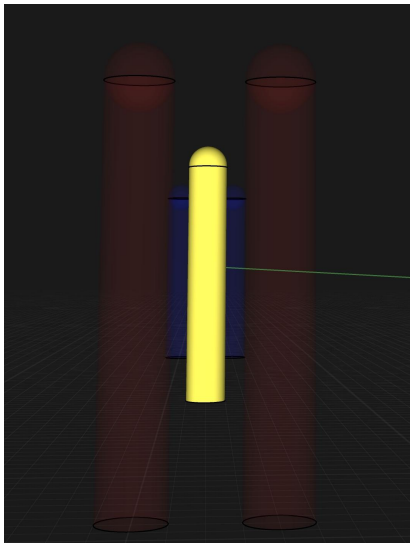
(5) A 2D perspective image, while looking fairly “depthy,” is fundamentally just 2D. To make it 3D, each point must be understood at its proper radial depth. In other words, the pupil of the eye (or the lens of a camera) is the sampling point in the physical universe, and each 3D vector light ray must 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.

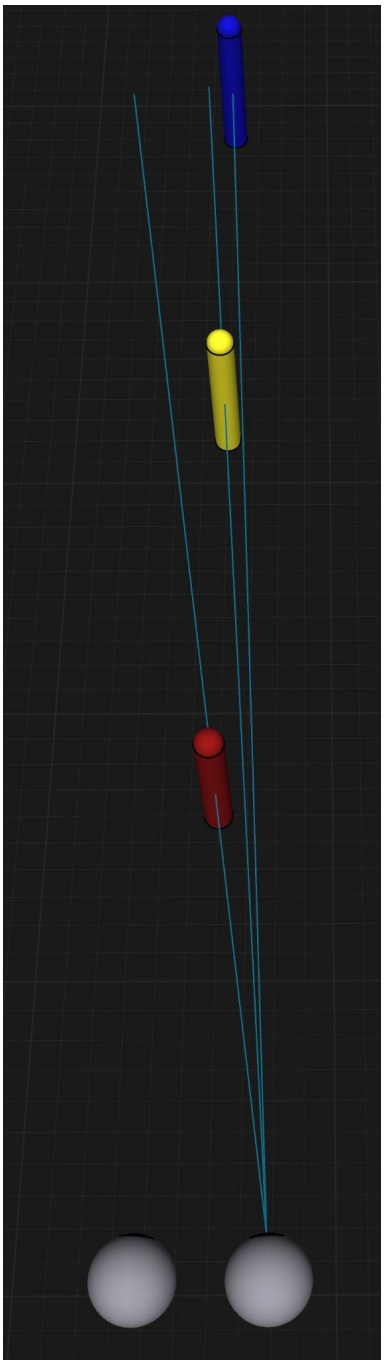
(6) But, once we introduce the fact that we have two eyes, it may become useful to consider the *foveal focal voxel* to be the center of rebuilding 3D rather than the eyes, for it is the fixed centerpoint that marries the two overlaid 3D scenes. Or, keep the centroid of each individual eye’s contribution, 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.

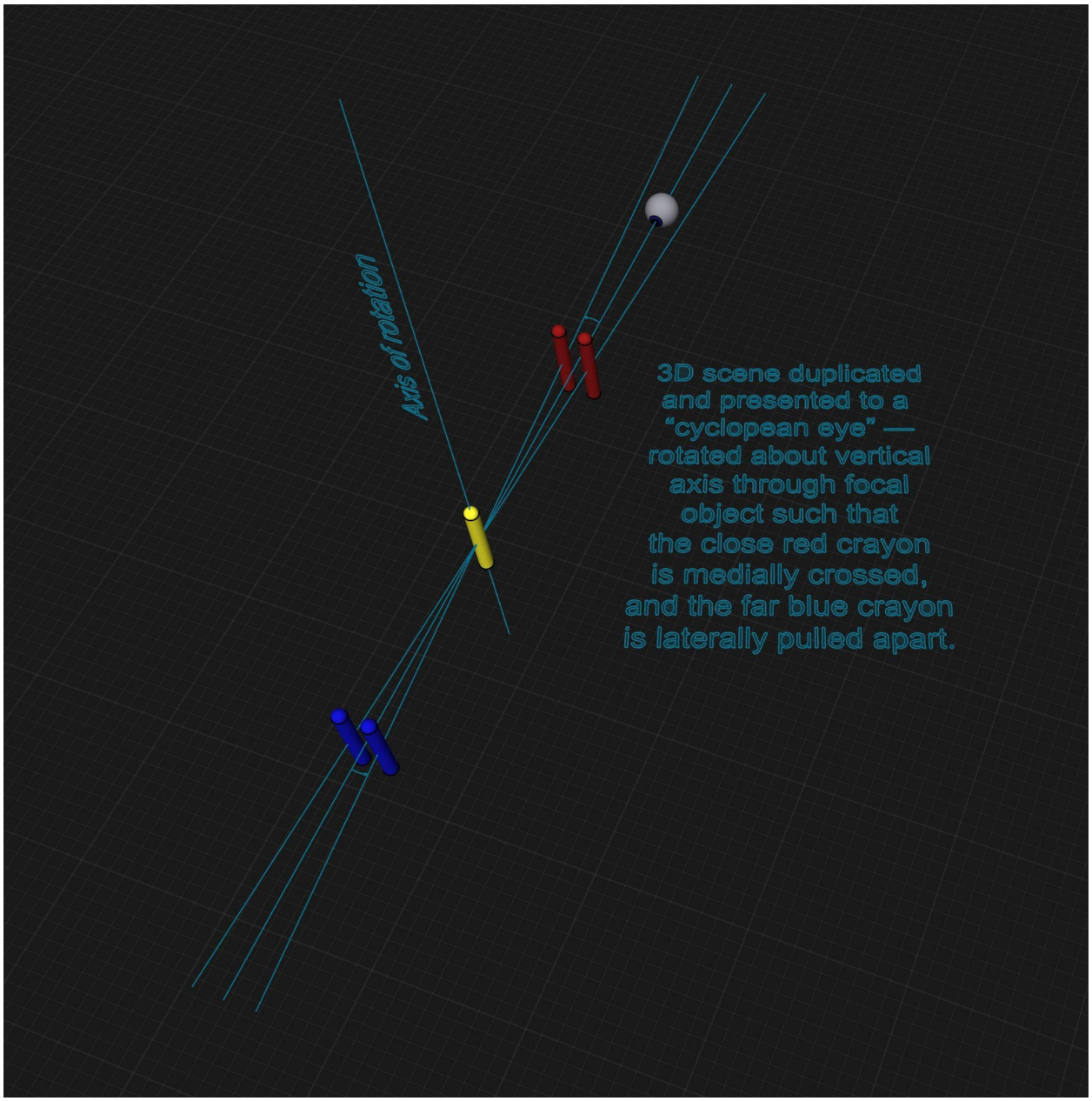


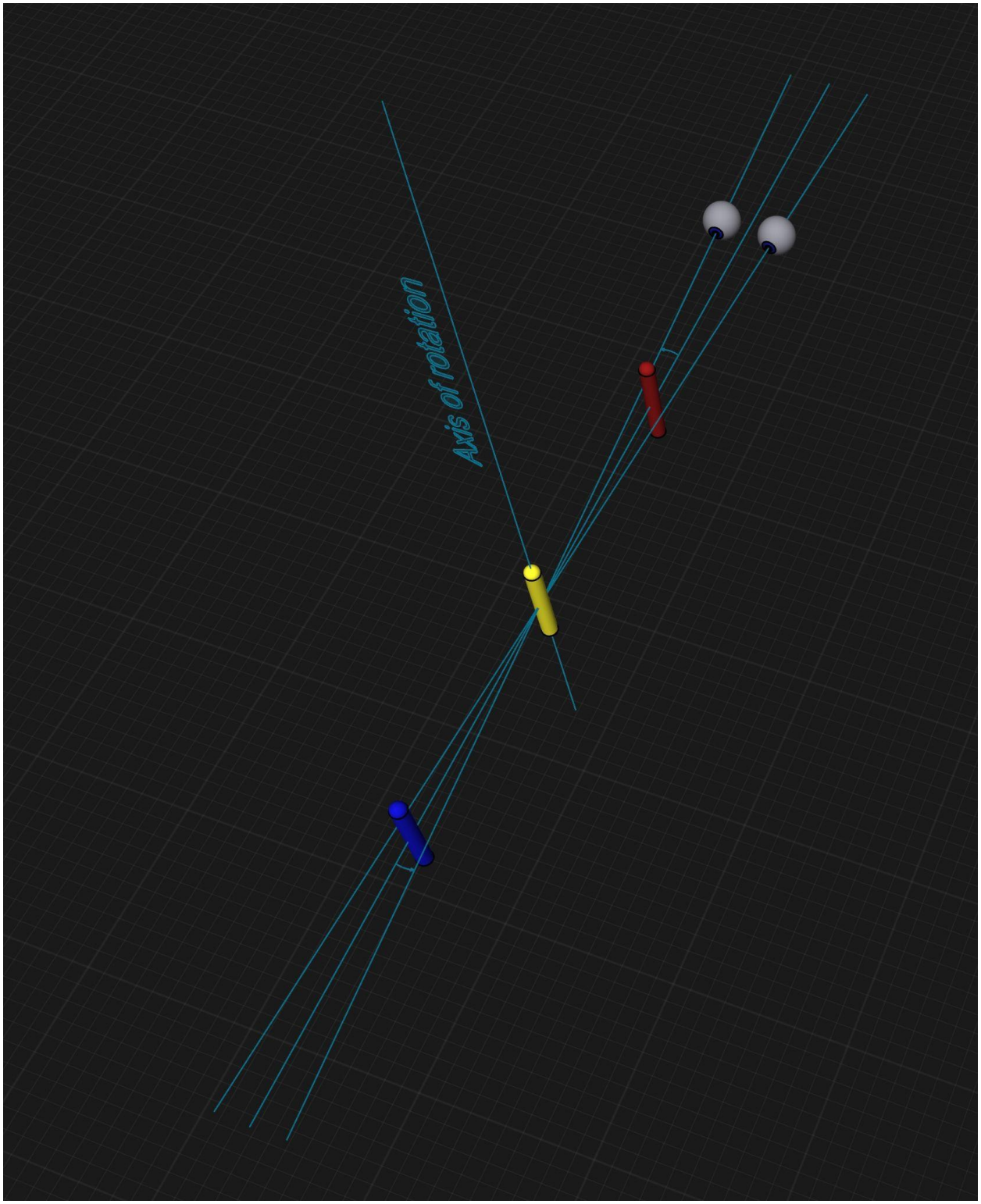
As a result, the red crayon duplicates and crosses medially (“crossed disparity”) and the blue crayon duplicates and pulls apart laterally (“uncrossed disparity”).



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.

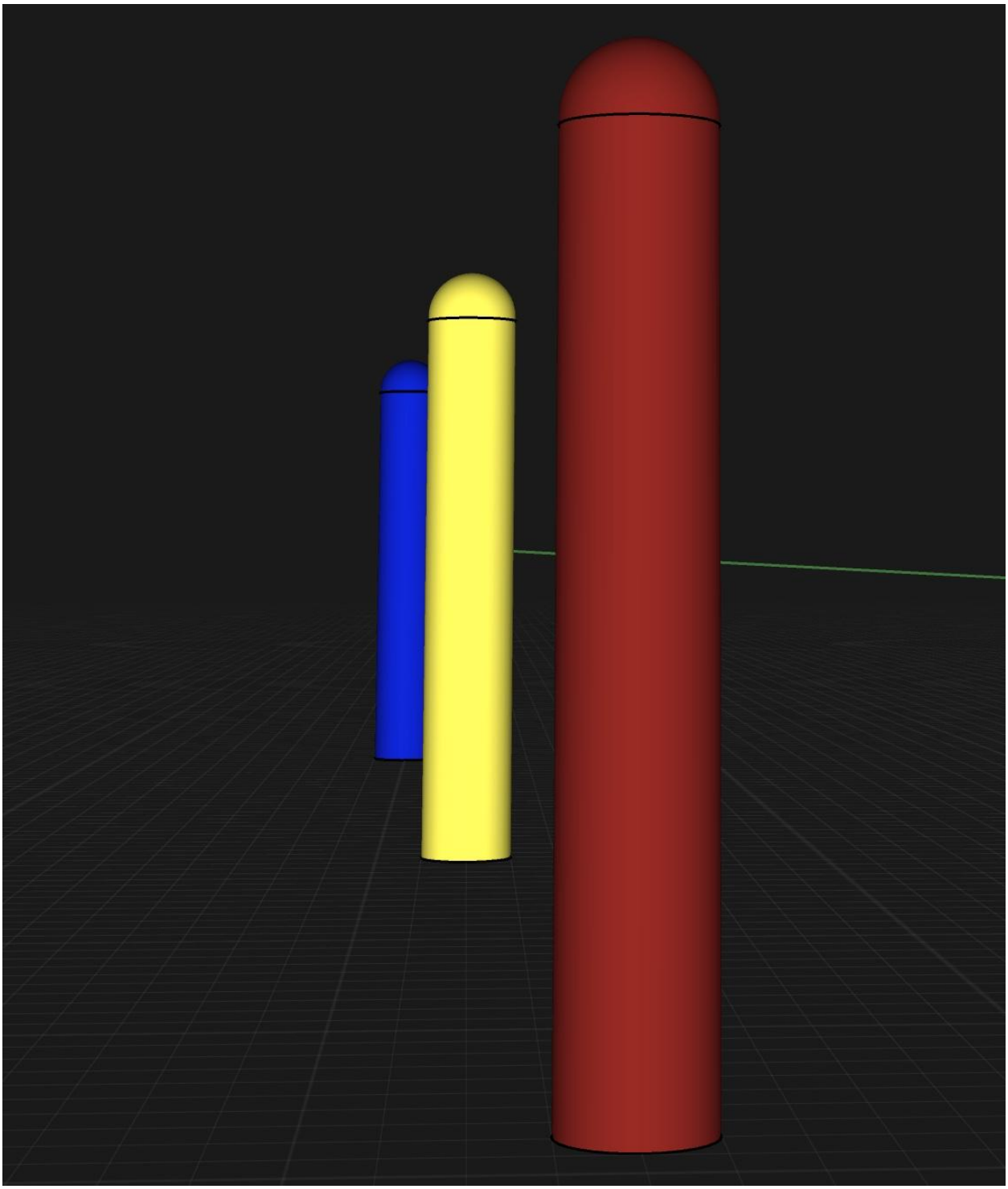
It seems nearly equivalent to consider ourselves as having a cyclopean (single, center) perceptual eyeball, with the two 3D copies of the perceptual world rotated about a vertical axis going through the focal object the amount that would offset the cyclopean eye into two copies that match the distance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



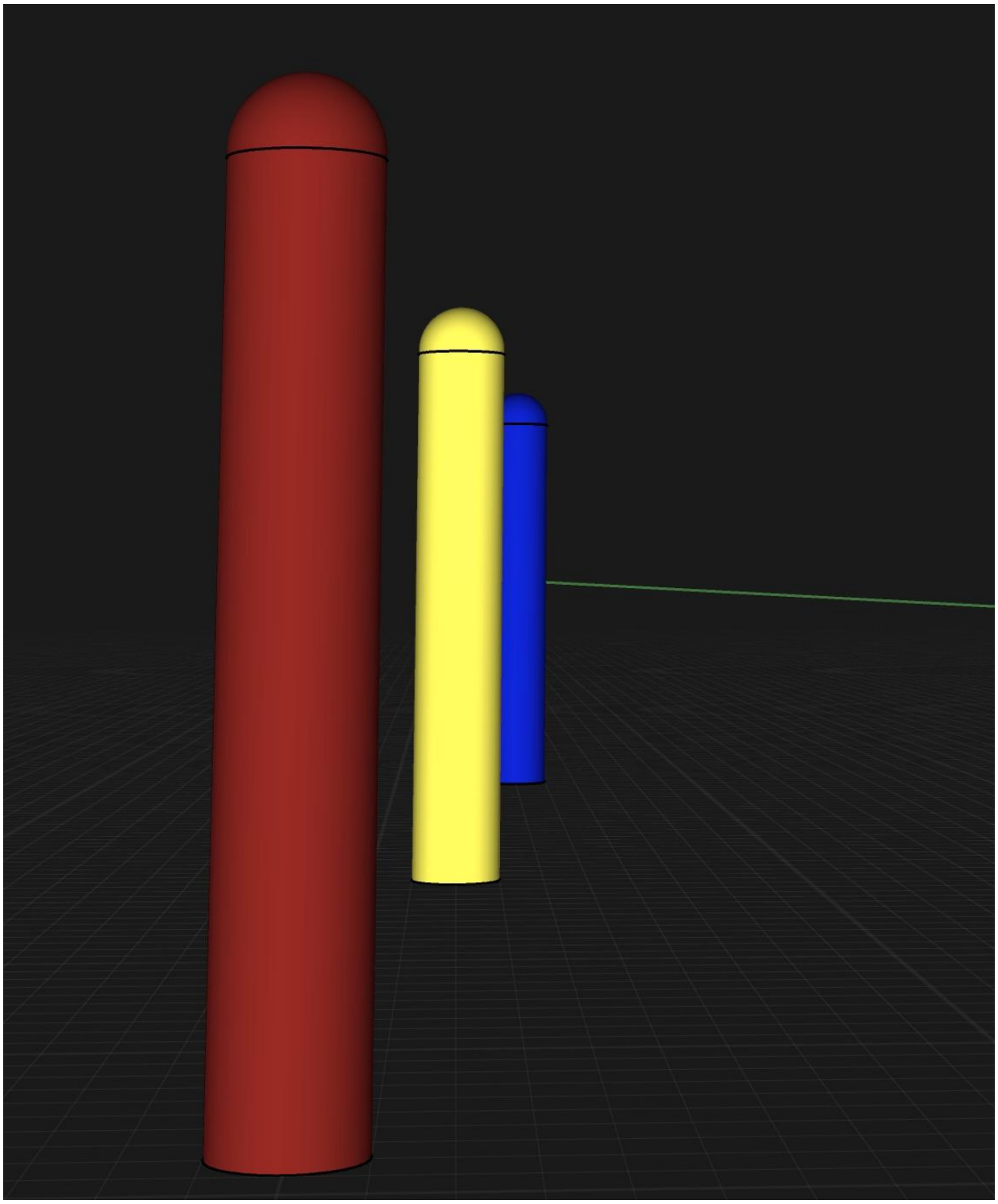


Here, let’s take another look at it. Below are what the left and right eyes see individually. Now it should make sense why together they create the “stereoscopic vision” of the ghost image a few pages previously shown.

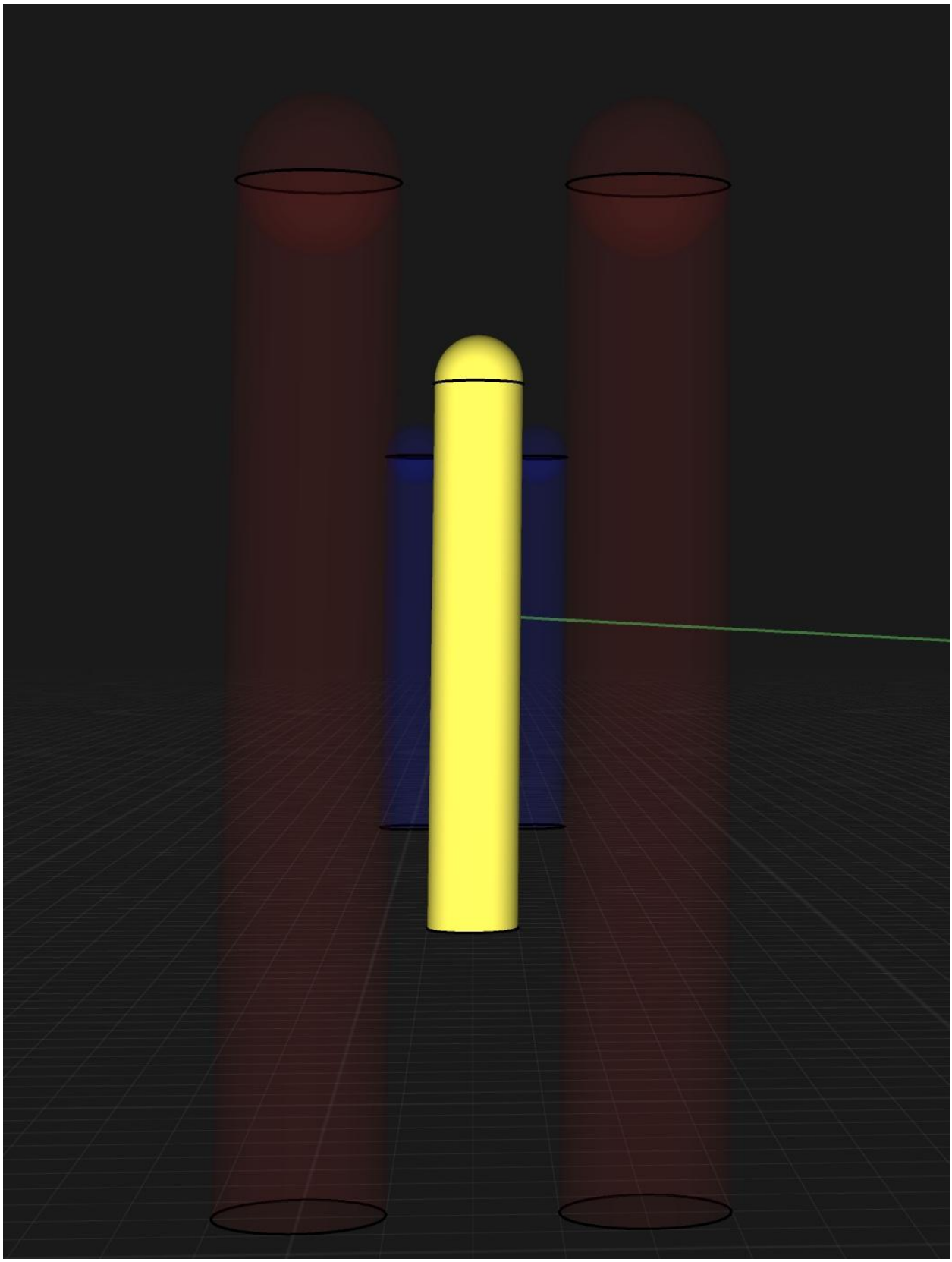
Left eye sees:



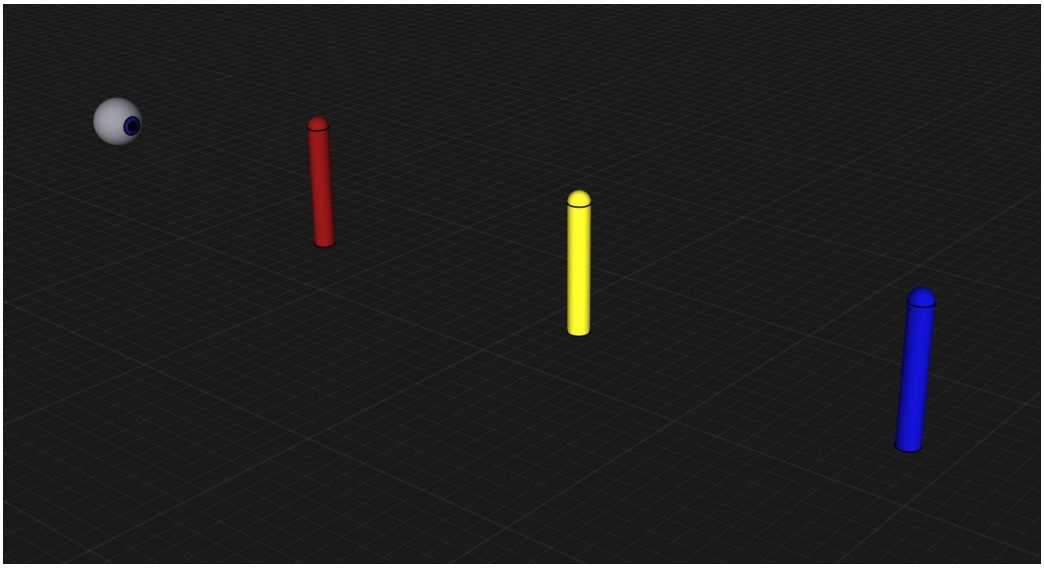
Right eye sees:

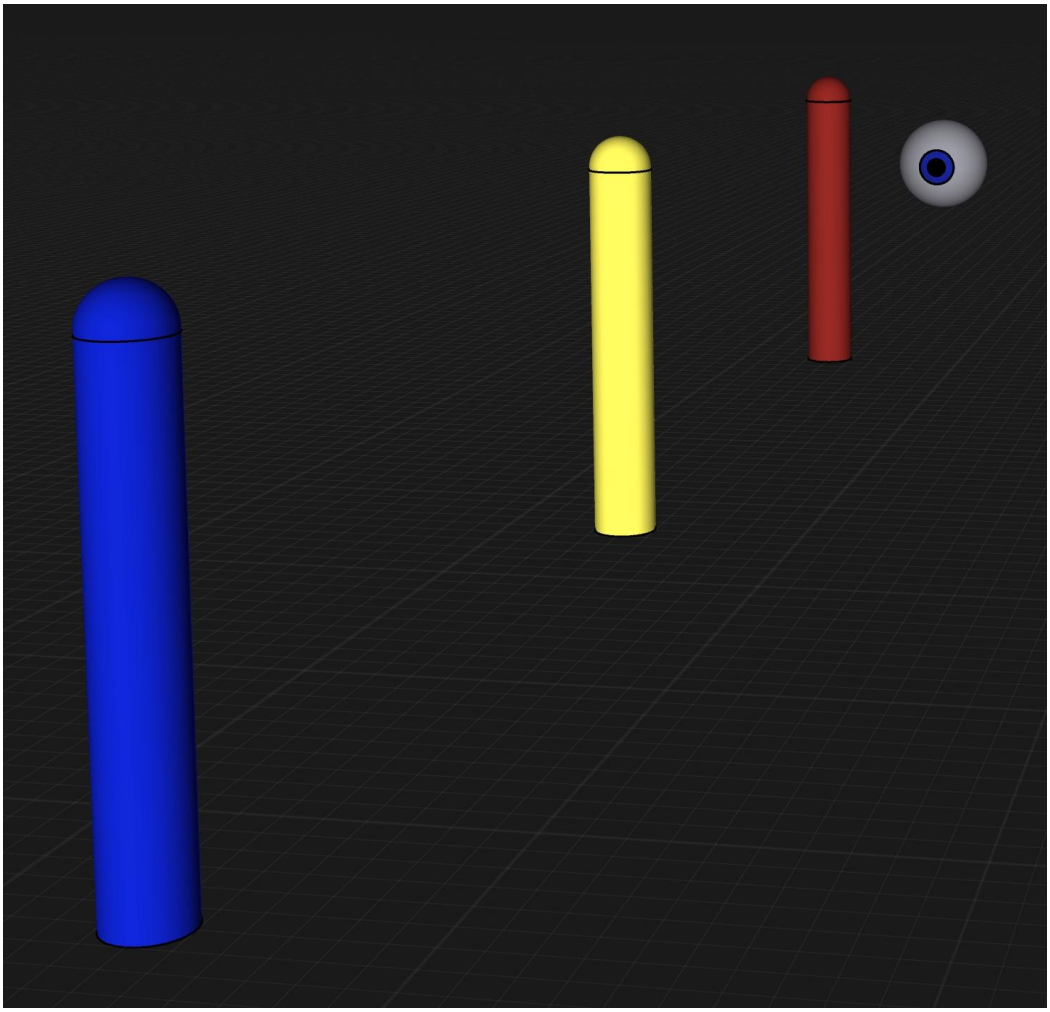


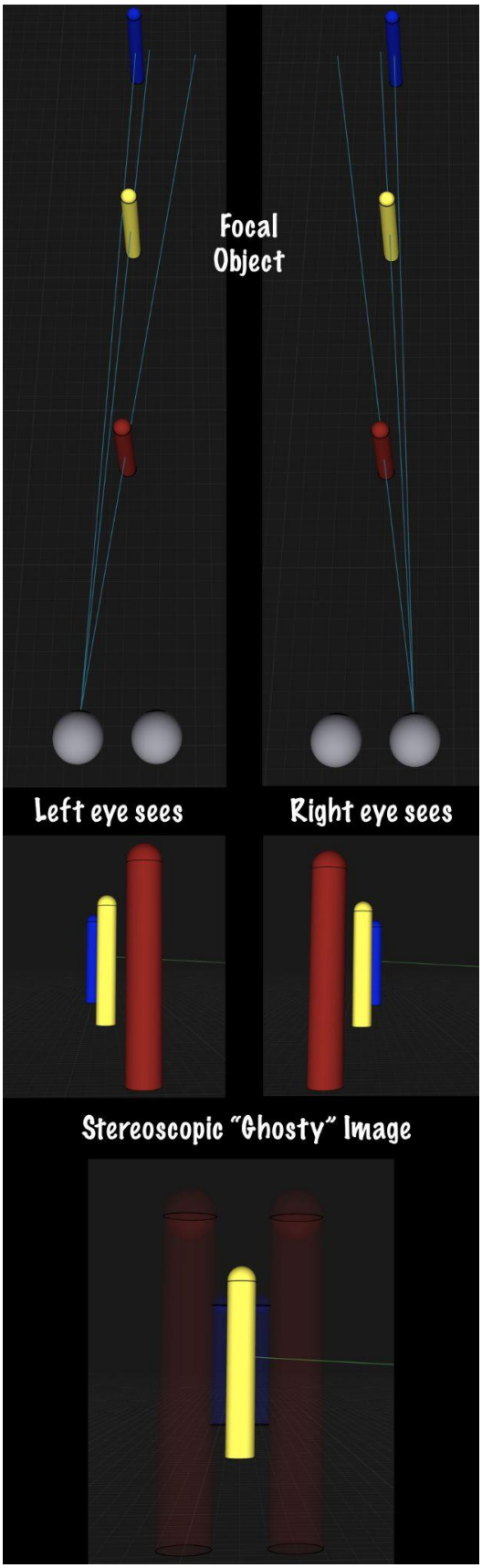
“Ghost” stereoscopic vision image, provided again:



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).