[SQUEAKING]

[RUSTLING]

[CLICKING]

JOSH MCDERMOTT: Happy Thursday, everyone. So today we're going to continue talking about lightness perception and then get into color. So remember, the problem that we're talking about is that organisms need to estimate surface pigmentation.

And the problem of lightness perception is the one-dimensional version of estimating pigmentation. So we're going to ignore effects of wavelength. So we'll talk about this quantity called reflectance that corresponds to the proportion of light that a surface reflects, and lightness is the perceptual correlate of reflectance.

And so last time we talked about how this really is-- this only really makes sense in the context of Lambertian reflectance, so these are surfaces that scatter light in all directions. And we talked about exceptions to that. There's specular reflections and then different degrees of diffuse reflections and so forth, translucency, but that many surfaces are approximately Lambertian, making it a pretty good starting point.

And so the problem of-- or the phenomenon of lightness constancy refers to the fact that people are pretty good at correctly estimating Lambertian reflectance. So a great piece of paper looks about the same shade of gray indoors as it does outdoors even though if you go outdoors, they'll be dramatically more light that's reflecting off of the surface. And then we ended the lecture last time with this demonstration of the kind of contrast between perceiving light variation as being due to illumination versus being due to reflectance. So you can take a shadow, you draw an outline around it, and it really changes how it looks.

So the formal problem that we're talking about here is shown here. So we've got a luminance image that gets measured by your eye. That luminance image is generated by a combination of illumination, comes from a light source, and reflectance. And so at every point in the image, there's some amount of incident illumination. And then there will be reflectance that's determined by the pigmentation of the surface in the world at that point in the image.

So the problem here is that we're trying to-- we would like to estimate R. The idea is that reflectance is an important property of objects, tells us about what they're made of. So we would like to estimate that.

What we observe is L. OK, L is the product of these two variables. And you don't know I either, so I and R unknown. So it's ill posed. You got one equation, two unknowns.

So the question is, how can we do this? And so people have been interested in this a long time. And many clues to how we solve this problem have come from, as you might expect, from illusions.

And so last time, we left off with this classic illusion of simultaneous contrast. So the two squares are the same physical shade of gray, but they look different. So the one on the left looks lighter than the one on the right.

All right, so historically, there were a variety of different ways of explaining this phenomenon. And people debated this pretty fiercely, back in the 1800s. So there was one fairly prominent perceptual scientist of the time, Hering, who favored explanations of this type of phenomena in terms of neural processing.

So remember, we talked about how there are these center-surround receptive fields throughout the early stages of the visual system, particularly the retina and the LGN. You can think of these center-surround receptive fields as making a comparison between the light that falls in their center and the light that falls into in their surround. And so imagine that you have a center-surround receptive field that's positioned on each of those two squares, it's going to give you a different response by virtue of the fact that the luminance in the surround would, on average, be different.

And so this particular receptive field, if it's centered on this, will give you a bigger response than if it's centered on this. And just as we explored these explanations of phenomena like Mach bands and the Hermann grid in terms of these kinds of receptive fields, they were also offered as explanations of simultaneous contrast. And so Hering was the person who favored these sorts of explanations.

All right, so Helmholtz was another very prominent perceptual scientist of the era. Helmholtz is normally credited with this idea of unconscious inference, the idea that, when you are observing an image or a sound, you're trying to infer the state of the world that is most likely given the observed sensory signal. And so the proposal here was that this illusion results from an attempt to discount illumination.

So the concept is that, when presented with a stimulus like this, your visual system is trying to come up with an explanation of the stimulus in terms of the things that the world is made of. And so in the context of this problem, the world is defined by illumination and reflectance. And so the notion is that you might observe an image like this and then hypothesize that what could cause that image would be a stimulus like this in the world, so a surface like this in the world where there are two patches that actually are different reflectances but that they are in different levels of illumination. So this one is in shadow, this one is in higher illumination.

And the notion is that the surrounds here could be interpreted as being due to the illumination. So the explanation of the stimulus would be that this is physically a light patch that's in dim illumination, and this is physically in a dark patch that's in bright illumination. And so the combination of the light patch in the dim illumination and the dark patch in the bright illumination produces the same physical gray level, so the same amount of light that comes off of the page, and that the explanation in the world is that you have the combination of causal factors and that your brain is trying to infer that and that what you see, when you look at this is your brain's estimate for what the reflectance plausibly is.

So that's the Helmholtzian explanation. And as I said, these things were debated. So Hering supported a mechanistic so-called bottom-up view of visual processing. Helmholtz supported a more inferential view based on experience, what we now refer to as priors and lightness perception. So the problem of reflectance estimation was a major battleground.

So this is a quote from a book by William James. William James is another pretty famous psychologist from that era. And he writes, Helmholtz maintains that the neural process and the corresponding sensation also remain unchanged but are differently interpreted.

Hering, that the neural process and the sensation are themselves changed, and that the interpretation is the direct conscious correlate of the altered retinal conditions. According to the one, the contrast-- so simultaneous contrast here-- is psychological in its origin. According to the other, it's purely physiological.

And it references this distinction between sensation, so the idea of the measurement process at the front end of the system, and perception. And so of course, we read this kind of stuff nowadays, and the debate seems a little bit silly because we could think of these two explanations as not being mutually exclusive. So the Helmholtzian explanation, of course, would have to be implemented neurally in some way. And potentially, the filtering that's happened in the early visual system could be part of that. But at the time, this notion of different levels of explanation wasn't really as clearly understood or acknowledged.

So this is just a little taste of the history of this kind of problem. Another major development in how people thought about lightness perception came in the '70s with Edwin Land. How many people have heard of Edwin Land?

Yeah, so he used to be pretty famous. So Edwin Land invented something called the Polaroid camera. How many people know what Polaroid camera is? OK, yeah.

So Polaroids have made a comeback. When I was a kid, everybody had a Polaroid camera. So Polaroid camera is this type of camera that instantly generates a picture. So you press the button, and it's got the film inside the camera, and it kind of develops like on its own. And so Edwin Land founded Polaroid and was a major presence in the Cambridge area a while back.

So he got interested in the problem of reflectance estimation because, in order to take pictures, cameras need to have the correct exposure. So they need to open the shutter for the right amount of time. So they need to divide out the effects of illumination.

And so he got interested in how humans solve analogous problems. And so he came up with what's called the Retinex Theory of lightness constancy, where retinex is short for retina plus cortex. And it's a pretty cool and elegant theory.

So the key idea of the retinex algorithm or theory is to take advantage of regularities in the world, i.e. scene statistics. And so the notion here is that some scene interpretations, the explanations of images, are more likely than others.

And so the proposal that Land and his collaborator McCann made was that it's common for illumination to vary gradually over space, like with distance from a light source. But by contrast, it's common for reflectance to vary abruptly, so to be piecewise constant. So the notion here is that a lot of objects, my shirt, are fairly constant reflectance. And so then you get to the edge of an object, and the reflectance kind of makes some large change.

All right, and so to explore the consequences of these generative assumptions, Land and McCann used this other very influential approach of employing a toy world. So they used images that they called Mondrians, so sets of random patches. So the proposal is that this is an approximation of the world.

So this is reflectance. So it's an image that consists of all these patches of different reflectance values, and then the illumination would vary gradually. And then to generate the image that would be observed, you multiply these two things. And this gives you the luminance image.

So this is the generative process. So we live in this simplified world. The illumination varies gradually. The reflectance changes discretely. And then they get multiplied, and you get the luminance image.

All right, so in this generative model, these things get multiplied. To simplify it further, you can take the log of both quantities to turn that multiplication into an addition. So we're going to work with the log of the illumination and so forth. We'll just, for simplicity, still call it illumination.

So this is what this looks like in 1D. So this is the image-formation process. That's the top half of the slide. So we've got illumination. So that varies as a function of space.

This is just a 1D depiction of this for simplicity. So illumination is on the y-axis here. Space is on the x-axis, position in image.

This is the reflectance. So the illumination varies gradually. This is just one example.

Here's a reflectance image where it varies abruptly. Those get added because we're now in the log domain. And so you get this luminance image.

So that's the image-formation process. That's this theory or an idealization of how the world generates images. So vision is the inverse process of starting with this luminance image and somehow estimating the illumination and the reflectance. Now, normally, you're mostly interested in the reflectance. But if you estimate the reflectance, then you probably could also estimate the illumination, and maybe you would need to in order to do that.

All right, so the key proposal on how you would solve this inverse process is to take advantage of this assumption about the world that illumination varies gradually and reflectance varies abruptly. And so given that assumption, if you look at spatial derivatives, so how the luminance changes as a function of space, it follows that big changes, abrupt changes should be due to the reflectance and small changes, small derivative values, should be due to illumination. So the idea is that you take your input, you compute spatial derivatives. And what you can see here-- so there's this constant illumination gradient that's been added.

So you have a small positive derivative here. And then you hit this big step, and so the derivative shoots up. Then you go back to the small positive value. Then it goes down.

So you get a larger derivative down and so on and so forth. So these are the derivatives here. 0 would be somewhere here. And so the idea is that you go through, and in any place where the derivative is large, so either strongly positive or strongly negative, you classify that as having been due to reflectance. Any place where it's small, you suppose that that was due to illumination.

So you then separate out the big derivatives and the small derivatives. And then you integrate the big ones to retrieve the reflectance. And then the remainder is the illuminance. So we take all of the derivatives and then integrate them. And you get this thing out that's an estimate of the reflectance, so super simple.

OK, so here's a flow chart providing another example of how this would work. So again, here is an example reflectance image and an example illumination image. Those get combined.

So here's the observed images. So again, the signals, in actuality, multiply. But we're going to work in the log domain, so they add. You get this observed image.

You take the derivatives, split them up into the strong derivatives and the weak derivatives. Integrate each of these as a function of space. And you get the estimated reflectance and the estimated illumination. All right, any questions about how this works or why? Yeah.

AUDIENCE: How was this pattern discovered?

JOSH MCDERMOTT: How was--

AUDIENCE: How was it discovered?

JOSH MCDERMOTT: They thought it up. I mean, yeah, it's just human intuition, really. Yeah.

Yeah, I mean, it's an interesting example where-- I mean, we'll talk about this a little bit later. I mean, part of why this was such an important development was it was just a nice example of how to approach this problem. So they simplified it down to something that's pretty simple but not so simple as to potentially be totally irrelevant. And then because they simplified it down, they could make these assumptions and build that into an algorithm that could actually solve the inverse problem.

But it was basically just a good idea. Yeah. Yeah.

AUDIENCE: Once they integrate, how did they know what constant to add to each of those bottom graphs?

JOSH MCDERMOTT: Yeah, it's a good question. And I think the major answer is the constant offset often doesn't matter that much. But, yeah, that's of an open question, I would say. Yeah.

AUDIENCE: Is that empirically true that natural images have this property?

JOSH MCDERMOTT: It's a great question. So hold that, and let's reevaluate that. That's a very good question. OK.

All right, so how could you implement this in neural hardware? So one idea that kind of resonated at the time was that you could use the kinds of filters that we find in primary visual cortex, which you can think of as taking an approximate spatial derivative. So what is a derivative? Well, it's a differencing operation.

It's telling you how much something changes from one point to the next. And if you think about a simple cell receptive field, you can think of that as taking a difference between two adjacent regions of an image. You got an excitatory lobe and an inhibitory lobe.

So you can think of that as taking a spatial derivative. So you get an input like this. And the cell will give a positive response when there's an increase-- a particular cell might give a positive response when there's an increase and then a negative response when there's a decrease.

So the kinds of ingredients in the algorithm, at least some of them in the way of derivatives, seem kind of compatible with what was known about early vision at that point in time. But another reason why people got pretty excited about retinex was that it seemed consistent with some illusions that were discovered-- or popularized at that time.

So this is a very well-known illusion called the Craik-O'Brien-Cornsweet effect. Craik, O'Brien, and Cornsweet were humans that, I guess, discovered this, I think maybe independently, or I can't remember if they were working together. And the stimulus is what's shown here, so this is showing the luminance as a function of position.

So it's just these two gradients, where the luminance kind of starts out high and gets low, starts out high and gets low. And the illusion is that, when you look at this, it looks like the thing on the left is darker than the thing on the right. But in fact, they're physically identical.

All right, so let's think about this. How is this consistent with the retinex algorithm? Well, what retinex is going to do when it looks at this is take all these derivatives. And so you get small negative derivatives here and then this big positive derivative and then a small negative derivative. And so if you just assume that that one big positive derivative is due to the reflectance and then you integrate that, it's going to tell you that this is lighter than this.

All right, so this kind of illusion and others like it indicates that the visual system gives very strong weight to edges in the computation of lightness. And not so much to slow gradients. And so this seemed like suggestive evidence that the visual system might be doing something similar to retinex.

And there are a lot of other illusions that are kind of the same flavor. This is another one that's kind of nice. This is the Cornsweet square wave grating. So again, the physical stimulus is what's depicted here.

So this is, again, luminance as a function of position. And so the point of this stimulus is that the middle regions of each of the squares are the same gray level. But then there's this scalloping here, creating these big step edges.

And so the point is that, when you look at this, it looks like this is lighter than that when, in fact, the middle portions of them are the same. And so this is showing how the stimulus is created. So you start with a square wave. You high-pass filter it, you get this is the physical stimulus. And then you've got these big derivatives here that are happening kind of at the edges and then an integration process that would cause things to fill in so again, consistent with what you might expect from a retinex type algorithm.

All right, so one of the major themes of this class is that illusions reveal the workings of our perceptual systems. So we call illusions illusions because they're, from a certain perspective, an example of the visual system getting things wrong. So we have these two stimuli that are actually physically identical, but they look different. But we actually think that these illusions result from the action of successful engineering designs that we place in unusual circumstances. So by studying illusions, we can study the hidden design rules underlying those engineering designs.

All right. Another kind of theme of the class is this idea of levels of analysis. So as famously articulated by Marr, so we've got the computational level talking about what the problem is that's being solved, the algorithmic level of what the approach is to solving it, and the implementation level, how to actually put it into hardware, and reflectance estimation is a nice example of this.

And so this is actually an excerpt of the book that Marr wrote. So this was his magnum opus. So Marr was a researcher here back in the '70s, and he got cancer at a very young age and wrote this book on his deathbed. And then the book came out, I think, right around when he died.

And it just has his worldview for how to study vision and maybe perception more generally. And the retinex algorithm is featured fairly prominently in there because it's a nice example of these principles. And in particular, he gives the Retinex Theory as an example where the same computation can be implemented in two different algorithms.

And so this is an excerpt from the book that illustrates this. And so I'll just read you what he says. He says, the retinex computation-- again, that's the computational-level description of this-- so the problem being solved here is the estimation of reflectance, leveraging assumptions about the nature of illumination and reflectance. That's what allows the problem to be solved. So that's a computational-level description.

So that computation has been implemented in at least two ways. So Land and McCann themselves used the one-dimensional approach illustrated in figure 383, which I've shown here for your convenience. If we trace the image intensities along any path from A to B, they will have the form shown as blah, blah, blah.

OK, so he talks about this approach, but then he says Horn, another vision researcher from back in the day, derived a two-dimensional analog of this algorithm illustrated in figure 3-83(b), which is shown here. Consisting of essentially the same three steps, but now in two-dimensional form. So these are two different algorithms for solving the same computation.

So a nice example of levels of explanation. So the ideas that we illustrate here with retinex are, number one, levels of analysis, so the same computation can be implemented with different algorithms. Number two, scene statistics help with ill-posed problems.

So this is what we think of as priors, the idea that certain things are a lot more likely in the world than others, in this case illumination varying gradually, reflectance varying abruptly. They're not explicitly defined as such, but same idea. Another thing that's illustrated with retinex that was very influential is this idea of using a toy world to help think through a problem. So the use of these Mondrians and illumination gradients.

And it was also very influential because it was this really nice convergence of ideas from computation, from psychophysics with these new illusions, and physiology, where some of the operations seem consistent with the physiological things that were happening, that were being observed in the visual system. Any questions about retinex or any of this?

All right, so another thing that's really important about retinex algorithm is what's called stimulus computability. So sometimes, the theories that you encounter when you're dealing with things related to the mind are verbally expressed. It's like there's an idea, gets described in words, and it's not always clear how to say, take an image, and actually figure out what the theory predicts.

The retinex algorithm has this property of stimulus computability. And what that means is that it is specified with enough precision that you can apply it to any arbitrary signal, and it will do something. And that defines the prediction of the theory. And that's really important because it enables you to actually evaluate the theory fairly exhaustively because you can test it on anything essentially.

And so that's what we aspire to always have. It's often kind of hard to do. But this was an early example of this.

So stimulus computable means that a model or algorithm or theory can be applied to any stimulus. Not all models or algorithms are defined in this way. Some are defined with words.

Others work on abstract variables. So we assume that there are these variables that are extracted by earlier processes. And so the theory is trying to model a second stage, but it means you can't really test it on actual stimuli.

All right, and so the consequence of this is that we can actually apply the retinex algorithm to real images and see what it does. And it does funny things. And this gets at the question that was asked earlier of the extent to which these assumptions that are embodied in the algorithm are really accurate. So here, we have an image that's just taken of a toy in the world. And you can apply the retinex algorithm, and it outputs an illumination estimate and a reflectance estimate.

So the reflectance estimate looks pretty reasonable. So the tires are dark, and the hair is dark. And the bucket is pretty dark, and the shirt is lighter. The background is lighter right. It looks OK.

The illumination estimate, on the other hand, looks pretty funny. And so some of what's happening here is some of the luminance variation, like, for instance, on the tires is not due to illumination, but rather, it's due to shape, so the surface orientation changing and, as a consequence, there being less reflected light. Now that happens gradually because it's just a curved surface, and the shape changes gradually, and so that produces gradual changes in luminance.

And because the world model that is implicit in retinex really only knows about illumination and reflectance. It doesn't know anything about three-dimensional shape. It assumes that every luminance change that happens in the image is either due to illumination or reflectance. And that's not true of the world. There's luminance variation due to lots of other causal factors, three-dimensional shape being a big one of them.

So on the other hand, it is getting this right. This looks like maybe there was sunlight through a window or something. But part of the importance of these things being stimulus computable is that you can test them on real images, and they break. And then you get insight into what's missing from the theory.

OK, so what's wrong with retinex-style theories of lightness constancy? So one of the big things that's wrong-- and so I just alluded to something else that's wrong about it, which is the world model is incomplete, that luminance changes can come from other things. But even if you accept the idea that you just have reflectance and illumination, this other big problem is that illumination changes are not always gradual. Sometimes they're abrupt, and so shadows in particular often have fairly crisp edges.

And evidence that retinex is incomplete as an explanation of lightness perception comes from the fact that humans appear to discount illumination despite the absence of gradual change in some cases. And so there were these early illusions from the '70s that seemed consistent with retinex. And then there was another wave of illusions in the '90s and 2000s that kind of revealed the limitations.

So this is one example. And a lot of this work was done by Ted Adelson, who is a professor in this department in CSAIL. He was my PhD advisor. And this was what he was working on back at that period in time.

So this is one example called the tips and strips illusion. So we have tips and strips, and it's an illusion because all of these gray regions are physically the same shade of gray. But when you look at this, I mean, these look dark, and these look light.

And one explanation for what's going on is that there is evidence that this region is in high illumination and this is in shadow. But you can see that the border between these two potentially different regions of illumination is quite abrupt. So if you gave this to retinex, what would the retinex algorithm do? Yeah.

AUDIENCE: It would assume that it's all reflectance, not due to different illumination.

JOSH MCDERMOTT: Exactly, yeah. So it would say, this is all reflectance. Now in actuality, this is all reflectance. This is just an image that was created in Photoshop or something, but, yeah. So it would not do what the human visual system is doing, which is essentially inferring that the reflectance is lower here than it is here.

This is another one with the same idea, the snake illusion. So the diamonds there are always the same gray level. And so you can have this variant here where you essentially just get simultaneous contrast. So there's a little bit of difference between the lightness of this and this.

But then this version here where things are just tweaked a little bit, and now there's a pretty big difference between the two sets of diamonds. And so again, the idea is that there's some evidence that this is in high illumination and that's in low illumination. So the same physical shade of gray would be explained as there being low reflectance and high illumination here and high reflectance and low illumination there, so the Helmholtzian explanation we talked about earlier.

So this also made it onto a T-shirt. So this is a T-shirt that I happen to own that was printed up 25 years ago or something. And this is a particularly cool version of this because these phenomena that we're talking about, they cause this to look darker and not to look lighter.

But in actuality, there's just some dark stuff and some light stuff that's printed on the same color cloth. So this and this are just physically the same stuff. So you can tell when you look at it it's the same cloth, but it ends up looking different.

All right, so how can we explain this kind of stuff? So I'll start by just saying we don't actually have complete stimulus-computable explanations that can work on all of these classes of illusions. That is something that would be good to revisit. But all of the pieces of evidence that we have suggest a potential alternative approach.

And one such approach is to not try to estimate illumination explicitly but rather try to find boundaries between regions that are likely to be in different illumination and then to look at the range of luminance values that you have within these regions that are plausibly different regions of illumination and then try to map that onto expected ranges of reflectance. So I'll just give you a little flavor of how this could work. So again, let's think about the world consisting of a set of random gray patches that are viewed in different illuminations.

And so if we have some patches that can range from black to white in terms of reflectance, in high illumination, that would create a big range of luminance values from low to high. In lower illumination, that would create a smaller range of luminance values because you get smaller numbers that are multiplying the same reflectances. There are also some viewing conditions where there's an additive component, if you're viewing something through fog, and then that causes everything to get shifted up. That's another interesting wrinkle.

And this is what these would actually look like. So we've got this set of random gray patches here. They're stripes.

And in the high-illumination condition, the reflectance here gets mapped onto this particular range of luminance. In the lower-illumination condition, it gets mapped onto a smaller and lower range of luminance values. And then here's the hazy condition where there's this-- it gets shifted up.

And so the notion here is that what could be happening is that if you have some way of finding boundaries between regions that are in different illumination, you could just sample the luminance values that you're getting in a particular region. And if you assume that there's some-- that those samples come from a particular range of reflectance from light to dark, you would then take the highest luminance that you would see and map that onto something that's pretty light and the lowest luminance that you see and map that onto something that's pretty dark.

And so the idea is that you're going to have these different ranges for different illumination conditions, but you would take this particular range and call this something that's pretty light and take this range and call this something that's pretty light. And again, this is all a little bit hand-wavy because we don't really have very precise implementations of this. But that's something that could potentially be developed.

So the idea is that you could potentially sample the ranges of luminances that you have in an image. Ideally, you'd like to get enough samples that would cover the full range of reflectance. If you make the window in which you're sampling things larger, you're more likely to accurately sample the full range of reflectance, but on the other hand, you could mix viewing conditions.

So the idea is that this whole thing only makes sense if the things that you're sampling are under the same illumination. And so here's a situation where this is pointing up in higher illumination. This is pointing down in lower illumination. And you really don't want to be comparing things that are in these two different states.

And so one idea is to try-- that the visual system might be trying to collect enough samples so that you could get some sense of the range of reflectance while avoiding mixing illumination conditions. And thus, it would be taking advantage of strong cues to illumination boundaries and such that your lightness percept would be really modulated by these strong cues. So we'll come back to how that might work in a second.

And so I showed you some of these examples where cues to illumination boundaries, so these junctions here that suggest that you might have, darker illumination here and high illumination here, usually produce lightness differences. You get these big illusions in cases where there's very strong evidence of illumination boundaries. And then these other cases where you have similar gray levels scattered throughout the image, but things are not set up so as to cause this sense of illumination boundaries. You don't really get these differences.

So if we take this back to what we started with, so this is the classic simultaneous contrast effect, which we started out with. So you got these two squares. They look a little bit different.

The one on the left is darker than the one on the right. So there's an illusion there, but it's a small effect. So here we've taken those same grade levels here but then set them up in such a way that you have all these additional cues that this and this would be in different regions of illumination.

So it looks like they're-- the surface orientation is different. There's these other regions here that are consistent with the idea that this is in high illumination. And so one way to think about the fact that the illusion here is a lot stronger than the illusion here is that there's stronger evidence for an illumination edge. And so the visual system is less likely to mix up the viewing conditions whereas, in classic simultaneous contrast, that could be an illumination boundary, but you don't have very strong evidence for that. So it could be partially influenced by the stuff on the other side.

So that's a hand-wavy indication of really how some of this might work. Now, you look at something like this, and it's natural to maybe suppose that the lightness computation is actually maybe influenced by a three-dimensional interpretation of the scene. So you look at this thing, and it looks like it's folded like this, with this being oriented differently than this.

So is that actually what's going on? And so there have been a few examples that kind of actually suggest that we're not really using a full 3D model of surfaces but instead using something that's a little bit simpler, maybe more appropriately described as mid-level. So here's one example.

So we've got these different gray patches here, shown here and here. And the things that the arrows are pointing to are always physically the same shade of gray. So up here, they look like they're roughly the same shade of gray. Down here, they don't. That looks darker than that, so that's the illusion.

So in this case, this makes sense. Why does it make sense? Well, the three-dimensional shape of this surface is such that it looks like this is poking out. So this surface would be angled up, potentially towards the sun.

This one is angled down, thus, in shadow. You get a lot of cues here. This is high illumination.

Got a lot of cues here. This is in low illumination. All the luminance values are low. So it makes sense that there would be a difference here.

All right, so this was discovered and published. And then somebody came along and came up with this variant where you can see that this has got the same shades of gray as this, but this one is reconfigured. And now the three-dimensional interpretation is different. And it looks like the orientation of this surface and this surface are actually the same-- they would both be pointing up.

So logically, that should suggest that they would be in the same illumination, pointing in the same direction, about the same place. But the lightness difference persists, so that looks darker than that by about the same amount that looks darker than that.

So that's arguing against this idea that you're using this model of 3D shape to determine whether or not the illumination is likely to be the same or not. Instead, it suggests that there's something simpler going on where you're using some local cues and again, in an unconscious way, supposing that this stuff all has high illumination and this has low. Yeah.

AUDIENCE: Doesn't that feel like a super convincing case where it's like everybody's going to view this as having the same illumination? You can also see it as something flat and then going up and then going down, to the right? So it feels like both are maybe possible.

JOSH MCDERMOTT: Yeah, that's a good point. Yeah, that's, potentially, an alternative explanation.

AUDIENCE: Is this super convincing in terms of discounting the 3D case?

JOSH MCDERMOTT: All right, yeah, so we have a skeptic in the audience who's arguing that this maybe is not the most compelling evidence that 3D shape is not being used. I mean, I think one question that you could ask yourself is, if you look at this and the 3D shape is bistable and you can see it in a few different ways, does the perceived lightness actually change when your three-dimensional interpretation changes? So for me, I see both of those as is pointing up. I get a pretty good lightness difference, nonetheless.

But we will see-- I mean, in the next lecture on color, I will show you some cases where three-dimensional structure actually does seem to actually influence color in some cases. So I would say none of this stuff is like perfectly settled. It's more just like, this is some evidence that this could be based on some local heuristics rather than something related to the global interpretation.

So the thought is that a lot of the action may be happening at a mid-level representation using sensible heuristics. Here's another piece of evidence that maybe is consistent with that. So here we've got a pretty strong illusion where this looks darker than this same physical shade of gray. Again, same idea, a lot of evidence for an illumination boundary here.

This is consistent with high illumination, this, with low illumination. And so under the story we were telling, you got evidence for illumination boundaries, and so within this region, this is the lowest luminance value. And thus, you would map that onto something dark whereas this is the highest luminance value, and that would look light.

Now, in this region, this is now the highest luminance value. And so that would get mapped onto something that is light. And in this region, these are darker and so forth. So again, you don't need to explicitly estimate the high and the low illumination necessarily.

All right, so that all makes sense. Now, in the one on the right, everything is the same. It's just that we've moved these ovals further away. So these kind of local cues here for the illumination boundary are pushed further away from the diamonds.

And the argument is that the illusion gets weaker. So the argument is that the difference between these two diamonds is not as pronounced as it is here. And again, logically, the evidence that there is an illumination boundary here is just as high in this image as it is in that one. You got all the same stuff. It just got moved further away.

So it suggests that there's some kind of limited spatial propagation that goes on and argues against a high-level cognitive interpretation of illumination boundaries. Does that make sense?

AUDIENCE: Why does it argue against high-level cognitive interpretation?

JOSH MCDERMOTT: Because if I were reasoning about this-- the evidence that there is an illumination boundary here is-- there's the same evidence here as here. You got all the same local features. They just got moved further away from the diamonds. So there's some effect of the distance that I think you wouldn't necessarily expect in that case.

AUDIENCE: Because distance has to do with mid-level?

JOSH MCDERMOTT: Well, no, it's just that the-- you got this straight line here. So the question is, what causes that straight line? And you've got these bits of local evidence that that's due to an illumination boundary because you've got these x junctions here. And all that same evidence is here and here. And so, yeah, I mean, certainly, if we were reasoning about it logically, there's as much evidence in one case as the other.

AUDIENCE: So you're saying that mid-level has to do with local things, like physiological effects?

JOSH MCDERMOTT: Very vaguely. Yeah, it's very underspecified. Yeah.

OK, so if you take a lot of different cues to illumination, different regions of illumination, and you combine them, you can get very powerful illusions. So here, we've got these ovals on the top and these ovals on the bottom. Those are the same physical shade of gray, but they look like they're really, really different.

And there's lots of cues that are all pointing in the same direction. So there's local contrast. So this is by far the darkest thing up here. This is by far the lightest thing down here.

You've got this thing called a penumbra. This is the edge of a shadow, which tends to be blurry, which indicates that this is in shadow. There are these three-dimensional cues to depth, which may or may not matter, that would suggest that this is on the underside of a cylinder and this is on the top.

There's a little bit of a specular highlight that indicates that there's a light source that's coming from above. So there's lots of cues here, and they're all pointing in the same direction. And that gives you a pretty big effect. And this is another one with the same idea. So the arrows here are all the same, but they look very, very different.

This one is also pretty cool. So the two blocks are actually the same color. So if you put your fingers covering this up, you'll be able to see that.

So just to summarize. Many surfaces have approximately Lambertian reflectance, though, as we talked about, not all. So that means that reflectance can be summarized as one number.

Lightness is the perceptual correlate of reflectance. Lightness illusions are probably caused by processes at multiple levels-- low level, mid-level, and high level. Although, exactly what that means and how those things operate and getting complete theories of this is still an ongoing thing.

We talked about retinex as one example of an influential model, illustrates some important principles that probably have some truth. The idea that abrupt edges have a big influence on perceived lightness and reflectance estimates, but it's pretty clearly not fully, fully correct. In particular, the main shortcoming is that it assumes that illumination varies smoothly, and that is not always the case. And you're able to actually make reflectance estimates that seem to incorporate the effects of abrupt illumination boundaries.

So we currently lack a stimulus-computable model of lightness perception that can explain most lightness illusions. We also don't have clear neural correlates of lightness. So that's an open question.

So we think that a lot of the action happens at a mid-level, where 2D heuristics about grouping contours, surfaces, and junctions are being used. But as I said, I haven't given you a complete theory of this. It's just some hints as to what some of the ingredients and what that kind of explanation might look like. What questions do you have about lightness? Yeah.

AUDIENCE: So some of those later illusions, like the snake cylinder and the one after it, there's a lot of squiggly lines that are being used. So is there a reason for those? Do they add to the illusion in some way? Or would it still-- I'm curious.

JOSH MCDERMOTT: Like these squiggly lines?

AUDIENCE: Yeah, on the cylinder too, is it to add more surrounding stuff to it [INAUDIBLE]?

JOSH MCDERMOTT: OK, so here, they definitely help give you this percept of 3D shape because you can see that the squiggles are closer together here than here, which is an indication that the surface is oblique. So they'll help with that. Again, how much that actually benefits the illusion I'm not sure in this particular case, but that's one reason. It makes the three-dimensionality of the thing a little bit easier to convey in a synthetic image like this.

And then in these kinds of things, I think the curves, they probably enhance the sense-- they probably strengthen the evidence for the illumination boundary. If you made these things rectangular, I think it'd probably be easier to explain this as two different things that happen to be placed down. That's just my intuition, but that's just a hand-wavy intuition. So, yeah, my guess is that it helps a little bit, but, yeah, I don't know. Any other questions?

We're now going to talk about color. So I'll talk about this for the rest of the lecture and then probably next time. So the world is full of color.

So this is different types of fruit. They're different colors. Color is really useful for lots of things. You can get insight into this by looking at images that are not colored and then switching to colored versions.

So it's a little unclear what is in that plant. And then you add color, and you can see the fruit in the bush. Here's two different versions of the same image, black and white version and a color version. And it's much, much easier to see the flowers in the color version than in the black-and-white version. So color serves to help us segregate objects.

Also helps us with object recognition. So if you had to identify what type of fruit our former president is eating here, you would have a hard time. But when we add color to the image, it becomes a lot easier to make out.

Color also gives us information about object state. So those different pieces of fruit all look pretty similar, but some are a lot riper than others. Similarly, depending on your taste, you'd probably want to have some of these pieces of steak and not others. So it tells us about object state.

Color also tells us a lot about the state of the world. So here, we have two different natural scenes. They don't look all that different in black and white, but with color, you can tell that one is quite lush. The other one is pretty arid.

All right, so color is really important. And it's also an important part of art. Some people have a favorite color.

It's really important to how you pick your clothes. Some combinations of colors look good and bad. It's a big part of life.

So color is a consequence of light. It's part of vision. And if you observe different wavelengths of light in isolation, they appear to be particular colors. So this is the spectrum of visible light, from roughly 400 to 700 nanometers.

The short wavelengths, the shortest wavelengths that are visible when you view them in isolation look pretty blue. The longest wavelengths that are visible when you view them in isolation tend to look red. In between, you get this spectrum. So this is why, when you take sunlight and you pass it through a prism, you see a rainbow.

All right, so different wavelengths of light in isolation appear to be particular colors. But we really believe that the function of color biologically is to provide information about pigmentation. And so these different wavelengths of light just really provide a signal for the different pigments that things can be made of.

So this is the elementary school explanation of color. I call it the elementary school explanation color because this is what I learned when I was in elementary school. Maybe they have an even more sophisticated version that is popular now.

But the elementary school explanation that I got is that objects appear colored if they reflect some wavelengths of light better than others. So the idea is that the light from the sun is approximately white, so it contains all wavelengths. So you get all of these different wavelengths of light that are kind of impinging on a surface. And if the object is red, it will tend to reflect the long wavelengths, the red light more than the shorter wavelengths. If the object is green, it will tend to reflect the medium wavelength, the green wavelengths more than the short and the long, and so on and so forth.

How many people learned this explanation in elementary school? Yeah, so I see it's still fairly popular. And there's a lot of truth to this. So you can actually measure the spectral reflectance of different surfaces in the world. So that's what's shown here.

So what this is a graph of is it's showing you for six different kinds of surfaces of the relative energy that they reflect as a function of wavelength. So for the things that don't look colored, so the white and the gray and the black surface, they differ. So that's what we'd be studying in the world of lightness perception, so no color there.

But they differ in how much light they reflect. So the black surface doesn't reflect much light. The gray surface reflects intermediate levels of light, again, across all wavelengths. And the white surface reflects a lot of light across all wavelengths.

But the colored objects, the leaf and the orange in the tomato, they reflect some wavelengths a lot more than others, so some wavelengths get absorbed and others get reflected. So the tomato, which is red, is predominantly reflecting long wavelengths of light, absorbing everything below about 600 nanometers. The orange is again predominantly reflecting long waves of lengths of light but a little bit bigger range than the red thing. And the leaf, which is green, is predominantly reflecting the middle wavelength of light.

All right, so this all seems pretty consistent with the elementary school theory of color. The additional wrinkle, which we will address a little bit later, is that, as in the case of lightness perception, reflected light is not just a function of the reflectance properties of an object. It's also a function of the spectrum of the illumination. So it's the same kind of problem that we talked about in lightness perception. Only now, we've got lots of different wavelengths to worry about.

All right, but it all starts with the fact that there are these different wavelengths of light. And our ability to detect differences in wavelength relies on the fact that we have three different types of cones. So remember the retina, the photoreceptors are either rods or cones. The cones come in these three variants, and those are differentiated by their spectral sensitivities.

So this is the spectral sensitivity of the three types of cones. The scientific term for them is S cones, M cones, and L cones. That stands for Short wavelengths, Medium wavelengths, and Long wavelengths. They're often colloquially called the blue, green, and red cones.

So this graph is plotting normalized sensitivity as a function of wavelength. So you can think of these curves as reflecting the probability that one of these types of receptor will absorb a photon at a particular wavelength. So it says that, for the S cone, it's really good at absorbing 450 nanometers and not so good and, really, not absorbing at all 600 and 700 nanometers.

So if you imagine that you've got a photoreceptor that's in the business of absorbing photons and generating a response, if you have a whole bunch of those receptors, so if you sum over many receptors or over time, this is going to be the magnitude of response that you would get for a given intensity of a given wavelength. So this graph is, again, showing the cone spectral sensitivities, but we've also added in the rods just for comparison. That's the dotted line. And there's only one of those. You've got three cones and one rod.

So one other way to think about these curves is that if you had only one receptor type, then the receptor would determine psychophysical sensitivity. And there's a scenario in which you have only one receptor type. And that is when the light level is very low.

So this is scotopic conditions, remember? So under scotopic conditions, the cones are not doing anything. You've just got this one type of rod.

And so this is a graph that is comparing the rhodopsin-- so that's the molecule in the rods that absorbs photons, the absorption probabilities of that-- to human scotopic sensitivity. So scotopic sensitivity, remember, sensitivity is the inverse of a threshold. So it's what the light level needs to be for you to detect it.

So you could, in principle, measure detection thresholds. And the point is that these two curves are essentially the same. So the ability of the rods to absorb-- the photons of different wavelengths determines how sensitive you are to them.

Another way that you could think about measuring scotopic sensitivity is to do an experiment where an observer adjusts lights of different wavelengths until they appear to be equal in brightness. And they will appear equal in brightness because they are producing the same response in the rods. So I can give you 450-nanometer light and a 600-nanometer light, and I could ask you to adjust their intensities, and you would be able to do that, and you will be able to make them look identical.

Now there'll be different intensities and at different wavelengths, but they'll look the same. And this is possible because of this thing that we talked about a little while ago called the principle of univariance. So once photons are absorbed, they have exactly the same effect on the photoreceptor.

So the only thing that really determines the response of the photoreceptor is whether the photons get absorbed or not. Some types of photons are more likely to get absorbed, and that means they give you a bigger response. So all the receptors can do is respond more or less.

So they vary along one dimension. So that's why it's called the principle of univariance. So it just goes up or down.

All right, so we can think of the initial stage of color vision that's happening in the retina with the cones as a mapping from the space of spectra to the space of cone responses. So this is an example to help us think about this. So what is depicted here is the spectrum of a particular light. So the y-axis here is power, and this is wavelength in nanometers from 400 to 700. That's the region that you can see.

So this particular spectrum is the sum of 31 monochromatic lights. A monochromatic light is a light source that emits a particular wavelength. And they're spaced every 10 nanometers.

All right, so to fully specify-- and you can see that the intensities of those 31 lights vary from one to the other. So to fully specify that mixture, you would need 31 numbers. So you'd need a vector of length 31. So we could also say there's 31 degrees of freedom, 31 dimensions. All means the same thing.

Now, most colors in nature have continuous spectra. So to specify a continuous spectrum, you would need an infinite set of numbers. So this is an infinite dimensional quantity. Now, the output of a single cone, when it is stimulated by a given light with a given spectrum, is a single number.

Thus, a cone maps an infinite dimensional vector into a single number. Now, you have three types of cones. So those three cones together map that infinite dimensional input, so that's the light at a given point in space to a three-dimensional vector. So another way to say it is that each spectral distribution is a point in an infinite dimensional space. The cones project this point to a point in a three-dimensional space.

OK, so here's a picture that depicts what happens when you create a spectral image. So you got a person. They are looking at something, a surface. There's a light source, the illumination.

So photons from the light source hit the surface. Some of them are absorbed. Some of them are reflected back. They enter the eye and are absorbed by the cones producing cone responses.

And so the bottom row shows the different quantities that are involved in this process. So here's the spectrum of this illumination. So this is a light bulb.

Light bulbs tend to have more power at long wavelengths. So you get something that looks like that. So that's the spectrum of the illumination in this particular example.

Here's the reflectance. So this is a blue surface. And so it reflects more light at short wavelengths than at long wavelengths.

Now, the scattered light that comes off the page is the product of the illumination and the reflectance. So you just pointwise multiply this and this. And so that gives you this.

And so you can immediately see that there's a tricky problem that the visual system faces. So the blueness of the surface shows up in the fact that there's a bump down here. But the fact that the illumination has so much more power at the long wavelengths means that the scattered light is very biased towards long wavelengths, or red light.

So this is the spectrum that enters the eye at a given point. These are the cone sensitivity functions, the things that we just looked at. This gets multiplied by each of these to yield the responses of the L, M, and S cones. And it would be a lot better if these were flipped. I don't know why they are not, but short, medium, and long wavelength cones.

So even though this is a blue surface, the response in the S cones is actually lower than the response in the L cones because the illumination is so heavily biased towards long wavelengths. OK, so that's process here. We've got this infinite dimensional thing here that's coming into the eye, the spectrum of the scattered light, that's caused by these two things in the world, the illumination and the reflectance.

That produces three responses in each of the three cones. And your percept of color has to be derived from these three numbers across the entire image. What questions you got about that?

AUDIENCE: So why is it blue if the red is higher? Why does it look blue?

JOSH MCDERMOTT: Why does it look blue? So this is the problem of color constancy. And so you would have to have some way of discounting the illumination, either estimating the illumination and dividing it out or doing something else. And so we're going to talk a lot about that in just a little bit.

OK, so the key idea here is you're critically dependent on the fact that you've got these three cones. So if you just have a single receptor, you can't tell anything about the wavelength composition of light. So here's an example where we just look at the responses of the M cones to these two individual wavelengths.

And so in this particular example, these two wavelengths have the same-- so the M cone has the same sensitivity to, it looks like about, 450 and about 625 nanometers. So you get the same response in the M cone. So even though these are two physically distinct wavelengths, you get the same response, and you wouldn't be able to tell them apart.

But when you add in the other two cones, you can tell the difference because the other two cones give you different responses. So even though the M cones give you the same response, the L cones respond a lot to the longer wavelength and not at all to the shorter wavelength. And the S cones respond a lot to the shorter wavelength.

And-- sorry, I said it backwards. So to the longer wavelength, the L cones respond a lot. The S cones don't respond at all. To the shorter wavelengths, the S cones respond a lot, and the L cones respond less. So the three-dimensional response is different.

All right, so even though you are mapping this infinite dimensional space down onto three dimensions, those three dimensions allow you to distinguish different colors to some extent. So it's often very convenient and clarifying to think of the color transduction process as linear algebra. So we've got the light that's projected into the eye. Think of that as a vector, a very long vector that tells you the power of the light at each different wavelength.

And then you have three spectral sensitivity curves. Those curves that were plotted on the graph. So that tells you, for every wavelength, what is the sensitivity of the L cone, the M cone, and the S cone? That vector gets multiplied by that matrix. And you get out a three-dimensional vector, which gives you the response of the L cone, the M cone, and the S cone.

All right, so one really important consequence and one really intriguing consequence of this, namely the fact that you have this very high-dimensional thing coming into your eye that gets projected down onto this three-dimensional space, one consequence is that there are many physically distinct spectra that will produce the same cone responses and thus will be indistinguishable to a human even though physically you could measure them and see that they're different. And these are called metamers. So metamers are physically distinct stimuli that are indistinguishable to an observer. And so in the case of color, anytime you excite the cones in the same way you experience metamers.

So these are two examples. These are two spectra. So power versus wavelength, they're physically distinct.

This is the approximate spectrum from a tungsten light bulb, the kind of light bulb we used to all have in our houses. This is a metameric match from a color monitor. So this is a physically distinct spectrum that produces the same cone responses and would look exactly the same to a human viewing it.

So metamers are different mixtures of light that excite the cones in the same way. And this is a consequence of, again, this property of univariance, the fact that the cones only respond on this one dimension and then the fact that you have only three of them. So these are three examples of stimuli that produce the same responses in the cone.

And so for simplicity here, these are stimuli that don't stimulate the cone, but they are cleverly arranged to produce the same response in the M and the L cones. And there are lots of these. OK, so metamers produce exactly the same neural response starting at the cones. So here's an imaginary example.

Lights A and B. So light A is projected on this half. Light B is projected on this half of the circle. They have different spectra.

But if they stimulate the three cones in the same way, they will be indistinguishable. So they'll look like the same color. And no later processing in the brain can tell them apart. So the thing that distinguishes them was lost at the moment of photon absorption because the only thing that your brain gets is the cone responses. So to tell these things apart, you'd have to get a photometer and measure it with something else other than your eye.

So here are some puzzles to think about. So suppose you adapt to a uniform red field. So I show you a big sheet of red, and I make you look at it for a minute, like we sometimes do in this class. That will reduce the gain, the sensitivity of the L cones, and the consequence of that is that if you then go around and look at some other stuff, everything will look a bit more greenish than it normally does.

So if you do that and you then look at lights A and B, both of them will look more greenish than they did before, but they'll look the same. So you're reducing the sensitivity of one of the types of cones. So it's response gets a little bit lower. But because of the linearity of this process, light A and light B will still generate the same cone responses even though they'll be different than they were before.

Here's another puzzle. So imagine we have lights A and B and then we introduce a wash light. So this is a separate light that gets added and is shown on the entire circle.

So you can think of that as adding some additional light to both A and B. So A and B will now, again, both look different than before, but they will still look the same as each other. And so I want you all to think about why this is and again, to try to explain this in these linear algebraic terms. So this is a good case where thinking about this as matrix multiplication ends up greatly simplifying these things that would otherwise make your brain hurt. So something like this will be on an upcoming problem set.

All right, so metamers are kind of cool. They're also practically really important because they are the reason why monitors, color TVs, cameras, film, et cetera work because what this means is that, with just three different color guns, so with three different light sources, you can create any color experience because the space of color is defined by this three-dimensional space. So if you have these three things that you can adjust, you can move the cone responses around to match any particular color experience. So super important practically.

So now we know about the different types of cones nowadays-- because, in part, you can chop up the retina and measure stuff and find the pigments and measure their responses and things. But that's a relatively recent development in human history. But the idea that color vision is trichromatic predates all of that kind of modern biology and was hypothesized, really, back in the 1800s by Young, a vision scientist, and Helmholtz, who we hear about a lot in this class.

So they proposed that different-color experiences are due to the activation of just three receptor types, again, long before anybody could actually measure the actual cone properties. And this was based on the observation that you could generate all possible colors just by mixing lights of three different wavelengths. So this just turns out to be a fact about color vision. So that led them to hypothesize that there are these receptor types. It turned out to be correct.

All right, but this was controversial at the time. And the reason for this is that there are some, what seem like, problems with this idea that color is purely three dimensional. So one problem is that sometimes people are colorblind. Colorblindness is fairly common, so you probably know somebody who's colorblind.

When people are colorblind, they are typically missing one or more receptors in the eye. But when this happens, colorblindness typically occurs in pairs. So you're not just blind to red or just blind to green. Typically, you're red-green colorblind or blue-yellow colorblind.

And the pairs are not arbitrary. So it's not just any pairs. So you don't ever get blue-red colorblindness.

So again, people were aware that colorblindness existed. And that led people to question the fact that-- or the proposal that vision was purely trichromatic. Trichromacy also really doesn't give a very clear explanation for the fact that it's impossible to create certain kinds of color mixtures. So there's really no such thing as a bluish yellow or a greenish red.

And so this gave rise to a competing theory. Again, this is from Hering, who we talked about in the context of lightness perception. So Hering made this observation that a red can look yellowish or bluish, but not greenish. A yellow can look greenish or reddish, but not bluish. And inferred that, in the visual system, there were opponent color axes, so something that was computing the difference between red and green and between blue and yellow.

So Hering thought that opponency was generated at the level of receptors. So he thought that Young and Helmholtz were wrong. It wasn't the case that there were these three receptor types. He thought that there were these opponent receptors, something that was measuring the difference between red and green and the difference between yellow and blue and the difference between white and black.

And so there were all these vigorous arguments about that. Nowadays, as is sometimes the case with these debates, we believe that both parties were correct. In fact, we start out vision with three cone types that then get translated into opponent channels.

And that happens actually in the retina. So we've got these three different types of cones and that those get combined in different ways, taking differences and sums to generate opponent channels that you can see in retinal ganglion cells with these on/off receptive fields where the center will have input from one cone type and the surround will have input from another. And again, the pairings here are not arbitrary but rather systematic into these red-green and blue-yellow opponent channels.

So we're going to end there today. And when we come back, we'll continue talking about color and particularly get into color constancy and then how color is mediated in the brain.