[SQUEAKING]

[RUSTLING]

[CLICKING]

JOSH MCDERMOTT: Greetings, everybody. Welcome back to 9.35. Last time, we started talking about depth perception, focusing on monocular depth perception. And we'll wrap that up today and then talk a little bit about stereopsis, binocular depth perception.

So just to recap what we talked about last time, the challenge here that we're talking about is that the world is three-dimensional. It's essential for lots of things that we need to do to be able to recover that three-dimensional structure. So that includes the three-dimensional shape of objects, where things are in depth relative to us. The problem is nontrivial because the input to the visual system are two-dimensional images. Now, we get two of them, but they're still two-dimensional images.

So we do take advantage of the fact that we have two eyes. And we'll talk about that shortly. But what we talked about last time, and what we'll finish today, are other ways in which we recover depth-- typically leveraging assumptions about the world that often are coupled with knowledge of the physics and geometry of the way that light interacts with objects to form images. And so we talked a lot about cues. So there are lots of different visual cues to 3D shape and depth.

And last time, we talked about shape from shading, linear perspective, texture gradients, the relationship between distance and size, height and field, aerial perspective, and more. So remember, shape from shading is based on the fact that many surfaces are approximately Lambertian. So the idea is that the amount of light that a surface reflects is proportional to the angle between the illumination and the surface orientation. And so if you have a surface whose orientation changes for a fixed illumination direction, the amount of light that's reflected will vary.

And so if you know the direction of illumination and you know that the intensity changes in the image are only due to shading, you can then infer shape. So we saw a bunch of examples of how we do this, talked about how the visual system seems to assume-- in certainly in the absence of other evidence-- that light tends to come from above.

And the consequence of that assumption is that in some cases you misperceive depth. So this thing is a crater but, from this orientation, looks like a bump. Here are some sand dunes that only really look like sand dunes when the light comes from above. We saw this demonstration of the same kind of thing.

We also talked about shadows. So shadows are created when an object blocks light. And so as an object changes in height above a surface, the position of the shadow will typically change. And so we use the positioning of shadows to infer depth. The visual system seems to have internalized the fact that shadows tend to be dark if you make them light. You don't really get depth as much. And then we saw all these examples where you can find things that either are shadows, but from some other object, or that just happen to look like shadows that cause objects to look like they're elevated up in the air.

And we revisited this idea that the visual system seems to model things at a local scale within the image. So we saw these demonstrations of situations where you perceive depth, but where it's really is logically impossible for the shadows to be oriented the way they are because it implies that there are these two different light sources that happen to be hitting the two sets of circles. But that doesn't really bother you at all. On the bottom, though, it does seem to bother you and interfere with the extraction of depth. So there's some degree of modeling of the shadows and illumination maybe, but on a local scale.

We talked about these regularities that have to do with the size of objects or texture elements and how there's also these relationships between where things are in the height within your visual field. So as things extend away from you, they're often on the ground plane. So they'll be higher in your visual field. And so if you get gradients where the size is varied with height in visual field, it gives you pretty compelling sense of depth. So these texture gradients are quite powerful. We saw some examples where artists have exploited this to give you the illusion that the floor is not flat, even though it is, and they just painted it in a funny way.

We then talked about Emmert's law. So remember, Emmert's law is this relationship between apparent size, visual angle, and apparent distance. And we saw this demonstration where we created an after image. So the idea is that the after image is something that's a fixed size on the retina, so the visual angle is fixed. And you can change the apparent distance by either looking at something that's very close or looking at something that's far away. The consequence of that is that the thing changes in apparent size.

So in general, when Emmert's law is at work, it's typically in situations where either size or distance is less ambiguous than the other one. And so changes in apparent distance will cause changes in apparent size or vice versa. So we saw a bunch of examples like that and examples in photographs where you misperceive things because of the application of Emmert's law in various ways.

So here's a situation where you seem to implicitly assume that the children are all the same size. And that means that they're all the same distance away, and that means that the girl has got to not have her feet on the ground. Saw these very large pigeons, where the distance is erroneously assumed to be very similar to that of the cars. And so they look pretty similar in size.

Talked about aerial perspective. It's an optical effect that happens in the air that you've learned is associated with distance. Linear perspective-- due to the fact that parallel lines in the world will converge in the image, unless you happen to take the image such that they lie in a plane that's parallel to the image plane. And we saw lots of examples of how these depth cues can interact with Emmert's law to create these funny illusions. So these two monsters are the same signs, but one of them looks like it's a lot further away, and so it looks a lot bigger.

We then saw some examples of what we see when we perceive the world are the three-dimensional structures that we infer. And so it's very hard to actually undo that inference. So you get these funny examples like this, where it's impossible to believe that those two parallelograms are the same size because you're not seeing them as they are in the image. You're seeing them in the way that you infer them in the world.

And then we wrapped up with these examples of bistability. So the amazing thing about perception is that despite the fact that we're constantly solving these ill-posed problems, we typically arrive at one pretty convincing solution to the problem. It's whatever you see. And usually, it's correct. But very occasionally, when you get these kind of impoverished stimuli, things can become bistable.

So the Necker cube is a very famous example of this, where you can see either face as being in front. And so if you look at it over time, you'll flip back and forth between seeing one and seeing another. And so that phenomenon of bistability is a pretty common thing in perception. It's super interesting, as we've commented before, that we don't actually just see one interpretation and stick with it.

So some people would argue that that maybe is related to the fact that there's multiple good solutions, and you're sampling from a distribution. We talked about a few different possible explanations of this. And we saw some classic examples of bistability where you can interpret things in two different ways and some jokes that are based on that.

All right. And so that's where we left things. So another kind of phenomena that is related to depth perception and ties into some of the other themes that we've talked about so far is the phenomenon of impossible objects. So these are drawings, or paintings, or photographs where you glance at the thing, and it looks fine. But then you inspect it in more detail, and you realize that it's not a possible object that could be physically realized in the world.

And so these things in the art of art of Escher are pretty common, where there are these staircases that look like they forever go up. And obviously, it's not really possible. There's another example where it looks like you have these arrangements of cubes. But then if you extrapolate what all the local relationships would imply, it's not something that could actually happen.

And so these things are kind of fun to look at and maybe perplexing. I think they actually tell us a lot about perception. And specifically, I think they speak to this idea that we've already hit upon in a bunch of other places in this class, that the visual system seems to model images in local neighborhoods. So you look at a particular point in an image, and it seems like you create this local 3D model.

As you could tell, the staircase maybe goes up, and maybe part of it intersects with another that keeps going up. And it doesn't seem like you model the entire thing at once, because if you did, well, I don't know what would happen, but you'd probably be aware that the thing doesn't make sense. So it suggests that you're not really requiring global consistency in the interpretation of the image.

So you can make these funny images that exemplify this idea. So this is another example of an impossible object that, at a glance, maybe looks OK, but doesn't really make sense. Here's an interesting case of this, where if you look at this, the local interpretation is forced to change. It's related to those figure ground illusions that we saw, where the motion makes it clear that something is a three-dimensional thing. And it's kind of painful to look at.

Oh, this is another one that I found. This is really wacky. So again, at a glance, anywhere that you look, this looks fine. But it's doing some things that are not possible, right?

It's related to depth because the reason that you're able to render this image is essentially because it's a two-dimensional image. And the thing that's not possible is the depth arrangement that's implied at every point in space. There isn't a global 3D structure here that is actually consistent with the images.

OK. Any questions about impossible objects or anything else we've talked about so far?

So why do you suppose that vision might work like this? Why do you think it might be the case that you make these inferences locally without really enforcing global consistency? Any ideas?

So one kind of idea that I think explains a lot of things about perception is that the world is always there. And so anytime you really you want to see something or know something about the world, typically what you do is you look at that. So we have this receptor lattice. It's very dense in this one part of the lattice. And so you typically look around the world, and you move that lattice onto the stuff that you want to see.

And because the world tends to be pretty stable-- and if you want to understand what the 3D geometry is, like, say here, you just look there. Or there, you just look there. Or there, you just look there, right? So there's a lot of aspects of vision that seem consistent with this idea that you only compute things as needed.

So there's something that you're interested in. You look at it, and then you do some computation that's related to the thing that you look at without necessarily analyzing, or representing, or remembering the entirety of the scene. And so we'll come back to that idea little bit later in the class.

So one other kind of thing that's related to depth perception is the hollow-face illusion. So this is a hollow mask. So you look at it, and it looks like the shape of someone's face. But it's actually a hollow mask.

And so these are cases where you misperceive the depth, in part because you have a pretty strong prior for what the three-dimensional structure of faces should be. And that has some pretty interesting consequences on the motion that you see. So this is another really interesting example where motion and depth seem to be really tightly

Yoked. So I'll show you this one video about this. And then we'll experience this ourselves in a sense. One second.

[VIDEO PLAYBACK]

- We've had an Einstein mask for many years. Now we have one that's fully painted, decorated by an artist. And in fact, with all the flesh tones and the white hair and so on, it's remarkable because it really does look like Einstein himself. But this is a hollow face, don't forget.

Here is the piece being turned around very slowly to show that it's hollow that side. And then this side is all in gray. You're not supposed to see this side normally. It's just the back of the mask.

We keep turning around. And there's a nice effect here as he suddenly pops out to you because although it's a hollow face, he's going to suddenly pop in your imagination to be a full face. An effect like that, also, you see very remarkable is when you tilt the piece back and forth like that. You get some very, very strange effects of the head tilting upwards and downwards. And the eyes seem to follow you up and down.

And now if we just leave the mask on the table and move the camera, you'll get some very interesting extra effects as well-- hollow-face illusion.

[END PLAYBACK]

JOSH MCDERMOTT: Greatly enhanced by the dramatic exposition. OK, so what's going on there? So this is a case where you look at this thing where the depth is actually the opposite of what you see it to be, right? So the actual structure in the world is the opposite of what you see it to be. And then it moves.

And what's so freaky about looking at this thing is that the face appears to move in the direction that's opposite to the direction it is actually rotating. And so this, at some level, is a simple consequence of geometry, which is that if you have a three-dimensional thing and it rotates in one direction, the points at the back are going to move physically in one direction, and the points at the front will move physically in the other direction.

And so if you take an image of this, that stimulus-- and we saw examples of this where-- remember that stimulus where you have that thing that looks like a sphere of dots that's rotating around? And it's bistable, right? You can see it as rotating either direction. But when it switches direction, the depth also flips. So in that case, it's not very striking because both depth interpretations basically look the same. It's just a sphere of dots.

But in this case, that has the consequence that-- and particularly, in this case, part of the thing is reversing, which is the part that's the face. And so because of this kind of geometric ambiguity about the depth of the points and the direction of motion, there are these two possible interpretations.

One is the actual depth of the thing-- depth interpretation of the thing, which is that it's hollow and rotating in one direction. The other is the depth that you see, which is where it's poking out. But in order for it to be poking out and for the image motion to be what it is, the actual face has to be rotating in the opposite direction from the direction it actually is and from the direction that the frame is. So that's essentially what is going on here.

And so what does this tell us about perception? Well it's an additional piece of evidence that the three-dimensional motion interpretation that we make about the world is very coupled to the depth interpretation. Those things go hand in hand. So the visual system in this case is tricked into the wrong depth interpretation.

And the way that it's tricked is in part because you have a very strong prior that faces are sticking out rather than hollow. And in the version in the video, the face was painted to also accentuate the three-dimensional interpretation. So you get tricked into the wrong depth interpretation. And then the 3D motion interpretation also has to be incorrect in order to be inconsistent.

All right. So I'm going to show you a real-life example of this. In this class, you learn a lot of party tricks, and this is yet another one. So what I have for you here is the wire frame that is affixed to the end of a champagne bottle. I have three of them. That's all I could find in my office. We keep these every time someone graduates. We open champagne, and then I keep these to show all of you.

So why is this important? Well, so this thing, you can think of this as like a three-dimensional Necker cube. So it's got two faces and these edges. Normally, you hold this thing up, and you look at it with two eyes. And so the depth interpretation is unambiguous. So right now, this thing is closer to me than this thing.

But if you close one eye and you wait a little while, the depth will reverse. Now, when you rotate this thing, the most psychedelic thing you can imagine will happen, which is that you will see the wireframe as rotating in the opposite direction as your hand. So the wire will look like it's non-rigid.

So there are three examples here. I'm going to pass this around, and you can try this for yourself. It may take a couple tries. So again, the method here is you're going to look at this with one eye and wait for the depth to reverse and then rotate it. And try not to scream, OK? There you go. And pass those around. OK.

So I'm going to keep lecturing while you all try this yourself. It may take a couple tries. And I should just say, just to be clear, this is exactly the same phenomena as this hollow-face illusion. What's happening is that we are waiting for your brain to arrive at the incorrect depth interpretation. We're then moving the thing. Your brain is then forced to make the correct motion interpretation to be consistent with the incorrect depth interpretation, even though it seems physically impossible. Does that work for you? Yeah, OK.

OK. Pop quiz here-- name some depth cues and depth-related phenomena in here. Yep?

AUDIENCE: Object size.

JOSH MCDERMOTT: OK. Are you referring to that one?

AUDIENCE: I'm referring to all of them, some being the object's closer or bigger than the objects further away. And that one, that's--

JOSH MCDERMOTT: Yeah. So that's kind of like Emmert's law, where you know that all of these things are the same size, and so that kind of implies that those things are further away. But down here, we have the opposite application of Emmert's law, where the inference, if this is a lot closer than that, is that that's very close. And thus, it must be a toy or something. So that's Emmert's law applied that way.

What are some other depth cues that you can see here? Yep?

AUDIENCE: I forgot the name. But in the background, the mountains are faded.

JOSH MCDERMOTT: Yeah. So that's what people call aerial perspective, yeah. So the stuff gets blurry when it's further away. Yep. Anything else?

AUDIENCE: Linear perspective.

JOSH MCDERMOTT: Yep. You've got some converging lines there. Yep. OK, good. How about here? What are some depth cues here? Yep?

AUDIENCE: Shadow and light.

JOSH MCDERMOTT: Shadow? Yep. So we got some shadows here that tell you that that's off of the ground plane. Anything else? How are you inferring the 3D shape of those objects?

AUDIENCE: Shading.

AUDIENCE: Shadow [INAUDIBLE].

JOSH MCDERMOTT: Shadows. And somebody said something else?

AUDIENCE: Shading.

JOSH MCDERMOTT: Shading, yeah. So there's a lot of shading here, right? And then there's also occlusion-- so the fact that this object looks like it's going under that one. So we didn't really talk a whole lot about that.

What are some depth cues here?

AUDIENCE: Shadow.

JOSH MCDERMOTT: Yeah. What else?

AUDIENCE: Size.

JOSH MCDERMOTT: Size, yeah. So this is kind of like a texture gradient. What else? Anything? Linear perspective because the knobs are arranged in these lines, and those converge.

This is actually a very interesting case where there's some shading on these objects. But these objects are not Lambertian. So you can see specular reflections here. And this is, in practice, what makes the problem of shape from shading very challenging, is that a lot of surfaces are not purely Lambertian. You get specular reflections, so you have to distinguish that from shading and from reflectance and so forth.

How about here? What are some depth cues here? Yep?

AUDIENCE: There's the aerial view of that or something.

JOSH MCDERMOTT: Yep. Aerial perspective, yep. Yep?

AUDIENCE: Linear perspective or [INAUDIBLE].

JOSH MCDERMOTT: Yep. Yep, linear perspective. Yep. Anything else? Yeah?

AUDIENCE: I think you can also assume that things are the same, so they get smaller and further away.

JOSH MCDERMOTT: Yeah, that's true. Yeah. What else?

AUDIENCE: The back is lighter.

JOSH MCDERMOTT: Yes, aerial perspective. Yeah, yeah. And then also some texture gradients. So the bricks gives you a texture gradient. Yeah. OK.

So just to summarize what we talked about in terms of monocular depth perception-- so as you could probably tell from the lecture, I mean, our understanding of these issues, I would say, is still mostly qualitative. There's some sets of cues and associated phenomena. We don't yet have working models that can account for human perception. This is something that I think merits revisiting in the current era.

So we talked about a bunch of cues and a bunch of phenomena that give us some clues to what an eventual working model ought to look like. So we talked about shape from shading. Key ideas there is that shape from shading typically assumes Lambertian surfaces as well as no variation in reflectance and known illumination. We often assume that the illumination is coming from above.

Shadows are a powerful depth cue. They have to be dark. It can help for them to be blurry. We talked about textures gradients. Talked about Emmert's law that relates size, distance, and visual angle, and different ways in which that can explain illusions that are related to depth perception.

Talked about how what we see are the inferred 3D shapes that we infer from 2D images. It's very hard to access the 2D image that we use to infer the shape. We saw examples of impossible objects and talked about how they suggest that depth representations are computed locally. And then we saw these phenomena that illustrate that motion and depth are jointly inferred. And we also saw some examples of that back in the lecture on motion perception.

What questions you got? Is everyone getting the wireframe thing to work? Did anybody not get it to work? All right, good. That's a pretty good success rate. Are you impressed? Yeah? OK, good.

All right. So let's turn to binocular stereopsis. So we just talked about how there's lots of monocular accused of 3D shape and depth. But another big piece of the puzzle is the fact that you have two eyes.

So if you look across the animal kingdom, organisms tend to come in two flavors when it comes to the positions of the eyes. So there are some critters that have the eyes positioned on the sides of the head. So here are a couple examples. There are others that have the eyes positioned at the front.

In general, the animals that have eyes in the front are what we consider to be predators. They hunt their prey. Animals that have eyes on the side of their head are often the ones that are probably worried about getting eaten.

So the notion here is that animals that are in the business of hunting for subsistence need to know how far away things are from them very accurately. And so they have two eyes on the front of their head because two eyes, as we will discover, really helps to give you precise depth perception. Whereas animals that are worried about getting eaten, it's more important for them to be able to see as much as they can at any one time, and so the eyes are positioned to give panoramic vision.

So there's a real cost to having two eyes in the front of our head like we do. So it greatly reduces the field of view. So if you just look ahead and you look at what the most peripheral thing is that you can see, you get, I don't know, a good chunk. But you definitely can't really see more than maybe 150 degrees, something like that. So there's this huge part of the visual world that is not visible to you.

But what you get in return for that is disparities between the two eyes. So your two eyes are in slightly different positions. And so the consequence is that they take slightly different images of the world. And those differences between the two eyes provide information that allow us to discriminate very small differences in relative depth.

So here's a demo that we're all going to do. So you're going to hold your two fingers out in front of you. Fixate the one that is close to you and then alternately close and open the two eyes. So you're going to look at it with one eye and then the other. And what you will see is that the distance between the two eyes in the image will be different for the left eye and the right eye. So that's an example of binocular disparity. And so in this case, that gives you information that one of the fingers is further away than the other.

All right. So binocular disparity refers to differences in positions between the two eyes. So in a situation where one object is being fixated-- let's say that's object a here. If b is not being fixated and is at a different depth, then the relative position of b will be different in the left compared to the right eye, and that's the disparity. And so we typically measure the disparity in degrees of visual angle or minutes of visual angle.

So here's a schematic. So here are your two eyes. You're fixating this red object here. Remember, what it means to fixate something is that your fovea is directed towards that object. And so the red point is projecting onto the fovea a portion of the retina.

So this dark-blue thing, which is further away, also projects onto the left and the right eye. But you can see that it's at a different place in the left eye compared to the right eye. So that creates a disparity.

So when fixating one object, the other object can be in front of or behind the focused or fixated object. If b is further away, we say that the disparity is uncrossed. By convention, we just consider that to be positive. If it's closer than a, we say that it's crossed. And by convention, we say that it's negative. And so the idea is that you would have to cross your eyes in one case to look at b.

So here's the idea. So in this case, the blue object on the left has crossed disparity. And on the right, the red object has uncrossed disparity. So you'd have to uncross or cross your eyes to fixate it.

So objects can also give rise to zero disparity or no disparity. So this happens when the lateral separation between objects is the same in both eyes. And the set of points that has this particular property is known as the horopter, also known as the Vieth-Muller circle. So there's this. And if you take a 2D slice here-- In reality, this is a 3D world-- but in this 2D slice, you get this circle here. And every point on this circle will project to corresponding points on the retina. And that would cause-- who would be willing to volunteer for this purpose? Preston. All right. Thank you.

So you're going to need two pencils or two pens-- two things that are pointy. Do you have that? Maybe somebody can lend. I should have brought those. Yeah, that'll do. OK.

OK, so just sit right here. So I want you to grab one in each hand. Point them towards each other. Now, using both eyes, I want you to bring them into contact with each other. Touch the tips. You're good. You're good.

OK, do it again. Start further away. Now bring them into contact. Very good. OK, now bring them apart. Close one eye. Now do it again. Oh, my goodness. It's a miracle.

[LAUGHTER]

Try it again. Try it again. Yeah. So it's much harder, right? OK.

So what you get from having two eyes is very, very, very accurate depth perception. So you have depth perception with one eye, But the acuity is much less. Preston, thanks a lot. Appreciate it. Let's give him round of applause.

[APPLAUSE]

And you should all try this yourselves. It works on everyone. So there's a really big difference here.

So just to give you a quantitative example-- so using binocular stereopsis, you can discriminate a 1-millimeter difference in relative depth at a distance of 1 meter. So that's very, very small. And so the disparity that is created by that difference is tiny. It's that really small number of millimeters. So it's less than the diameter of a single photoreceptor-- so these very, very small differences between the two eyes that the brain is really attuned to because they tell you about fine-grained differences in depth.

So one very compelling demonstration of binocular stereopsis makes use of this device here, called a stereoscope. So this was invented in the mid-1800s by Wheatstone. And so what the stereoscope does is it's a device for taking two images and projecting one to the left eye, one to the right eye. And so the two images are taken from slightly different positions, so they have slight disparities in them. And they cause you to see things in 3D. So this was a demonstration that you can get a vivid impression of depth from two flat pictures.

And so here's another diagram of this. This particular version of it actually works a little bit differently. This uses prisms. But the classic version involves mirrors. So you have two images on either side and then two mirrors.

SIRI: [INAUDIBLE]

JOSH MCDERMOTT: My computer thinks I asked Siri about something, which I didn't. OK. OK, I must have said something that sounds like Siri. So you can tell that machine speech recognition is not quite where it needs to be. All right.

So we've got these two images and two mirrors. And so each eye sees a different image. So in this particular version, there are two prisms here. And then there are these cards, each of which has a different image taken from a slightly different position. And so you look at it, and each eye gets an image. And you can see this as being 3D. So I'll pass this around in just a second.

So stereoscopes were invented around the same time as cameras. And so they became a popular fad. So they were very common back in the 1800s.

People were once crazy about stereo views. Many families had a stereoscope and a collection of stereo views, which they used to entertain friends, educate their children, and see the world. Oliver Wendell Holmes called the stereoscopic photograph the card of introduction to make all acquaintances. So this was like, you'd have someone over to your house, and then you'd show them stuff on your stereoscope. Yeah, old school.

So I'm going to pass this around. So you can still get these things on eBay, and Amazon, and things like that. So here's just one example, and you can try it out for yourself. So you just look through. There you go. So that's how they did it back in the day.

Nowadays, with modern technology, we have other ways to experience binocular stereopsis. One such way is called an anaglyph. So an anaglyph uses colored ink and colored glasses to present different images to the eyes. And so I'm going to hand these things out to you. And you can try this out for yourself. Some of these are actually-- you want one that has one red and one blue. Actually, TAs, maybe I can put you in charge of making sure everybody gets one of these. There's some more in here. I don't know how many of these we need. There you go. OK.

So we'll pass this out. And you're going to put them on. And the little thing in the bottom is supposed to tell you which side the red goes on and which side the blue goes on.

So here's the idea. So we've got these two different images. These are just synthesized. So this one is supposed to go to the left eye. This one is supposed to go to the right eye. You can see that what distinguishes them is that in the left eye, the letter A is a little bit to the left of the letter B. And the letter C is a little bit to the right. Whereas in the right eye, that's reversed. So that's going to create disparity.

And so now the anaglyph just puts one of these images in the red channel and one of these in the blue channel. And then these colored glasses are going to pass one of those images to one eye and the other to the other eye. And when you do that, lo and behold, if you're not stereo blind, you should see depth differences.

AUDIENCE: What's stereo blind?

JOSH MCDERMOTT: Stereo blind means that you can't make use of stereo. So we'll talk about that in a minute. So 5% of people supposedly are stereo blind. Everybody, keep these on. Let me take a quick picture here. Everybody looks very stylish and hopefully is experiencing very vivid impressions of depth. Does this work for you? Raise your hand if this works. Most people, maybe. Yep. OK, good. All right, we'll see some more. So this one is not that exciting.

This one is a little bit better.

AUDIENCE: Ooh.

[LAUGHTER]

JOSH MCDERMOTT: Yeah. Yeah, I like it. So if you flip the glasses, the depth will switch because you're going to reverse the disparity.

AUDIENCE: Whoa.

[INTERPOSING VOICES]

JOSH MCDERMOTT: These are just generated synthetically. You can do this with photos. These are like the photos that go into the stereoscope-- so two different photos from two different positions. I'm not entirely sure what will happen if you flip the glasses here.

AUDIENCE: How does this work if it's all gray? There's a little bit of--

AUDIENCE: It's not all gray.

AUDIENCE: [INAUDIBLE] the red. It's very subtle.

JOSH MCDERMOTT: Yeah. I mean, if you add the blue and the red together, you're going to see something close to gray. But the glasses separate them out. Yeah.

[SIDE CONVERSATION]

JOSH MCDERMOTT: I see murmurs of something-- or hear murmurs. What's that?

AUDIENCE: It looks a lot better.

AUDIENCE: Oh, no, these do.

JOSH MCDERMOTT: Oh, I see. Yeah, OK. Our projector here is possibly not helping us too much. Here's a Mars rock.

JOSH MCDERMOTT: Whoa.

[LAUGHTER]

[SIDE CONVERSATION]

JOSH MCDERMOTT: Here, show me. Yeah, no, it is definitely better. I don't know what's wrong with our projector. So just to be clear, the reason why this might work better on your iPad or your laptop screen than the projector is that all of this depends on these things being the right colors relative to these glasses. And so if the projector is a little bit off, then it may work less well.

This one is pretty cool. Yeah, Christa?

AUDIENCE: Would these types of 3D effects work if you had different glasses?

JOSH MCDERMOTT: Oh, yeah. There's nothing special about red and blue. Yeah, you just need filters that will get one image to one eye and another to the other eye. Yeah. Yeah.

So I think the moral of all of this is that people like 3D. And that's consistent with the fact that 3D movies are a recurring fad. So nowadays, if you go and you see *Avatar* or whatever in 3D, the glasses are not like these. They're a little bit different. So the glasses actually use polarizing filters.

So you can present two projectors that polarize light at different angles, and then that light will be passed by one lens or the other. And so the advantage of that is that the film can be in color. So if you're going to use this anaglyph technology, then you're stuck with a black-and-white film, which people don't like. Yeah?

AUDIENCE: This this the cool thing. After wearing these for a few minutes and then taking them off, and you close one eye, when I close my left eye, I feel like my right eye sees more red. Then when I close my right eye, the world went blue.

JOSH MCDERMOTT: Yeah. So that's photoreceptor adaptation. Yeah, yeah. Yeah, yeah. Yeah, the miracles of vision. OK.

OK. So binocular disparity-- so that's the basis of all of this, these small differences between the images. So the intuition here is obvious, right? So the question is, how does this actually work? How do you actually compute disparity?

So one of the big challenges is what's called the correspondence problem. So in order to measure disparity, that assumes that you know what in the left image goes with what in the right image. And so when I show you these diagrams, it's obvious because there's only two things, and they're lines and stuff. But in actual images, it may not necessarily be obvious what in one eye corresponds to what in the left eye.

And so in particular, what are the features that would be matched? And where in the visual system might this occur? So for a long time, people supposed that stereopsis happened after shape analysis. So the way that the vision would work is you'd get the left image, and you get the right image, and you do some analysis of each of the two images to figure out where the shapes were. And then on the basis of that, you would then compute disparities. So the idea with stereopsis happened late.

So the modern view is the opposite of this, where we now think that stereopsis is something that happens earlier, probably prior to what you might normally think of as shape analysis. And really, the key demonstration that caused this shift in how people think about this was random-dot stereogram.

So this was invented by Bela Julesz, who was a perceptual scientist who was working at Bell Labs at the time. And they had just started the-- it had only been a few years that computers had come online and it was possible to actually generate these synthetic images. And so the random-dot stereogram consists of images that are random dots. So each dot in the image is randomly assigned black or white with 50% probability. And then you take that image and then take a square and shift it over by a dot or two. And so that gives you the other image.

And now we have our anaglyph. And you should-- oh, yeah. You can flip the glasses either way, and you'll either see a square in front or behind. And yeah, this definitely does not work quite as well as it should, I think because there's a mismatch between the color guns and the projector and these glasses. But you hopefully get the idea.

So each eye by itself gets an array of random dots. And so the point of this is that there are no obvious objects or features that can be identified and matched between the eyes. There's just these random-looking textures. One second.

But when the eyes are combined, there's a square that becomes evident, either floating above the background or behind it. Yeah?

AUDIENCE: I was just curious how eye dominance plays into this because I noticed that when I flip it over, it's definitely whatever color it is on my dominant eye. My right eye is where I'm seeing more. Does eye dominance also play into depth at all? Or is it just the color that I'm seeing?

JOSH MCDERMOTT: It probably does. And when you say eye dominance, what does that mean to you/

AUDIENCE: When on the team, you have to see which eye your brain is using more.

JOSH MCDERMOTT: Oh, yeah. OK, yeah.

AUDIENCE: If I close my left eye, my view doesn't shift. But if I close my right eye--

JOSH MCDERMOTT: Yeah, yeah, yeah, yeah, yeah, yeah, yeah, yeah. I don't know how that's going to affect stereopsis. I mean, my intuition is that it might make it weaker. So there is an extreme case-- and we're going to talk about this at the end.

There's a phenomenon called amblyopia, where one eye completely dominates the other. And this typically happens when the eyes are misaligned, usually at birth. And if that's not corrected-- there's a critical period. And if it's not corrected before the critical period is over, you get this phenomenon called amblyopia where one eye dominates, and you don't have stereopsis after that. But there's probably gradations of that, and that may be what you're experiencing. So yeah, I don't know how that's going to affect normal stereopsis. Yeah.

OK, so this is the random-dot stereogram. And so you can look at this as an anaglyph. You can also look at it as a stereo pair. So people who study stereo vision train themselves to free fuse. So what that means is that you train yourself so that your eyes can diverge or converge at will.

So normally, when you look at something like a screen, you just fixate the screen because normally, when you fixate something, you want to bring that into focus. But you can train yourself to have your eyes actually not fixate the screen, but to fixate a point either in back of the screen or in front of the screen. And that can cause the left image to be registered at the same location as the other image in the other eye. And that will then give you disparity that you can then use to see depth.

And so, yeah, you can try holding up your thumb and get this to work. It usually takes some practice. I never got to be very good at this, even back when I was of doing vision for a living. But some people are very skilled at this. Yeah, it's a little tricky. You can try it at home.

So this is the actual stimulus just to make sure everybody gets this. So here's the stimulus to the left eye, and here's the stimulus to the right eye. So it's just the square of dots that's offset. And each image on its own doesn't really have any kind of meaningful features or structure to it.

So the significance here is that it suggests that you certainly don't have to do anything like object recognition before matching things between the two eyes and getting binocular disparity. So maybe there's some kind of shape analysis, but it's not the interpretable form analysis that you might think of.

And in particular, this is also important because it demonstrates that you can get depth from binocular cues alone. So even like the stereoscope, it's showing that the depth kind of gets enhanced just from these two images into the two eyes. But these are actual photographs, and they got a million different monocular depth cues. And so the random-dot stereogram isolates binocular cues.

So here's just another example to illustrate the correspondence problem. I've been showing you all of these pictures of how points in the world project to points on the retina. And it's a little bit misleading in the sense that the visual system, when it is making use of binocular information, it doesn't know this thing. All you get is this. So you get these two images and the two eyes, and then you're trying to infer what's out there in the world. And so you could create these three dots from three dots that are in a row at the same depth. But there's other kinds of arrangements of dots at different depths that could produce the same thing.

And so, in general, I mean, this is another area of vision where I would say we don't really have complete models or theories. I mean, the correspondence problem is acknowledged as a challenge. The random-dot stereogram places some constraints on how that might work, because it shows that you don't have to have interpretable shape.

One idea for helping solve the correspondence problem is to do the matching on low spatial frequencies first. So the idea is that lower spatial frequencies don't have as many possible matches between the two images. And so you can of settle on a good match, maybe with a easier search, and then hone in on a more fine-grained match that would cause the higher spatial frequencies to line up as well. And so that's what's often referred to as a coarse-to-fine strategy. You see this in lots of areas of vision where you first try to solve a problem at coarse spatial scales and then refine the solution with higher spatial frequencies.

So there's lots of funny phenomena that are associated with binocular vision. One that you may have encountered is called the wallpaper illusion. Maybe you haven't encountered this because wallpaper is a thing that people don't do anymore. But the previous generation, they used to decorate their houses with wallpaper. So instead of painting the walls, you'd put paper up on the walls that would have particular patterns in it. So maybe your grandparents had had this. Mine did. And there would typically be these repeating patterns.

And so this is a situation where there are lots of possible matches. So if you're fixating the screen and you-- this particular feature is matched between the left eye and the right eye, then everything is good. But if, by chance, your eyes diverge or converge a little bit because of the ambiguity and this feature ends up being at the same location in the left eye as this in the right eye, then you'll get disparity with respect to the rest of the world. And that will cause the wallpaper to jump out in depth or jump back in depth. So that's the wallpaper illusion. And that'll happen sometimes.

Another thing that you may-- how many people have encountered in autostereogram? Yeah. So these were a really popular thing when I was a kid. So these patterns that allow for the possibility that you could get an aligned image if your eyes converge or diverge a little bit. So again, you look at this in the right way where your eyes diverge a little bit.

So instead of actually fixating the depth of the screen, you're fixating a point that is in back of the screen or in front of the screen. And if you get your eyes in exactly the right position, the images will come into register. And you'll get something that's a little bit like a random-dot dot stereogram, but it just looks different. In this case, it'll look like the shape of a horse. So you might have to look at it for a long time. We can come back to this at the end of class.

All right. So another pretty cool and important phenomena that's related to binocular vision occurs when you can't match the images between the two eyes. So if the two images in the two eyes are too different, you get what is called binocular rivalry. So hopefully the projector is good enough for us to see this.

So the left eye here-- oh, yeah. This is good. The left eye is going to get one orientation, and the right eye is going to get another orientation. And if you look at this, what will happen is that, over time, you will see one image and then another, and then back to the other image, and then back to the other. Is this working for everybody? Yeah? Good. All right. So that's binocular rivalry.

And it's an interesting-- and I can play another example here. This one is a lot of fun to look at. It's interesting because the retinal stimulus is constant, but your percept-- what you're consciously aware of-- will change over time. This one is interesting because there are these two different words, but you may actually mix up the letters from time to time. So it's blue and red, but you might see bled sometimes. Yeah, question?

AUDIENCE: What the glasses work if you were colorblind?

JOSH MCDERMOTT: What's that?

AUDIENCE: If you were colorblind, would this work at all or no?

JOSH MCDERMOTT: I think it should because this is-- yeah, it's an optical effect. So it just relies on the fact that these are filters that are going to absorb certain wavelengths and let others pass. So it doesn't depend on having color vision. Yeah.

All right, so binocular rivalry, it's a thing. It's something that the brain does. It's an interesting case of bistability, right? Really well-known case.

One reason why this is important is it's been a tool that people have used to look at the basis of conscious awareness. So this is a case where you can have the stimulus be completely fixed, but what you will see, what you perceive, fluctuates over time. So you can ask the question, what happens in the brain when your percept changes like this?

And in particular, you could say, record from neurons in different stages of the visual system and ask whether the responses of those neurons is changing when what you see changes, even though the input of the visual system remains the same. And so that's, in fact, something that people have done fairly extensively. And we will talk about that in a couple lectures. Any questions about binocular rivalry or matching? Yeah?

AUDIENCE: This might not be a very answerable question. But if we had eyes like prey that didn't necessarily-- or like older or black, would binocular rivalry what we perceive? Or would we just have a much broader view? What would that be?

JOSH MCDERMOTT: Yeah, that's a great question. So I would have to assume-- I mean, it's a difficult question to answer because we can't really talk to squirrels very easily. But I would have to assume that they just would not experience binocular rivalry and that they just see a much wider field of view without just using monocular depth cues. I mean, it would be dysfunctional, I think, if they experienced rivalry, so yeah. So I think this is something that's probably specific to people that have stereo to organisms that have stereopsis.

So in order to do the experience that I talked about-- I mean, some of those experiments have been done with humans using fMRI by our own Nancy Kanwisher, who some of you may know. But others were done on macaque monkeys, who were reporting what they saw at every moment in time. And prior to engaging in the physiology work, they made sure, as best they could, that the properties of the rivalry that the monkeys were reporting was similar to the rivalry that you see in humans.

And so in particular, if you measure how long it takes to flip between one percept than another, that has pretty established parameters. So there are certain distributions for the duration of each percept, and it changes the stimulus parameters and predictable ways. And so I think they checked that that happens in monkeys as well. So as far as we could tell, or they could tell in those studies, monkeys have a fairly similar experience of binocular rivalry.

OK. Another cool phenomena that is related to binocular vision is luster. So luster occurs when the two eyes receive different luminances. So when that happens, you get this shimmery appearance that's known as luster. So in this particular case, the two middle squares-- the one on the left and the one on the right-- will generate different luminances in the two eyes because one is bluish and one is reddish. And so they'll be filtered differently by the glasses. And you should see those as having a funny shimmery appearance. Does that kind of work? Yeah? OK.

So here's the idea. All right. So these are two different photographs taken from two different positions. The idea is one camera is at the position of the left eye. One camera is at the position of the right eye of some keys. So keys are metal. Metal is shiny and generates specular reflections. So it's not Lambertian.

And the consequence of the specular reflections is that there can be pretty significant differences in the luminance that hits the two eyes just because they're at slightly different angles with respect to the image. So the idea is that what typically causes these differences in luminance would be the fact that there's a shiny surface in the world that is generating a specular reflection.

So here's just those two images. And I can flip back and forth between them. So you can see that the camera's at these two slightly different positions, and there's some slight shifts in the position and then this big change in luminance.

I mean, the perception of shininess is super interesting. But you're getting this cue in the image, which is these two different luminances. You've learned to associate that with material. And then that gives it a certain look, the look of gloss. That's what gloss is in some way.

So you're always fixating something, typically, right? So you fixate a particular object. That induces this horopter. It's the set of all points that project a corresponding points on the two retina-- zero disparity. And so when you fixate an object at a certain depth, everything on the horopter, plus and minus a small amount, will appear to be fused. So you see a single thing. So that's called Panum's fusional area.

So objects that are outside this region will appear as double images. And so you can check this out yourself. So if you fixate one finger and then you move the other finger sufficiently far away, you'll see two of it. So if it's right next to the finger that you fixate, you see one thing. And then you start to see double.

How many of you have noticed that before? A lot of you haven't. So it's this thing that you're always seeing. You're constantly seeing double images. Most of the time, you don't notice them. You have to really pay attention to them.

And so this gets back to this thing that we were talking about a little bit earlier. So why don't you notice this crazy thing about your vision, which is that you have two copies of most of the objects in your world? And it's probably the same reason that you don't notice that the periphery is blurry. It's probably the same reason that you don't notice that some kinds of paintings depict impossible objects.

Whenever you're interested in something, you move your eyes, and you place the thing that you're interested in at the fovea. And then, in this case, because of binocular convergence, the two images line up. You're always fixating the stuff that you want to see. So typically, what you're paying attention to will be fused. And the stuff that you're not fixating won't be fused, but you're not paying attention to it, and you don't notice it.

OK, so how does this stuff work in the brain? So remember that the two eyes is remain segregated at the level of the LGN. So the LGN consists of these six layers. Remember, the top four layers are parvocellular. Bottom two are magnocellular. But then three of those layers are for the left eye, and the other three are for the right eye in each of the two hemifields. So there's strict segregation in the layers of the LGN. A neuron will exclusively get input from the left eye or the right eye.

Now, when you get to V1, the left and the right eye begin to mix. Now, the cells have varying weights from the two eyes. So some are purely monocular. Some have equal contributions.

Remember, we talked about ocular dominance columns. So even though you begin to see binocularity in neurons, it's still the case that at a given point in the cortex, that column will tend to be dominated by one eye or the other. It's called ocular dominance column. So here's an image of looking down at the cortex of where the animal's got one of its eyes open. It's being stimulated with checkerboards or something, probably. And these are the bits of the visual cortex that are activated by that one eye. So those are ocular dominance columns.

But it's nonetheless the case that you see plenty of neurons that get input from both eyes. And some of them are disparity selective. So here's an example of a neuron that is selective for both orientation and disparity. So what you're seeing here is the position of the stimulus in the left eye and the right eye. And so these numbers here are the minutes of arc that separate the images in the two eyes. So that zero, that would correspond to zero disparity-- 15 minutes, 30 minutes, 45, 60, and so forth. And then this is the right eye and the left eye on its own.

So this is an oriented bar, and it's being swept in one direction and another. And this is a trace from a recording. So these will show you the action potentials that result.

So they've already figured out what the preferred orientation is for the cell, and so the bar is fixed at that orientation. But you can see that there is a disparity that elicits the biggest response where the left and the right eye images don't correspond perfectly. So they're offset by a little bit.

And if you present either the right or the left eye on its own, you also get a weak response. So this happens even in early visual cortex. And again, that's consistent with the random-dot stereogram and the idea that stereopsis is something that at least starts early in the visual system.

And so it's easy to imagine how you might wire up a disparity-tuned neuron. So the idea is just that you would have neurons that have receptive fields that are in different locations in the two eyes and that get wired together. You have a thing in the world that happens to project to the right combination of points in the two eyes, and those converge on some downstream neuron that's disparity selective, then you get a response.

All right. So as I was mentioning earlier, there are some fairly common problems that can occur with binocular vision. So strabismus is a condition that people are often born with where the eyes are slightly misaligned. That can produce double vision. And so typically, in the United States, this would be detected very early on and corrected. Sometimes I think it just goes away. Sometimes there's a surgical procedure that is performed to get the eyes in alignment.

But if it's not corrected, you develop something called amblyopia, where one eye suppresses the other. It's typically caused by strabismus. So strabismus is usually correctable by exercises or by surgery, like I said. So this is what it looks like. The eyes just won't quite be at the right relative positions.

And childhood amblyopia is a big deal because if it's not corrected, it can cause permanent losses in binocular vision because it interferes with the development of binocular cells in the visual cortex that are selective for disparity. And so there's a critical period up to about six years of age. So the idea is that if you can correct the problem prior to six years, the person will be able to develop normal binocular vision. But if you wait beyond that, even if it's corrected, there's less plasticity after that point in time, and you don't develop the necessary abilities. Yeah?

AUDIENCE: Under the stereo part, is that the suppressed eye development longer if someone had anybody at 20 years old lost their eye that was suppressing? Do they just continued normal vision without the stereopsy? Or is that the weaker and they'll be able to see less?

JOSH MCDERMOTT: So I don't know. My guess would be that you would be able to recover some other aspects of vision and that it would just be stereo that you would lose, but I don't know for sure. I'm sure the answer is known because this is something that's been pretty well studied. It's just not my area of expertise.

And so there are tests that will actually diagnose whether people have normal binocular vision. You can give them displays of that we've been looking at. And by the standard diagnosis, about 5% of people are what's considered to be stereoblind. So you give them moderate binocular disparities, and they won't be able to perceive depth based on those. So that's a pretty good percentage.

Any questions about binocular vision or stereopsis? Yeah?

AUDIENCE: If you did have amblyopia, for example, like, you were given a 3D image and these glasses have particular distance, would you still be able to have 3D effect? Or are you just not able to see everything with the glasses?

JOSH MCDERMOTT: Yeah. These anaglyphs just wouldn't work. So I mean, you'd presumably have normal monocular depth perception, but you just wouldn't really get the benefit of stereopsis. I mean, these glasses are not doing-- what the glasses are doing is essentially recreating what happens normally, where there's a 3D world, and you get two slightly different images. Because the two eyes are in different positions, then the world is 3D.

So what's happening is that what's useful about the glasses is you can look at a 2D screen and still get disparity. If you're looking at things normally, there would be no disparity looking at the screen because everything is at the same depth. So the anaglyphs are just a way to artificially get disparity that you would normally get from three dimensions.

AUDIENCE: So as a follow-up, does that mean that this type of binocular depth perception is more cognitive than physical?

JOSH MCDERMOTT: No, I think it's pretty automatic. I wouldn't call it cognitive at all. It just happens, and you're not even-- you don't really think about it. Yeah. Yeah?

AUDIENCE: For the stereo blindness, is it only caused by problems with the development of the different context? Or does it also include problems caused by strabismus? Is that a level eye or complex?

JOSH MCDERMOTT: Well, it can be both. So with strabismus, if the eyes kind don't align in the right way, then, yeah, you wouldn't necessarily be able to normally get images that are in register, which is why one-- and the solution to that is that one eye ends up dominating. That's how the system copes with it.

AUDIENCE: So it's like colorblindness that's a cortical cause [INAUDIBLE]?

JOSH MCDERMOTT: Yeah, I would say-- OK, so the way that I would think about this is that there's this cortical mechanism that normally develops if the eyes are positioned in the correct way such that the depth cue is available. So if that depth cue is there in the input, then your cortex develops, and you learn to detect the disparity. Now, if the eyes are misaligned and the depth cue is not available, then the cortex doesn't develop the disparity analysis.

Now, subsequent to that, if you're, say, 10 years old, you can then correct the strabismus. So now you have disparities between the two eyes. So the cue is available, but the cortex is past its critical period for this particular thing. And so the cortex can no longer learn to develop the disparity sensitivity. Make sense?

AUDIENCE: Yeah.

JOSH MCDERMOTT: OK. OK, have a great weekend. Problem set's been extended to Monday, as you probably heard. Have a great weekend, and we'll see you Tuesday.