[SQUEAKING]

[RUSTLING]

[CLICKING]

JOSH MCDERMOTT: All right. Let's talk about perception. So perception is the task of determining what is out there in the world from sensory input. So as an organism, you need to know what's happening around you. And things in the world, they give off different types of clues to their existence. And organisms have sensory organs that detect these clues. And the nature of the clues is different for the different senses.

So in vision, photons that originate at some kind of light source reflect off objects and are absorbed by the eye. And the pattern of photons gives you information about the objects that they reflected off. In hearing, objects cause vibrations in the air. They travel through the air. And they're absorbed and measured by the ear.

With the sense of touch, we bump into things, intentionally or unintentionally. And that stimulates receptors in your skin that respond to pressure and other things. With the sense of taste, we lick things. And molecules that are contained in the substances that we lick interact with the taste receptors in your tongue. And turns out they also interact with the olfactory receptors in your nose, as we'll learn. And with the sense of smell, substances in the world give off molecules that float through the air and interact with the receptors that are in your nose.

OK. So we've got these sensory organs that measure these different types of clues that are coming from the world. Now, the task of perception is to take input from these sensory receptors, and then, with that input, figure out what is out there in the world. So the first important point I want you to take away from this lecture is that perception is deceptively hard.

Now, normally, you just open your eyes or you listen and you effortlessly apprehend the world around you. I look out at you all. And I'm seeing all these chairs, and people in them, and people looking at me. And I don't really have to try to do it most of the time. It just kind of happens. And this often makes it non-obvious, all of the complicated stuff is going on that actually enables us to derive that kind of information about the world.

And one way to get some perspective on this is to actually view the sensory input in a slightly different way. And so here, what we have is an image. But instead of representing it like a normal image, where there will be pigmentation on the screen at a particular point, here we have numbers.

So this is a grayscale image. And the number here represents the gray level at a particular point in the image. So you can think of this as if you snapped a picture with your phone. And if it was a black and white camera, this is what the CCD from your phone camera would output. It's an array of numbers.

And so your task here is to take input like this, and then to figure out what's out there in the world. And so when you look at it like this, it makes it clear that, well, this is actually kind of non-obvious, because normally what happens is when you look at the image, well, your brain is doing its thing. So there's all this complicated stuff that happens. And that causes you to see. But it all starts with this.

Now, when you talk to people on the street and you tell them that you study, for instance, vision, they often think that you must work on the eye. So here's the eye. So the eye is amazing. The eye is this incredible device that evolved to form an image and to measure that image. So we've got a lens there. Light passes through the cornea, and then the lens. And it gets focused onto an image on the back of the eye called the retina.

So this is a close-up schematic view of a little piece of the retina. So the light rays come through here. So one weird thing about the retina is that it actually is, in a certain sense, wired up backwards. And so the light has to pass through a whole bunch of stuff. But then it eventually gets to the things up at the top. Those are the photoreceptors. And so the photoreceptors are a special type of cell that absorb photons, and then take that absorbed photon and turn that into a change in voltage. So they take light and they turn it into an electrical signal.

All right. And so this is happening across the entire retinal image. And so you effectively get a spatial array of voltages. So it's really analogous to this array of numbers here, except that it's happening in your eye. So you can think of your eye as something that is measuring light, turning that into a bunch of numbers, and then sending that through the optic nerve to your brain.

So that's just the start of it, though. So the eyes just measure light. They don't interpret it for you. And that's really the job of the brain. And so one piece of evidence that this is a pretty complicated thing comes from the fact that really a large fraction of the brain is devoted to seeing, so roughly 50%, depending on how you measure it in humans, more in monkeys. There's another pretty big piece that's devoted to hearing. So there's very similar issues with audition. So just listen to this particular sound signal.

[AUDIO PLAYBACK]

[INTERPOSING VOICES]

[END PLAYBACK]

JOSH MCDERMOTT: OK, so what did you just hear? Yeah, just shout it out.

AUDIENCE: Background noise.

JOSH MCDERMOTT: Some background noise? Yeah, there was some background noise. Yeah, what else?

AUDIENCE: A question.

JOSH MCDERMOTT: Yeah, somebody asking a question. Yeah. How many people were talking in that?

AUDIENCE: Two.

AUDIENCE: At least two.

AUDIENCE: Four.

JOSH MCDERMOTT: Two or three. OK, yeah. What kind of-- where do you think that was recorded?

AUDIENCE: A restaurant.

JOSH MCDERMOTT: A restaurant, OK. All right, so just from listening to that, you just immediately know all these things that are happening in the world. But the sensory input that--

[AUDIO PLAYBACK]

[INTERPOSING VOICES]

[END PLAYBACK]

JOSH MCDERMOTT: Oops, kill that. The sensory input that you received was a pressure waveform. So there was a sound signal. It started off in my computer. And it traveled to the speakers in this room. And there was a diaphragm that wiggled back and forth. The sound wave traveled through the air. And it caused pressure variation at your eardrum. And that made your eardrum wiggle back and forth in some particular pattern.

So you can think of this waveform here as the eardrum displacement as a function of time. But really, at some level, it's just a time series of numbers. And so from that time series of numbers, you were able to determine all those things about what was going on in the world to cause that signal when it was recorded. And that's your brain doing its thing.

So the ear, kind of analogous to the eye, is a really remarkable device for transducing the mechanical energy from sound into electrical signals that get sent to the brain. So this is a schematic of the ear. We've got the eardrum here. And then there's an organ called the cochlea that ends up turning the sound energy into electrical signals. But your brain then-- so this is the auditory nerve, analogous to the optic nerve. And your brain then gets the signals from the auditory nerve and does all this complicated stuff to cause you to hear what is there.

All right. The second important point that I want you to come out of this lecture with is that perceptual problems are usually ill-posed. So ill-posed means that there's not enough information to uniquely determine the answer to the problem. And most perceptual problems actually are of this nature.

So one classic example derives from the fact that the world is three dimensional. And we usually are pretty good at correctly perceiving its three-dimensional structure. So I can reach and grab this coffee cup. And that requires that I know the shape of the cup. Otherwise, I'd knock it over. Or my hand wouldn't be closed tight enough, and so forth. And that's primarily something that happens visually. Similarly, I know roughly how far away each of you are from me and so forth. So we're pretty good at perceiving three-dimensional structure.

But the input to perception, in particular to vision, is two-dimensional. We form an image on the back of our eye. Now, we have two images, one for the left and one for the right. And that actually is part of the solution. But you can close one eye and depth perception is still pretty good right.

All right. So the depth information that is there in the third dimension is lost in the projection that forms images. And so there's lots of different shapes, for instance, that have the same three-dimensional-- have the same 2D projection onto an image. So each of these is like a different three-dimensional shape, but they line up in just the right way and would all cause the same image.

There's lots of other examples. This is another one that's kind of interesting. So this is a bunch of moving dots. And the same dot motion is consistent with a whole bunch of different possible objects. So here, we have triangles. Here, we've got some squares. This is something else. This is something else. So it's exactly the same the same motions. But they could be grouped differently and either represent fixed points of objects or things that are moving along other things.

Another classic example is that of auditory scenes. So usually in the world, there will be more than one thing that's making sound at the same time. And the vibrations that are caused by different sound sources-- so my voice and the rustling of your neighbor turning the page of their notes, for instance-- they sum together at the ear. So the signal that you get at your ear is a mixture of the sounds that would have been caused by the individual events on their own.

But as an organism, typically, what you want to hear and what you need to hear are the individual sounds. You need to understand what I'm saying. You want to understand whether somebody is turning the page, or walking close to you, or whatever it may be. And so the problem that you really have to solve there is akin to me giving you this equation and asking you to solve for x.

And so if I put that on the exam, you're all going to complain because there isn't a unique solution. There's lots of different combinations of x and y that could sum to the same number-- one equation, two unknowns. But that's exactly the problem that is happening when you have a mixture of sounds and you have to understand one or more of them.

And so somehow, in the case of this auditory scene problem, we can usually hear the constituent sounds reasonably well. And maybe the classic example of this is what's called the cocktail party problem. How many people know what movie this is from?

AUDIENCE: *Breakfast at Tiffany's.*

JOSH MCDERMOTT: *Breakfast at Tiffany's.* Yeah, you got it. Yeah, good. Yeah, it's classic, Audrey Hepburn. So at a cocktail party, you're often trying to talk to somebody. So there's somebody who you want to understand. And so maybe it's somebody saying this--

[AUDIO PLAYBACK]

- She argues with her sister.

[END PLAYBACK]

JOSH MCDERMOTT: And next to that is a picture. It's a way of turning a sound into a picture, very similar to what's called a spectrogram. So we have frequency on the y-axis and time on the x-axis. So there's all this structure in that speech signal that allows you to understand what the person was trying to say. But the problem that you might encounter is that what actually enters your ears might be this.

[AUDIO PLAYBACK]

[INTERPOSING VOICES]

[END PLAYBACK]

JOSH MCDERMOTT: There's another person talking there. And so now, that picture is kind of complicated, because it's got these two sound signals that are on top of each other. But it could also be that there's even more people talking.

[AUDIO PLAYBACK]

[INTERPOSING VOICES]

[END PLAYBACK]

JOSH MCDERMOTT: Or even seven other people talking--

[AUDIO PLAYBACK]

[INTERPOSING VOICES]

[END PLAYBACK]

JOSH MCDERMOTT: So by the time you get down to here, it's really a pretty serious mess. But you could hopefully probably tell that you had a pretty good ability to actually hear out that target voice throughout these examples. So in general, in many cases, speech remains intelligible, despite the presence of other speakers. This is a problem that humans still solve substantially better than machines. So present day speech recognition algorithms, like in your iPhone, they work pretty great now, if you're in a quiet room. But in a situation where there's a lot of other people talking, they'll typically still perform poorly.

OK, so that's another example of an ill-posed problem. So one of the amazing things about perception is that, despite the fact that we're constantly confronted with these ill-posed problems, which means that there's not usually a unique solution, usually we arrive at a single unambiguous interpretation of a stimulus. So you just open your eyes. And you see what's there. And it's usually correct. Again, this is what enables you to pick things up, and avoid running into things, and so forth.

There are, however, interesting cases where perception can be ambiguous. And those often suddenly give you this insight into, oh yeah, it's not always completely determined. So this kind of looks like a cat. Wait, actually, no, hang on. That's a crow, right? I don't know which it is. It could be either. This one kind of makes your brain hurt. Which way is this person facing?

AUDIENCE: I'm going to say they're facing forward and holding up a cutout piece of paper.

JOSH MCDERMOTT: It could be, yeah. It could be. But I kind of feel like the person's looking that way. Wait, actually, no, they're looking that way. Yeah. And if you look at this for a while, it may. It may change. What is this?

AUDIENCE: [INAUDIBLE]

JOSH MCDERMOTT: Is it a dog or is it a person?

[INTERPOSING VOICES]

JOSH MCDERMOTT: And are this person's legs shiny?

[INTERPOSING VOICES]

JOSH MCDERMOTT: How many people think this looks shiny? Yeah? But I think what actually happened is this is suntan lotion. So these, again, are interesting examples in the sense that they highlight the fact that these problems are ill-posed, because there are these two interpretations of the image, things that could actually be happening in the world, that are consistent with the image.

So this person could have put suntan lotion on their legs. Or their legs could be like wrapped in cellophane. This could be a photo of a dog or of a person running into the woods with some kind of backpack on. This could be a face looking forward or a face to the side and so on and so forth.

So there's also these unusual cases where individuals may be confident in their interpretation, but they'll disagree with other people. So this was an example that took the internet by storm several years ago now. It's the dress. And there were all these fights over whether the dress is actually white and gold or blue and black. How many people think it's white and gold? How many people think it's blue and black? OK, yeah. Disagreements.

So this is actually really interesting because the fact that people disagree about this indicates or suggests that there are different assumptions that the white/gold people are making about the world compared with the black/blue. And these assumptions are made to resolve the ambiguity.

And in fact, that's one of the central themes of perception, is that the way that we are able to solve these problems is by making assumptions about what the world is like. And that constrains the solution space hopefully enough that you can reach a unique solution.

The third important point I want you to walk away from this lecture with is that perceptual systems must be invariant. And this is because the sensory input that is caused by a single type of thing in the world typically varies enormously. So these are all images of a car-- a car viewed from different distances, and different viewpoints, and on different backgrounds. They're all cars. But the actual image that is here and here and here is totally different.

So that array of numbers-- If you think of this, that image as an array of numbers, it's going to be totally different. So the question is, how do you actually build something that can take that array of numbers and tell you that that is a car given that the numbers are changing so much from this instance to this instance to this instance? So that's one of the key problems of invariance. And so somehow or another, you're immediately able to look at these things and tell each of these things is a car and each of these things is not.

The same problem exists in speech. So I'm going to play you an example of what we call dry speech, so without a whole lot of reverberation.

[AUDIO PLAYBACK]

- They ate the lemon pie. Father forgot the bread.

[END PLAYBACK]

JOSH MCDERMOTT: So this now, the second thing is going to be speech and reverberation.

[AUDIO PLAYBACK]

- They ate the lemon pie. Father forgot the bread.

[END PLAYBACK]

JOSH MCDERMOTT: That's like in a subway station or something or a really big bathroom. The sound waveforms there are now very, very different. If you actually look at spectrograms, you can see the effect of the reverberation is to smear the structure out in time. So it's like a massive distortion. It's totally different sound signals. But you can listen to them and tell that they're saying the same thing.

Now, the ill-posed nature of the problems that your perceptual systems have to solve and the difficulty of acquiring the right type of invariance is why perceptual problems present a computational challenge. And so for really many decades, these were computational challenges that were insurmountable.

So I taught this class when I was a PhD student, which was in the early 2000s. And one of the things that you would always comment on is how amazing human perceptual systems are compared to computer vision systems or speech recognition systems and so forth. And things have changed a little bit.

So in particular, in the last five to 10 years, contemporary machine perception systems have become pretty good at certain types of perceptual tasks, specifically classification tasks. So object and face recognition now work pretty well. There's arguments over do they work as well as humans. Those are interesting questions we could talk about. Speech recognition works pretty well. Again, most of us talk to our phones all the time. You can dictate emails, and texts, and things like that. So that's remarkable and a game changer.

And of course, that's been a big deal in the world of engineering. It's also been a very interesting development for the study of perception, because the resulting systems now give-- you can treat those as models of perceptual systems. And they exhibit many interesting parallels with human perception. So this has given rise to a new generation of models of perceptual systems.

And so that's going to be a theme that we will talk about periodically throughout the class-- namely, can we obtain better models of the brain, in particular of sensory systems in the brain, using contemporary technology from AI? So one of the main engines of all of this is artificial neural networks.

So these are systems that consist of the repeated application of pretty simple operations, all of which were loosely inspired by things that people saw in the brain-- filtering, pooling, normalization. We'll talk more about what each of those things means. And we now have really effective methods to optimize the parameters of systems like this to cause them to correctly classify input signals.

So just to give you a little example of something that was really unimaginable back when I was a student, I'm going to show you a comparison of speech recognition by humans and by an artificial neural network. And so this is an experiment where humans are played, short excerpts of speech superimposed on background noise. And they just have to say what the words are that the person is saying.

So the y-axis here plots the proportion of words. And the x-axis is the signal to noise ratio. So as you move from left to right, the speech becomes louder relative to the background noise. And so you expect that people will get better. And indeed, they do. But you can see that, in this experiment, there were four different types of background noise. So the green one is music. For instance, the purple one is what's called speech babble. That's like crowd noise. And you can see that some types of noise for humans are much easier to recognize speech in than other types of noise. So that's just what people do.

And so next to it, this is the results of running a neural network model on the exact same experiment. And there's two main things to take away from this. One is that the model is doing about as well as humans. That's the thing that was inconceivable 15 years ago. And nowadays, it's pretty commonplace.

But the other thing that's interesting is that the conditions that are easy for humans are easy for the model and vice versa. So the phenotype of speech recognition seems to be shared across humans in this model. And so the question is, well, can we use these things to model sensory systems? And that's just something that we will talk about throughout the class.

AUDIENCE: I have a question.

JOSH MCDERMOTT: Yes?

AUDIENCE: You mentioned that 50% of the brain is [INAUDIBLE]. Is that by mass or volume? And is that the entire brain or just the neocortex?

JOSH MCDERMOTT: I was actually talking about the cortex. That's a very approximate number. And it would be both mass and volume. So mostly, it's pretty much the back half of the brain, more or less-- I mean, this is very, very crude-- more or less is involved in vision.

And again, it's a little complicated to give you a very precise number there, because the question is, what does it mean to be involved in vision? And there's lots of parts of the brain that respond when you're looking at things, but have other functions as well. But yeah, roughly half is what I would say. Thanks for the question.

All right. So the fourth important point that I want you to walk away from is that perception is unconscious inference. All right. So we've talked about how one of the key things to know about perception is that the problems that we're solving are ill-posed. So that usually means there's not like a unique solution.

So the information in the sensory input does not uniquely specify the structure of the world. And so the consequence of that is that the brain has to make its best guess as to what is out there. This is inference. So when you see or when you hear, we think your brain is choosing the most probable interpretation of the sensory input that you are getting.

Now, so this is inference. But you're not aware of the inference. So it's very different from-- some kinds of inferences you make consciously. You might reason about a problem to work out what might likely have happened to explain something. So the inferences that your perceptual systems make, they just happen automatically, really without you being aware of them.

And so that's why we call it unconscious inference. So that is a term that is due to Helmholtz. So Helmholtz was a giant of 19th century science, did lots of stuff, and had made many important contributions to perception, including this idea that perception is unconscious inference.

So let me give you an example or two of that. So I'm going to play you a bunch of moving dots. And so this is a little bit like that thing that I showed you a little bit earlier in the sense that the moving dots-- because really all that you can see there are these moving dots, there's lots of potential explanations of the motion. OK? Shh.

So you look at this. And you can see the dots moving. It may not really be completely obvious what caused those dots. But now, what I'm going to do is show you the same thing flipped upside down. And when you see it in this orientation, it becomes quite obvious that this is a person walking.

And so in fact, this stimulus is an example of a very famous type of stimulus known as a point light walker. And so originally, the way that they would generate something like this is by putting these little lights on the joints of a person, and putting them in a dark room, and then filming them. So now, of course, we can do this with computers. And so the really remarkable thing is that, when the orientation is correct, which means what you're used to, you can take that pattern of motion, and then perceive a form. But when it's not what you're used to, that's a lot harder to do.

So what does this mean? Well, OK, so it's an example of an ill-posed problem, because there are lots of potential explanations for this motion. So one explanation is that there's a person walking upside down. And we verified that because you saw it right side up. And you can tell that could be caused by a person.

But another explanation is that these dots are just moving around. Maybe they're fireflies or something. They're not necessarily related to a single thing. And there are these multiple interpretations. And then your brain is choosing one of those interpretations based on what it thinks is likely.

And so when there's an interpretation that there's a person that's right side up, that seems to be what you see, presumably because usually when people walk, they're not walking on the ceiling. They're walking on the ground. And you're mostly right side up. And so most of the time, you're seeing people walk right side up. So that, in some sense, is an unconscious inference. Any questions about that? Yeah?

AUDIENCE: Is there the nature/nurture thing going on in the fact that unconscious versus conscious-- is any of this things that we expect are hardwired versus learned by experience of like watching thousands of people walk?

JOSH MCDERMOTT: Yeah, that's a great question. So you will be able to ask that question of many of the things that we will talk about in this class. And in most cases, we don't really know the answer very clearly. So there's bits of evidence that young infants can experience some of the kinds of phenomena that we'll talk about. It's not that easy to actually determine whether a baby can tell if this is somebody walking because they can't talk and so on and so forth. So you have to do really clever experiments to assess that.

These assumptions that your brain brings to bear to constrain these ill-posed problems, they could be something that you acquire through evolution, that you're born with. It could be something that you acquire over development. It's probably some of both. In most cases, we don't really know. Yeah.

There are a handful of cases where there is very clear evidence that you can learn assumptions. So part of why this is hard to study is it's not that easy to have somebody inhabit a world where people walk on the ceiling. So just in practice, the interventions that you would need to do the experiments are impractical. But there are cases where those interventions are less impractical. And in some of those cases, people have done them. And we'll talk about some of those.

So here's another interesting example of an unconscious inference. So I'm going to play you a sound that consists of something called a sound texture. This is the sound of a lot of people clapping, applause. And then that will be interrupted by noise. And then there'll be a little bit more of the texture. And the amazing thing about this particular sound is you're going to listen to this and you will have the sense that texture continues during the noise. But in fact, it's completely not present. So listen to this and see what you think.

[AUDIO PLAYBACK]

[APPLAUSE]

[STATIC NOISE]

[APPLAUSE]

[END PLAYBACK]

JOSH MCDERMOTT: Did you hear it continue? Yeah? Again.

[AUDIO PLAYBACK]

[APPLAUSE]

[STATIC NOISE]

[APPLAUSE]

[END PLAYBACK]

JOSH MCDERMOTT: So it's so crazy that you might not believe me. You might say, well, hang on, how do I know that that thing is not actually there? So this is a variant that will hopefully convince you of this. So in this particular case, it's exactly the same. But there's a little gap here between this texture and the noise. And so now, I think your experience will be very different.

[AUDIO PLAYBACK]

[APPLAUSE]

[STATIC NOISE]

[APPLAUSE]

[END PLAYBACK]

JOSH MCDERMOTT: So it wasn't there then. All right, so what's going on? Well, it seems to be the case that this, what we call, illusory continuity-- the fact that you hear that sound continue during the noise-- one second-- is an unconscious inference about what most likely is happening during the noise. You had a question. Yeah?

AUDIENCE: So you can hear the texture in the noise. But why can't you hear the noise in the texture afterwards?

JOSH MCDERMOTT: That's a very good question. The presumptive reason is that the noise is actually more intense. So the sound level or the intensity or the loudness is higher. And so what we think actually happens in these examples of illusory continuity is that the reason that you actually hear the texture during the noise in this example is that the noise is sufficiently intense that a phenomena called masking would occur if the texture were there.

And specifically, what that means is that when two sounds play at the same time, if one of them is loud enough relative to the other, you will not be able to detect the presence of the other one. That's known as masking. But that's just a phenomenon. And so the consequence of that is that, when the noise is very high in intensity, you actually wouldn't-- the stimulus would not be any different whether or not the texture actually was there.

And when I say it would not be any different, it means it would not be any different from the standpoint of the ear. The stimulation at the ear would be the same due to this phenomenon of masking. And the consequence of that is that this is the situation is ill-posed. So the stimulus that you would get if the texture was there is basically the same as the stimulus that you would get if it wasn't there.

And so now, the brain just has to make its best guess as to what is happening. And presumably with these texture sounds, they're of that they just tend to go on and on and on and on. And so your brain infers that it's a good bet that that sound actually continued. You had a question there, and then there.

AUDIENCE: Yeah, I was wondering, does the type of noise-- let's say it was a trumpet instead. Would you still have the same phenomenon, like if the texture was really-- let's say clapping, but it's a different type or intensity of noise?

JOSH MCDERMOTT: So are you asking whether the effect would be the same if this was a trumpet?

AUDIENCE: No, the noise. Like, if the noise was not the static that it was.

JOSH MCDERMOTT: Yeah, OK. Got it. So the thing that seems to be really important for this illusion to happen is that the frequencies in the noise have to overlap with the frequencies in the textures. And they have to be kind of loud enough that they could mask the texture.

And so in fact, what happens is that if you actually make the noise quiet, if you lower the amplitude of the noise, at some point, this stops happening. And you stop hearing the texture there. You had a question, [? Hannah? ?] In the back?

AUDIENCE: If you were to just play the texture and then play the noise for a long time, would you eventually stop hearing the texture?

JOSH MCDERMOTT: Yeah. In fact, so Richard McWalter, who was a postdoc in my lab, he did these experiments. And so the amount of time that the texture lasts depends on the texture. But in some cases, it can go on for three or four seconds. But it fades out eventually. Yeah, yeah. Oh, yeah?

AUDIENCE: This question is asking about priming. And it reads, would I hear the same clapping in the noise if Josh didn't tell me that you will hear in the classroom? Thinking back to before page 25 of the slides, when I was listening to the 7-people overlay, I didn't hear the initial sentence before realizing that the sentence was uttered when Josh said you can still hear the sentence.

JOSH MCDERMOTT: Yeah, good question. So for this phenomenon, the priming-- I believe the priming has a very small effect. And the reason I think that we've done lots of experiments where we didn't tell people that they were going to hear this thing in the noise. In fact, we don't even tell them that the texture is not there. Sometimes, the texture is there. Sometimes, it's not. And we ask them whether it's there. And when it's not there, they say it's there essentially. So the illusion is real in that sense.

And it's certainly true that, with that cocktail party effect, part of what helps you hear what we call the target voice-- so the very first one that I played. Part of what helps you hear that is that you heard it in isolation the first time. So that is priming. So in that situation, the priming actually does have a fairly substantial effect. So that can definitely be important in a lot of settings.

So the fifth important point I want you to leave the lecture with is that illusions illustrate perceptual mechanisms at work and can help us study them. So illusions are fun. We love looking at illusions because they make us realize that the world is not always how it seems. And they're lots of fun. But they're are also scientific tools. And so a lot of what happens in perception research actually utilizes illusion. That's partly why studying perception is so much fun. And so throughout this class, we're going to be looking and listening to and feeling a lot of perceptual illusions. And you'll be making some yourself.

So this is an example. This is a classic illusion, one of my favorites, where shadows are being used to manipulate your perception of depth and thus your perception of motion. And so what is going to happen here when I play this is you will see this ball move across the screen. It'll go from here to here.

And you're going to see two different versions of this. And the versions will be differentiated by the trajectory of this dark spot that is supposed to be-- it's supposed to look like a shadow. And so in one case, the shadow is going to move in one direction. And in the other case, it's going to move in another direction. And that will cause you to perceive a totally different trajectory for this ball, even though the physical trajectory along the image is exactly the same. So check it out. And I'll loop this a couple times.

So it's on the floor, right? Whoa. All right, let's do it again. On the floor, up in the air. All right. So what's going on here? So I think I love this particular effect because it illustrates a whole bunch of important things.

So one of the things that illustrates is the ill-posedness of depth perception. So the image and the trajectory of the ball is exactly the same for those two totally different trajectories in the world. Three-dimensional trajectories, they project to the same two-dimensional image sequence.

However, there is a relationship between the location of a shadow and an object. So in general, if something is sitting on a surface, the shadow that it casts will be right next to that object, just like the geometry of the way the optics work. If something is way off of a surface, then the shadow will be more distant in the image.

And so your brain, either over evolution or development, has learned the relationship between shadows and depth. And it just automatically estimates the three-dimensional structure of the world using the shadow and its knowledge of the way that optics and geometry work.

And this illusion really-- it demonstrates the role of shadows in depth perception in a way that you wouldn't really have known otherwise, because you have this stimulus that's otherwise the same. And you just manipulate the shadow. And it changes what you see.

Another great example of this is lightness perception. So as organisms, we typically want to infer what things in the world are made of. And that is partially signaled by their pigmentation. Now, the problem that we face is that the light that reaches your eye from an object, it depends not just on the pigment of things in the world, but also on the amount of illumination.

So right now, the light level in here is kind of moderate. And so this surface is white. And I'm getting a certain amount of light that is reflected off of it. That's a function of the fact that this thing is white. And that means that it reflects a high proportion of light, but also the fact that the illumination here is modest. So if we walked out into the center of the atrium, where the light level is much, much higher, the number of photons that will be coming off of the surface will go way up.

Now, the amazing thing about your visual system is that the color of this thing doesn't change by and large. Most of the time, we have what's called color constancy or lightness constancy. And so in order to do that, your visual system has to somehow discount the illumination. And so illusions are one way that people have gotten a lot of insight into this.

And so my PhD advisor, Ted Adelson, did a lot of the pioneering work on this. And this is one of my favorites. So this is an illusion because these ovals and these ovals are the same shade of gray. You look skeptical. So they are. And so this little test patch can be dragged over here, and then dragged up here to verify that they are actually the same. Do people actually want to-- do you want to see that?

AUDIENCE: Yes.

JOSH MCDERMOTT: You don't believe me? OK, well, we're going to have to satisfy your curiosity. All right, so let's see if I can do this. So we've got our test patch here. That's the same, right? OK. Oh. I tell you the truth, right? I only say true things in this class. All right. So why do they look different? Yeah?

AUDIENCE: Because we make a lot of inferences based on shadows. And if we think that even in the shadow the bottom part is so light, we automatically associate higher than [INAUDIBLE] the shadow was typically with [INAUDIBLE]?

JOSH MCDERMOTT: Yeah, so that's basically the way that we think about this, that this contains a lot of evidence that the illumination level down here is very, very low. And there's a lot of evidence that the illumination there is actually very high. So it's on the top side of this thing. You can see a little highlight here indicates there's a light source. This is in the lower region of this thing. So all this evidence for different illumination levels.

And so the same physical amount of light in these two different cases is best explained by their being very different colors of paint on the object in these two settings. So again, it's an illusion. Why is it an illusion? It's an illusion because these things are physically the same, but they look really different. And it teaches us something about the way your visual system works.

And so you often think of illusions as, in some sense, your perceptual system is making a mistake. But as a perceptual scientist, we typically think that the illusion represents a sensible engineering solution that is in your brain that most of the time is going to cause you to see the world as it is. So we think that the reason that this works the way that it does is, most of the time, this is actually helping you to correctly tell that this thing is white. Yeah?

AUDIENCE: I'm just interested in knowing or seeing your opinion on how much of the other information plays a role in causing dilution because there are squiggly lines around it. And there's stuff under, on the ground. And so how much did that really contribute to-- if it was just a plain oval, would we still have that sense of illusion?

JOSH MCDERMOTT: Yeah, so Ted and other people who have worked on these things have looked at things like that. And in general, there's lots and lots of little cues in these displays that add up to create this, like incredible illusion. And so if you take away some of those clues, it'll get a little bit weaker and weaker and weaker.

In fact, when we start, we'll have a lecture on the problem of lightness perception in a couple of months. And we'll see the original version of this that was discovered 150 years ago is way simpler. It's essentially you take the same gray square and you surround it either with white or something dark. And there, you get a small effect. That's called simultaneous contrast. And this effect is huge because there's lots and lots of these other cues that have been added to the thing. And so your brain is just using all the different bits of information it can when it does this, as far as we can tell. Yeah?

AUDIENCE: And actually, going off of her question, I'm wondering, is there-- because for the top dots, they're on the light portion of the wavy lines, whereas for the bottom dots, they correspond to a dark portion. Does that contribute anything? Or is it unrelated?

JOSH MCDERMOTT: Yeah, no. That's one of those cues that will have some effect. Yeah, so there's these local cues. There's these geometric cues related to shape, like the fact that there's this blurriness here because a lot of shadows tend to be blurry. That also helps. There's a lot of stuff that's working towards this.

OK, here's another example in the motion domain, which I will show you real quick. So here, we have some line segments. And it looks like there are these two pairs of line segments that are moving more or less independently. So what I'm going to do is I'm going to add some other shapes to this display. And it's going to totally change the motion that you see.

So now, you look at this and you probably see a diamond that's moving around in a circle. Are people seeing that, more or less? Yeah? OK. And so if I get rid of them, you're going to go back to seeing these line segments.

So again, it's another example of an ill-posed problem because it's the same image motion. And it can be explained in two different ways. It can be explained by these two pairs of line segments that are moving separately or by a single shape that's moving around in a circle.

And in these two different settings, your brain is preferentially choosing one of those two interpretations, presumably based on what it thinks is most likely. So in this particular case, you've got these surfaces that could be occluding the corner of a diamond. And so the diamond perception seems pretty reasonable. But without that, it seems like the most likely explanation is something different.

And so again, what do we learn from this particular illusion? Well, this shows us that, when interpreting motion, your brain seems to be taking into account information about occlusion, what other surfaces there are, and their potential depth relationships to things.

OK. So to give you a firsthand taste of the role of illusions and perception, we have these illusion laboratories. This is a web page that we made for the version of this a couple of years ago. And so what you will do is you'll be in these small teams. And you will make your own illusion to try to answer some kind of question about perception.

And so usually, what people do is they riff on things that they encounter during the class. And so over the course of the class, you'll encounter lots of these effects. And so many of you have asked me questions about how these things work. Well, that's an experiment that you can do. So you can try this out yourself. And then you look at it or you listen to it. And you ask your friends. And then you learn something about how you see or hear.

AUDIENCE: I have a question.

JOSH MCDERMOTT: Yeah?

AUDIENCE: Would a person who has an unknown broken link between sound input and sound processing-- so someone that could generally hear fine, but maybe misinterprets what they're hearing-- would they still be able to perceive an illusion the same as a hearing person?

JOSH MCDERMOTT: I don't know. So sorry, the question is somebody whose ears work fine, but maybe the downstream auditory system is not working normally? So yeah, to my knowledge, that has not really been studied very much.

The main thing-- I mean, one of the main things that can cause the downstream auditory system to malfunction is brain damage that would typically happen after a stroke or if you hit your head really hard, potentially. And so there is lots of instances where people will report difficulty understanding speech.

And it might well be that if you tested people on some of these kinds of examples, that would also work differently. But in practice, most of the time, you're worried about trying to get them to communicate better. And so you focus on the speech deficit. So I'm not aware of that having been studied very much, to be honest.

OK. So why would you want to study perception? Well, one thing that's kind of cool about it is it's very relevant to everyday life. So there's just lots of things that you experience on a daily basis that really relate to perception. And then as a consequence of that, everyday experience can give you insights. You'll be walking around. And you'll see something. And it looks a little bit funny. And then you realize that it's not necessarily how it seems.

There's also lots of exciting applications. So one important application domain is prosthetics. So we're going to talk about this in a few weeks. But lots of people lose their hearing as they age, in particular. And so hearing aids are a huge industry. They work OK, but not great. We'd like to make them work better. There's also cochlear implants for someone who's deaf, retinal implants for someone who's blind.

There's lots of exciting new directions in all of those domains because of some of the developments that have been happening in terms of computational models of sensory systems. It's also very relevant to the design of displays. Right now, in particular, it's kind of an exciting time to be working on perception because there's lots of interesting analogs to AI systems.

And the other thing to note is that sensory systems, they have homologues, typically, in non-human animals. So other things that we do with our brains are fairly uniquely human. Most animals don't talk. It's not clear how much they think. But most animals see, and hear, and smell, and taste, and touch. So sensory systems have classically been model systems for neuroscience. And so knowing about perception is important in that sense.

All right. So just to summarize what we talked about, we talked about how perceptual systems measure energy from different sources in the world, and then try to infer what caused that in the world. We talked about how perception is deceptively hard. So you've got these sensory organs that measure energy. But then the rest of your brain is there to really take that information and figure out what caused it in the world.

We talked about how perceptual problems are usually ill-posed. That means there's typically not a unique solution. And as a consequence, perception is unconscious inference, where you are unconsciously trying to choose what we think is the best explanation of the sensory data that you get. And we talked about how illusions illustrate perceptual mechanisms at work. So they're fun to look at, but they're also powerful tools that we can use to help understand your brain. Any questions about what I've talked about so far? Yeah?

AUDIENCE: Well, we talked about perceptions of unconscious inference. But sometimes, especially with illusions, sometimes when we notice an illusion or see it, we can understand it consciously as well, like in a different way. And is that why it's in parentheses, like unconscious inference? Because sometimes it is kind of conscious?

JOSH MCDERMOTT: That's not what I meant. I was just-- yeah, I could take away the parentheses. And that would be equally appropriate. But the point is perception is inference. But it's a particular type of inference that is typically unconscious. And so sometimes when you're made aware of an illusion, you can then think about it and try to understand it.

They're often actually quite hard to understand. And part of that is because the perceptual systems are often fairly, what we call, encapsulated. You don't really have a tremendous amount of conscious insight into why does that look light and why does that look dark. You have to think about it and stuff.

So that's just an interesting characteristic of perception, that you're-- yeah, it's often fairly impenetrable. But there are interesting cases of what people will often refer to as top-down effects, where maybe what you think then influences what you see or hear. So there are interesting cases of that, yeah.

OK. So the sense of hearing exists to measure sound and use it to infer what's happening in the ear around you. So sound is produced when objects in the world vibrate. They transmit acoustic energy through a surrounding medium in the form of a wave. The task of the ears is to measure sound and transmit that to the brain as an electrical signal. And the task of the brain is to interpret that signal, and then to use it to figure out what is out there in the world.

So sound waves are longitudinal. So that means they consist of regions of high and low pressure that move away from the sound source. Here, we have a tuning fork that's been whacked. So it's vibrating. And there'll be this longitudinal wave that moves away from it with regions of compression and rarefaction.

All right. So some facts about sound-- so one is that it needs a medium to carry it. How many people have seen this movie? Every year, it's a couple. Yeah, you should all go watch this movie. It's one of my favorites, OK? So *Alien* is a movie that came out in 1979. And it takes place far in the future. There's a whole bunch of people that are on this spaceship.

And they get an alert beacon from this planet they're passing. And they go down there and some bad things happen. And they end up with this monster on the spaceship. And then they have to deal with that. And lots of bad things happen. And so this is the poster that was originally used to publicize the movie. And you can see that it's got this slogan, "In space, no one can hear you scream." So why is that?

AUDIENCE: [INAUDIBLE]

JOSH MCDERMOTT: There's no medium. That's right. Space is a vacuum. So there's no sound in space. So they won't be able to hear you scream. So sound needs a medium to carry it. So sound travels through the medium. The speed of sound is proportional to the density of a medium. So sound in air at room temperature moves at 343 meters per second. So in water, because water is denser, the speed is faster, 1,500 meters per second. In solid objects, it gets really fast.

So the intensity of a sound is the energy that's transmitted per second through a unit area. But the speed is independent of a sound's intensity. So whether you speak quietly or yell really loudly, the sound will travel at the same rate away from you. What does change with distance is intensity. So it's often approximately the case that, when something is making sound, the waves that travel away from the source are approximately spherical.

And so the consequence of that is that, as you move away from the source, by conservation of energy, the energy that's in the sphere of remains the same. But the sphere gets bigger. And so the energy per unit area drops as you move away from the source. And it drops with the square of the distance because the area of that sphere is proportional to the square of the distance. And so that's known as the inverse square law. So there are these very predictable relationships between how close you are to something and how loud it is.

All right. So the last thing we're going to talk about today is the decibel scale. So sound, sound level specifically, is measured in decibels. So one of the remarkable things about the auditory system is that it can deal with a huge range of sound intensities. So typically, instead of talking about intensity directly, we use a logarithmic scale of the ratio between two sound intensities.

So a bel is defined as a ratio of 10 to 1. So if you want to know the number of bels that differentiate two sounds-- one with having pressure P1 and the other having pressure P0-- you take the log to the base 10 of their ratio. So it turns out that bels are impractically large. So a lot of the differences that actually matter in sound are smaller than that. And so we use the decibel scale instead. So a decibel is a tenth of a bel. And so the number of decibels by which two sounds are different-- the pressures P1 and P2-- is 10 to the log 10 of that ratio.

So if you increase the sound level by 10 decibels-- this stands for power, not pressure. I misspoke. So these are power measurements. So if you increase the sound level by 10 dB, 10 decibels, that means a 10-fold increase in power because log to the base 10 of 10 is 1. 20 dB is a 100-fold increase in power.

So typically, we will describe particular things in terms of their sound level. So I might say, well, I went to a rock concert last night. And it was 120 dB. Whenever we say that, we are implicitly measuring things with reference to an agreed upon reference sound level. Here, it's called P0.

So there is a Bureau of Standards that's responsible for all these kinds of references. And if you go to the Bureau of Standards and you look up this reference, you'll find there's this particular number that's been chosen as the reference level for what are called sound pressure level measurements. And that particular value, P0, is chosen to be close to the minimum detectable sound level for humans.

So 0 dB is defined to be that reference sound level. So if you have a sound that is at 0 dB, that means that ratio in there is 1 Because log of 1 is 0. And so that means that sound is at that power level. And so that's intended to be the quietest thing that you can hear. Of course, this is all approximate, because everybody's hearing's a little bit different. But roughly speaking, something that is 0 dB SPL is supposed to be just the quietest thing that you could detect under perfect conditions if you were in a really quiet room.

And so when we measure sound levels in this way with respect to that standard reference-- so if you have a sound meter, like we have in our lab-- we will say that the sound is, in this case, 120 dB SPL. So when you say dB SPL, it means that you're using this special reference that is close to the threshold of human hearing.

All right. So here are some example SPL values. As we said, 0 is designed to be close to the threshold of hearing. Normal breathing would be around 10 dB. A very, very soft whisper would be 30 dB; quiet conversation, 50 dB; busy traffic on Massachusetts Avenue, for instance, might be around 70 dB. Somebody shouts right next to you, it'd be close to maybe 90 dB.

Once you get into the hundreds of decibels, you're getting sound levels where prolonged exposure can cause hearing loss, where you should really be wearing earplugs. 120 decibels is a propeller plane at takeoff. And 140 decibels is roughly the sound level of a jet at takeoff. And supposedly, that's the threshold of pain. I've never experienced that myself. And I don't intend to.

So this is the last thing that we're going to talk about. And then I'm going to end. So another reason to use the decibel scale, and the main one that is very, very common in perceptual science, is that human discrimination thresholds-- so that means the smallest change in intensity that you can detect. Human discrimination thresholds are roughly constant when you measure them in decibels. And usually, they're on the order of 1 decibel in optimal conditions.

So what does that mean? That means that if we have a pretty quiet sound, so like, say, 40 or dB, you'll just be able-- so if it's 40 dB, you'll just be able to detect a change in intensity if I move it up to 41. Similarly, if I have a pretty loud sound, 90 dB, you'll just be able to detect a change in intensity if I move it up to 91 dB. And so the decibel scale ends up being pretty convenient because the changes that you can detect are pretty constant as a function of sound intensity. All right. So I'm going to quickly play you some demos that illustrate the decibel scale. Here you go.

[AUDIO PLAYBACK]

- The decibel scale. Broadband noise is reduced in 10 steps of six decibels. Demonstrations are repeated once.

[END PLAYBACK]

JOSH MCDERMOTT: Oops. OK, sorry. So I wanted to pause this just to say so this is going to be 6-decibel steps. So I told you that your threshold is around 1 dB. So these 6-decibel steps will be very, very obvious. OK, let's do it.

[AUDIO PLAYBACK]

- The decibel is reduced in 10 steps of six decibels. Demonstrations are repeated once.

[PULSING STATIC]

[END PLAYBACK]

JOSH MCDERMOTT: OK. Now, in this example, the next one, I think this is in 1-decibel steps. So these steps should just barely be discriminable. And a lot of times, you might not be 100% positive if the level is changing. But you'll be able to then hear it over a couple steps.

[AUDIO PLAYBACK]

- Broadband noise is reduced in 20 steps of one decibel.

[PULSING STATIC]

JOSH MCDERMOTT: OK. So that's very close to your discriminant--

[PULSING STATIC]

[LAUGHTER]

[END PLAYBACK]

JOSH MCDERMOTT: All right. So those steps were very close to your discrimination threshold. And you could probably just barely tell that that was changing. So that's the decibel scale. That's how we measure sound intensities. It's always based on ratios. There's got to be a reference. And we use it because it's pretty closely related to discrimination thresholds.

We're going to end there. Welcome to 935. We're happy to have you here. I'm going to have office hours upstairs. You're welcome to come by, no obligation. And we will see you Tuesday.