[SQUEAKING] [RUSTLING] [CLICKING]

JOSH MCDERMOTT: Today, we're going to finish up this lecture on mid-level vision, and then start talking about lightness perception. And so I want to start-- I know it's been a little while since we last met. So I want to just start by briefly recapping where we left things last time.

So we talked a little bit about the broader structure of the visual system, the idea that it's divided up into a whole bunch of areas, that these areas are delineated by having retinotopic maps. We talked a bit about how those are measured. We reviewed the relationship between the left and the right visual hemifields and the representation in the cortex.

The central idea is that the left side of visual space is represented on the right side of the brain and vice versa. We saw some examples of this. And then talked about how the visual system continues to be organized in pathways. So you see these differences that start at the retina, also very salient in the LGN. And they continue into the visual cortex.

And there's this broad proposed organization into dorsal and ventral pathways. But there's lots of visual areas that are connected in lots of ways. But one of the characteristics of the organization of the visual pathway is the fact that we think of it as being hierarchical. So there are some regions that are closer to the input, some regions that are closer to the output. And as you move up the hierarchy, in general, the responses get more complicated, selective for complex structures that are often behaviorally meaningful. So we saw a few examples of that, where you can see neurons that are selective for particular individuals that you might see on TV or in movies.

And then we talked about this idea of moving from early vision to mid-level vision, where we think of early vision as involving a big set of measurements, often starting from the retina and going up to roughly V1. These are measurements that we often think were made with filters that are relatively linear. And then there's a lot of perceptual phenomena that we can link to these measurements-- things like the tilt aftereffect, the contrast sensitivity function, effects of adaptation, stuff like that.

By contrast, mid-level vision refers to things that are a little bit more related to inferences that we make about the world, that are based on the measurements of the early visual system that lead up to object recognition and scene perception, which are often thought of as high-level vision. And unlike early vision, mid-level vision is less well-linked to specific anatomical stages and to individual neurons. But there are nonetheless some linkages. And we'll talk a little bit about that.

So one key idea that we talked about last time is this idea that local measurements are ambiguous. This is important because we think of the early visual system as making local measurements. Neurons typically have associated receptive fields. These are small regions of visual space to which the neuron responds. These might be a fraction of a degree.

So it's like every neuron is looking at this tiny little portion of the world. And when you look at tiny little portions of the visual world, what you see-- what is evident in that small region is often ambiguous. And we saw a whole bunch of examples of that, like not being able to tell the difference between edges that are due to paint versus shading, or sometimes object edges that may be very apparent when you just view the image as a whole. If you zoom in and look at what the evidence is in a local region, it might be pretty weak.

And then we got into this idea of perceptual grouping, which is a particular perceptual phenomenon that we often view as part of mid-level vision. That refers to the fact that different elements of a visual display, in many cases will tend to group together. And typically, they'll be grouped in part because they're similar in some way. And so this was a major focus of gestalt psychology. And the phenomena continue to be important today.

So we saw a bunch of examples of grouping by similarity. And when we talk about the phenomenon of grouping, we're referring to the fact that, if you look at a display like this-- and this display is constructed by putting down, laying down a whole bunch of different elements. But when you look at it, there's a sense in which you perceive it as being organized into rows. And it's just subjectively obvious. So there's some aspect of how you perceive this in which those elements are grouped.

So you can also group things by common fate, as shown here, by texture, by proximity. We talked again, revisiting this idea of levels of analysis that we've talked about several times previously in this class, how you can take some problem that's being solved by the visual system or any sensory system and think about it in terms of these different levels-- computational, algorithmic, and implementation.

And you can often explain phenomena at these different levels. And in some cases, it's easier to explain things at some levels than others. And grouping is a great example of this, where you can dream up these implementation-level explanations of grouping, in terms of how neurons in the brain might be wired up. But you can also talk about grouping as a process of probabilistic inference, whereby what you perceive to be grouped are actually the latent causes of these elements in the world, the idea that there are these two clumps of dots here because maybe there's one process that generated one clump and another process that generated another clump.

We saw other examples of grouping, like good continuation, where you tend to see things as grouped if they result in continuous contours. So you look at an image like this. And you perceive this as being a circle on top of a square, even though there are alternative explanations that are possible. Grouping by closure-- so this is a whole bunch of elements that you could be perceiving as a repeating unit of different sorts. So each of those three building blocks could potentially be essentially an explanation of the elements that you see there. But almost everybody tends to see these things as circular.

And then we wrapped up by talking about this phenomenon of popout. So popout you can think of as the flip side of grouping. So things that are similar tend to group together. Things that are different tend to pop out as separate. So when you look at this, it's just immediately obvious that there's this one thing that's kind of different from everything else. And popout tends to work for simple, easily computed properties like color, or polarity, or brightness, or orientation, or size, or motion, or flicker, or depth-- here, we're using a drop shadow to indicate depth-- or shape.

So the point is that, as soon as I pop this display up, everybody immediately knows that there's one thing that's different from the rest. So that's the phenomenon of popout. You don't really have to search around looking for it. It just is immediately apparent. And that's not true for all dimensions by which you could distinguish one element for another. So I'm going to show you a display. And I want you to raise your hand when you see the thing that's different.

OK, good. You got it. But it took you a little while. You have to look around a little bit for it. So one of the S's is backwards. And that doesn't really pop out to the same extent. We're going to do the same thing here. Raise your hand when you see the thing that's different. Again, it took a little while.

So this is an example where the odd one out is distinguished by a conjunction of basic features. So there's a bunch of-- there's some things that are blue. There's some things that are green. There's some things that are X's. There's some things that are O's. But there's only one thing that's a blue X. And that takes a little while to find.

So element grouping and popout, again, this is a sloppy, intuitive description. But they're usually based on simple properties where we think of simple properties as plausibly being computed by the machinery of early vision that could potentially involve comparisons between filter outputs.

So we talked about how there's a computational level description of grouping, in terms of probability, where the Helmholtzian view of these things is that what we see is our best guess as to what's in the world based on the input data and our prior experience. And so grouping by proximity could potentially be explained in this way, by the notion that when things are close to each other in the image, there's a good chance that they were part of the same object in the world. And so we have a tendency to see them as part of the same object.

So that's an intuitively appealing view of grouping. And the question is whether that is something that can be made a little bit more precise. And grouping is, in fact, a nice place where of computational level description has been worked out in a fair amount of detail. And the essential idea is that you can actually measure-- at least for certain simple visual elements or properties, you can measure the likelihood that two elements are actually part of the same thing in the world.

And so here's the essential idea, that we've got a line segment here, and then a bunch of other line segments. And we can ask, just given the nature of images in the world, what is the likelihood that each of these other four line segments is actually part of the same thing as this line segment?

And the intuition here is that the edges of things in the world tend to be pretty smooth. And so this line segment is actually pretty likely to be part of the same thing as this one, whereas this one is not as likely. And there's some relationship there based on the difference in the orientation, maybe the position. So that's just an intuitive explanation.

You can actually measure this kind of thing and quantify it. And that's what is shown in this beautiful looking picture here. So what this picture is showing is the empirical distribution of edge orientations given a horizontal edge at the center. So supposing there is a horizontal edge at the center, we can then ask, what is the probability that there would be an oriented line segment at each of these possible positions and at each orientation at each possible position?

And so the way that this is displayed is that the line segment is color-coded based on how likely it is to be present, given, again, that there is a horizontal edge right here at the center. So what does this reveal? So the idea is that this is measured empirically. So the way that you would compute something like this is you'd have a big set of images. You'd have people go through and label the edges of objects. And each edge gets divided up into these little segments that are approximately straight and have orientations.

And so from that huge set of actual object contours in images of the world, you can compute this thing. And so what this tells you is that if there is an edge element that is here and that is horizontal, it's pretty likely that you'll have other edges out here that are pretty similar in orientation. So the idea is that edges of objects tend to be straight and smooth. And as you move away, there's also some likelihood that you have edges that are at nearby orientations because, again, there could be a smooth contour that passes through those points.

But if the orientation is very different, like it would be at these other orientations, either here or here, the probability is much lower. So this is a property of the world. It just says, how likely is it that there are edges at these different orientations and positions, given that there's one here? So it's a property of the world.

Now, the proposal, that computational level explanation of grouping, is that perceptual grouping, our tendency to see things as belonging together, has implicitly internalized these probabilities so that we would tend to see this line segment and this line segment as being part of the same thing because, empirically, they are likely to be part of the same object, just given the way the world works.

And so the proposal is that, either over evolution or over development, we have internalized this probability distribution, this property of the world. And in general, for grouping and for really most other things in perception, we still don't really know the extent to which this is a product of evolution versus development. It's probably some mix of both. It's usually pretty hard to tease those things apart. But the idea is that, over the course of evolution, and then as you grow up in the world, you get lots of experience looking at objects, picking them up. You learn what objects are. And one way or another, you internalize these relationships.

So the prediction then is that, if that is really how grouping works and why it is the way that it is, that you ought to see traces of these kinds of distributions in perceptual grouping. So we've already seen some of simple examples of this, grouping via good continuation, where it's just obvious from looking at these things that when you see something like this, you're going to see these two continuous curves rather than this kind of thing.

This has actually been studied in more detail and more rigorously. And this is one pretty well-known method for looking at this. So this is the experimental stimulus here. It's a bunch of Gabors. Remember, Gabor functions are the product of a sinusoid and a Gaussian. So each one of these things is a Gabor function. They're all the same spatial frequency, but they're different orientations.

And the idea is that this is a stimulus that contains a continuous contour amid a whole bunch of other Gabors. And the continuous contour is this one here. And what makes it a continuous contour is that there is this smooth relationship between position and orientation. And what you're supposed to see when you look at this is that that continuous contour pops out a little bit. You can look at that and see that there's this continuous thing amid this field of random-ish Gabors.

And so the argument that is a consequence of perceptual grouping. And so you could measure perceptual grouping by showing people displays like this and asking them to detect whether there is a continuous contour in there or not. And then you could vary the properties of how that contour is defined and see whether the grouping occurs when it is defined by the sorts of properties that you see in the world.

And so it turns out that this task of detecting the snake or the contour is easy if the orientation varies smoothly from one Gabor to the next. And it becomes more difficult if it varies less smoothly. So this is a diagram of how these experimental stimuli are constructed. So they're on a grid. And you can vary the orientation and the position.

And so this is a graph that shows what would happen when you ran this experiment. So again, you show people these displays. Sometimes, they have one of these contours in them. Sometimes, they don't. People have to say which is which. So this is a graph that shows proportion correct as a function of the angle between adjacent Gabors on the contour.

And so here's a case where the adjacent Gabor has changed by kind of a lot. So it's less smooth right, turns around like that. And what you're supposed to see when you look at that display is that it's harder to see the contour, because the relationships between the elements are less like the statistical relationships that you see in the world. And so what the graph is showing is that performance on this task gets worse as the angle between the elements goes down. These are two different participants. The two curves are two different durations, which, as the slide says, is not important for our purposes.

So the task is also harder if the orientation and the position don't co-vary so as to form a smooth contour. So here, the Gabors are all arranged such that their positions form a smooth contour. But the orientations don't follow that orientation or the global orientation. And so then you look at that display. And it's a lot harder to see that thing.

This one is also kind of cool. So here, the way that the contours are constructed is-- it's like the same as this initial one, where it's a well-behaved contour, but then you just rotate each Gabor by 90 degrees. So now, the orientation is orthogonal to the direction of the contour. And that is also hard to detect. Yeah?

STUDENT: So in these experiments [INAUDIBLE]

JOSH MCDERMOTT: No, we're cheating. Yeah, in these class demonstrations, yes. No. The experiment involves just seeing this and saying whether there's one of these contours in it or not. Yeah, and there's almost surely going to be some effect of looking at that and then looking at that, as you probably could detect. The problem is that if I just showed you this, you wouldn't believe me that there was something in there. So I've got to show you that one. I guess I could've animated it.

So these are all things that make it harder. On the other hand, randomizing the phase of the Gabors has no effect. So you can still see this thing really easily. But if you look carefully, the phase is different from Gabor to Gabor. So the black and the white stripes don't line up exactly. Again, we haven't really given you a clear explanation for why that is, but it's just part of the phenomenon. But presumably, you might be able to relate this to actually measurements that you would make in actual images.

So big picture here is that this phenomenon of grouping, which seems like this loose, subjective thing-- you look at these displays. And they kind of are organized in a certain way-- we think is actually rooted in the way the world works. So there are probabilistic relationships between the elements that you see in images and whether or not they're likely to belong to the same thing.

Your brain has internalized those. And that causes you to see these structures that are likely to actually correspond to objects in the world. And if the world were set up differently, with different statistics, then your perceptual grouping would presumably work differently. Again, that's a very difficult experiment to do. But that would be the expectation.

So as we said, we can talk about these things at different levels. So I just gave you a computational level explanation of grouping. The question is, what is the problem that's being solved? And what are the constraints that get imposed to solve it? The problem consists of an inference of which things are likely to be part of the same thing. You employ this constraint of knowing that the relationships between the elements of things in the actual world are not random. They have this structure. And you internalize that and apply that to make an inference. So that's the computational level description.

But we can also ask, how would this be implemented in neural circuitry? And there's a fairly well-known proposal for how this particular aspect of grouping of elements into contours might arise in the brain. The main idea is really simple. And it's that you could set the brain up in such a way that neurons that would be stimulated by the same contour would be wired up so as to excite each other, so as to enhance the representation of the contour.

So this is a simple diagram. So here, each one of these things now represents a receptive field of a single neuron. So it's located at a particular position in space. It's got a particular orientation. So these are different receptive fields of different neurons. The solid lines here represent excitatory interactions. The dashed lines represent inhibitory interactions.

So the notion here is that this receptive field and this one, or this one, and this one, they would excite each other. What that means is that, if they're both stimulated-- so let's say there's some orientation energy here and some orientation energy here-- then they boost the responses of each other. And that ought to cause that set of neurons to then accentuated in the neural representation. By comparison, these inhibitory interactions wouldn't cause that to happen. So this is an implementation level explanation of this kind of phenomena. All right. Any questions about grouping or levels of explanation?

So another place where grouping processes are evident are cases where we perceive contours in the absence of local image evidence. These are also often known as illusory contours. And these two displays illustrate this. So in this particular case, when you look at this, you all probably have the impression that there's a white strip that's vertically oriented.

And we call this an illusory contour because there's a sense in which it looks like there's an edge here. But actually, there's no contrast in the image at that location. The whole thing is just white. And in fact, the way that you create this illusion, if you will, is just by taking a bunch of these semicircle things and laying them down so that the endpoints all align. So that's one way that you could create this display.

Another way you could create the display is like the way that we see it, which is you take a whole bunch of ovals that are concentric. You put a bunch here and a bunch here. And then you put a white strip on top of them. And so that's really the explanation that you are implicitly inferring when you look at this. So you see this edge here, even though there's not actual contrast in the image.

Now, there's this other situation over here, where you're also perceiving an edge in the absence of explicit local information in the image because, in this case, it actually looks like this light gray rectangle extends beyond or behind the dark ovals. So again, in this region here, there's no contrast here. But you have the sense that there's a thing extending behind them.

So the difference between this and this-- well, there's a couple differences. One is that, in this case, the thing that you're perceiving is in front of the other stuff. In this case, the thing that you're perceiving in the absence of local evidence is behind the other stuff.

But the other thing, the other difference, which I think is deep and cool, is that, in this case, you actually see the edge. It actually looks like there is an edge there that you can see. In this case, like no one's going to say that they actually see the edge there. You sense that the edge is there. You perceive that it's there at some level. But you don't actually see it.

So these are often referred to as completion, in the sense that there's local evidence here and here. And then there's something that gets completed that causes you to see something in the absence of local evidence. This one is often referred to as modal completion. This is typically referred to as amodal completion. So modal-- I think it's Latin or Greek that refers to sensation. And so the idea is that you're actually seeing the contour here. Here, you're not actually seeing it. It's amodal. But there's still some sense in which you're representing a completed contour.

And so both of these things happen all the time. Again, most of the time, you're just not aware of it. But sometimes, there are these funny coincidences in images that reveal the completion processes at work. So this is a funny one, where there's a cylinder or something that's placed in front of this person's hands. And it makes it look like they have a really long finger.

Here, this is a case where there's a bunch of objects on the table. They're coincidentally covered up by this purple thing. And so it looks like there's a continuous table behind the purple thing. This is a really long cow. So it's these sorts of examples that really drive home that this is a real thing. On one level, it sounds very subjective, this idea that you're sensing this contour. But it really does affect the way that you perceive shapes.

And because occlusion-- so occlusion refers to the fact that something can be in front of another and block your view of it and part of it in some region. And so objects in vision occlude each other. So you have one object that's in front of the other. It blocks your view temporarily. And so this business of amodal completion happens all the time because occlusion is just everywhere in vision.

You might ask, well, what determines whether this completion happens? And one important property is what is depicted here, which is referred to as relatability. And this was popularized by two people, Kellman and Shipley. And so the idea is that when you look at this one, it looks like the two gray things are connected and are a single object, whereas, in this one, it really doesn't. It really looks like there's two different gray rectangles that are just sticking under the same black blob.

And so the intuition is that the contours that are on either side of the occluder, they should be aiming towards the same place behind the occluder and shouldn't be at too steep of an angle. So that's what this term relatability really refers to. It's whether the contours can be connected in a smooth way.

So here, we have an example of something that is relatable. So you can imagine this smooth curve joining the things behind an occluder. And these are unrelatable. There's an inflection point in the curve. And so again, this is something that's subjectively evident that is probably due to the statistics of the world.

It's probably a function of which edges are empirically likely to actually correspond to objects in the world. And the idea is that this kind of thing is more probable than this kind of thing. It's not that this could never happen. It obviously could happen. But it's just less likely. And you're making a probabilistic inference when you look at one of these displays.

So I want to conclude this lecture by talking about some evidence for how these illusory contours might be represented in the visual system. And one really interesting finding is that you can find individual neurons in an area of the visual system, V2, which is one stage beyond V1, that seem to respond to these illusory contours. So these were discovered by Rüdiger von der Heydt. he was a vision scientist who worked at Johns Hopkins for many years. And the very first paper was from back in 1984.

And so the idea is to measure responses to stimuli that have an illusory contour, so cases where you see something there where there's no local image evidence. And so the oval here represents the receptive field of a neuron. And the rest of the stuff here is the stimulus.

And so this is just a regular line that gets swept through the receptive field. And this is what's called a raster plot. So each row represents one trial where you sweep the thing through the receptive field. And the dots represent spikes that are fired by the neuron. And so the point is that there's just a lot of dots there, which is what you would expect because there's an edge in the image. You sweep it through the receptive field. And you get a response.

And down here, D is just what happens if there's nothing in the receptive field. So it's just spontaneous activity. So you don't have a whole lot of spikes there. And so the two interesting conditions are the ones that are shown here. So this is a really old figure. And it's a little hard to see. But this is an illusory contour.

So we've got a black line here and a black line here. And then this region here is also black. So there's no image contrast here. But if you looked at this, especially if this receptive field diagram wasn't there, you would perceive there to be a black line extending across here.

This stimulus, maybe if you're sitting all the way in the back, this actually looks exactly the same as this. But it's actually different in an important way. And that is that there's a thin white line happening here and here. And so this is an analogous example shown at better resolution, so you can get the idea.

So this is a stimulus where you would see an illusory contour. It looks like there's a white rectangle extending across between the two circles. Here, we've drawn these thin white lines, here and here. And that kills the percept of the illusory contour. If you look at that, you might actually see that there's a white rectangle that's behind some other white thing, like maybe you're viewing something through two round windows. But you don't see the illusory contour here that you see here.

And so the point here is that this is a very small change to the image that makes a big change to the subjective perception. And so von der Heydt used this to actually probe for responses to the illusory contour. So the idea is that you have these two stimuli that are almost the same, except for this really small difference. And in this case, you see the contour. In this case, you don't.

And what he found was that, in this case, you get a big response, so lots of dots on the raster here, action potentials. And in this case, you don't. So it's evidence that these neurons are responding to these illusory contours. And importantly, so this is interesting both just because it's interesting, but also because this is something that apparently differentiates V2 from V1. So you don't see this, or not very much of it, in primary visual cortex, something that emerges a little bit downstream.

Here's just another example. This is a different type of illusory contour, where, if you have these gratings that are slightly offset, you see an edge here, even though all of the actual orientation energy in the stimulus is perpendicular to that. And so this is showing you that this is-- so this is a neuron that's being stimulated with these lines. And so when you have a line that's nearly horizontal, you get almost no response. When you have a line that's nearly vertical, you get big responses.

And here, we have illusory contours with those same orientations. And the illusory contour, when it's almost vertical, gives you a big response, even though all of the lines that are in the stimulus are mostly horizontal. So the neuron's really responding to the percept of the-- or it's responding to some correlate of that illusory contour-- again, happening in an area of V2.

All right. So it's natural to wonder like, well, what are the mechanisms that give rise to these illusory contours. And I would say we don't entirely know. There have been proposals for models of illusory contours based on local low-level features. Here's a diagram for one such paper.

And so the idea is that you could have local receptive fields that would detect the elements of these displays-- the line endings, for instance. And you could imagine that you would try to create a neuron that would respond to these illusory contours in a way that's analogous to the way that we think orientation selectivity is produced by a bunch of center surround receptive field inputs.

Remember, we talked about that model of orientation selectivity where you have these center surround receptive fields in the LGN. And they converge on some downstream neuron. And if you have the right combination of receptive field inputs, you get an orientation selective receptive field.

So you can imagine the same principle applying here, that you would have a bunch of receptive fields that would be tuned to the individual elements of, say, a display like this. And then if those provided input to a downstream neuron, you would then get something that would respond to the illusory contour.

And that might be part of the story. But one really interesting property of these illusory contours is that they're very sensitive to manipulations that you wouldn't really think would affect the low-level responses that much. And so in particular, in this case, you actually don't really see a very strong illusory contour, whereas in this case, you do.

And so you have roughly the same number of line endings and the same amount of alignment in these two different cases. But in this case, the fact that the line segments are randomly oriented makes the illusory contour a lot stronger. And intuitively, one way to think about this is that there are other ways to actually explain the fact that these things are all lined up that don't necessarily require you to postulate that there is some object here that's occluding them.

So in fact, these are all exactly the same length. And they're all lined up. And so you might imagine that there was some other process that caused all these things to be lined up in this way, whereas, in this case, really the most likely explanation is probably just that there's an occluding surface that happens to be white. So it's really not so obvious how to account for these kinds of differences in terms of these low-level circuit-style models.

And as I've noted here, so this model here is probably-- I don't know-- 20, 25, 30 years old, something like that. This is the kind of problem that I think is ripe for revisiting in the current era. So we now have very different modeling tools than were available back when there was a lot of interest in this, in the '80s and early '90s. And so yeah, one of you could work on this, potentially.

All right. So we think that these are-- at some level, these illusory contours, they're illusions, right? We call them illusions because you're seeing something that's not there. But as is the case with most illusions, we actually think that these represent engineering solutions that help us see in the world.

And probably because of phenomena like this, that it just happens all the time, that there will be the edge of an object that just because the object happens to be pretty similar color to the background, at least in some place, there's not much local evidence for the edge. And so here, you have a log. And it's pretty similar in gray level to the background. And so locally, there's just not a whole lot of orientation energy.

But you can still see the edge pretty clearly, probably because, well, there's some high contrast stuff here and some high contrast stuff here. And so you get these mechanisms that contour completion. So this sort of completion is probably happening all the time without us really thinking a whole lot about it. What questions do you have about modal or amodal completion? Yeah?

STUDENT: This is just a clarifying point about the von der Heydt paper.

JOSH MCDERMOTT: Yeah?

STUDENT: When it's talking about some neurons in V2 responding to illusory contours, is it responding similarly to illusory contours as it does to just a line? Or is it [INAUDIBLE]

JOSH MCDERMOTT: Well, apparently. So that's kind of what this shows, right? So this is the response to just a line. And this is the illusory contour. I mean, it's not shown in these graphs. But this would be a neuron that is orientation-selective. So you change the orientation of the line and the response would go down. And you'd see the same orientation selectivity to the illusory contour.

So yeah, it does seem like it's like the same code for a regular contour that has some image contrast as for an illusory contour. And the next example, I think, also shows that maybe more clearly, where here you can see that this is orientation selective to the visible line. So it responds to this orientation, not this one. And here, you see comparable selectivity to the illusory contour. Yeah. Yeah?

STUDENT: [INAUDIBLE]

JOSH MCDERMOTT: What do you mean how are we supposed to interpret the two halves?

STUDENT: Like the two circles [INAUDIBLE]

JOSH MCDERMOTT: So this is a stimulus, OK? And the stimulus is just-- it just looks like this. So it's just a line. And this is being presented to-- essentially, presented to a neuron. So the neuron's receptive field would be here. And then these are raster plots. So a whole bunch of these are presentations of this stimulus. And the point is that you don't see very many dots here because the dots are spikes. And so the neuron is just not responding to the stimulus.

These correspond to presentation of this stimulus. And there's a whole bunch of dots here because it's orientation-selective. And it's responding to that line and not to that one. And this is the same thing, only presenting these illusory contours at these two orientations. So these are called raster plots. These are very common in old school neurophysiology papers. So again, each row represents a trial-- so a presentation of a stimulus. This is time. And each little dot here is a spike.

OK, so one last related concept is this notion figure and ground. So empirically, we often interpret images as consisting of figure and ground. And this is a really famous illusion that illustrates this point where, when you look at this, sometimes you see a vase and sometimes you see two faces. And you tend to see one or the other. And it's bistable. So if you stare at this, sometimes you'll see two faces. And then it'll switch and it'll look mostly like a vase. It goes back and forth.

And the point is that, at any moment in time, you tend to see one or the other. So that edge is always there. But the edge is either owned by one side or the other. So since the edge is caused by one side or the other, and that side has caused the figure, the other side is called the ground.

So if this one doesn't work on you, look at this. So this is one of these displays that makes your brain hurt. And what is happening here is that these little crosses, for brief periods of time, move. And that causes them to look like objects and figure. And then they become static. And something else moves. And so the what's figure and what's ground is changing from moment to moment. But you'll initially see these white crosses, and then see these black crosses. And so it's just supposed to make the point that you tend to assign edges as belonging to one side or the other. And the side that belongs to them, that owns them is "figure."

So what's figure and what's ground is inherently ambiguous. So it's another example of an ill-posed problem. But the visual system relies on many cues. And again, these are things that were studied by the gestalt psychologists. So for instance, things that are symmetric tend to be seen as figure. Things that are smaller tend to be seen as figure. So the dark orange things here look like figure. And the light orange things look like ground.

Parallel lines tend to be interpreted as the borders of figure. So the things that are parallel here tend to be seen as figure. And so as with perceptual grouping, this was initially just studied with this very empirical, subjective, intuitive methodology. People would make these displays and look at them. And oftentimes, the effects were just obvious.

And intuitively, we think-- people thought for a long time that these relationships are the way they are because of the way the world is. But it was hard to make that really argument precise. But nowadays, we have the ability to make these large scale measurements of images and measure probabilities of things. And so we can actually give teeth to these kinds of computational-level explanations.

And so this is just an example of a paper that provides evidence for the idea that the cues to figure and ground are actually rooted in statistics of the world. And so the way that this works is-- this was a group at Berkeley that took lots of images of the world, asked human observers to label all of the edges of objects in the image. And they probably had a computer vision algorithm that would help them out.

You find all of these edges. And then you get humans to label the sections of edges that correspond to a figure. So in this particular case, there's a person. And so that boundary is the boundary of the person. And so you have a way of labeling which side of the contour is the figure and which side is the ground.

Here's another one. This is like the edge of the cow. And so you have a way of labeling that. So you do that for all of the different edges here in the image. So for every point on every edge in a big set of images, you have what, in theory, is ground truth of which side of the edge is actually the figure, where ground truth here is derived using the entirety of a bunch of human visual systems.

So the question is whether the kinds of cues that people classically talked about out actually fall out of these kinds of probabilities. So remember, we talked about the idea that things that are smaller are more likely to be seen as figure. And so you can take a point on one of these contours. You have a little circle that you draw around the point. There's a contour that runs through that. And you can compute the area on either side of the contour.

And so in this particular case, the F stands for figure, the G stands for ground. And the figure side of the contour has less area than the ground side of the contour. And you can quantify that as the ratio between the two areas. Similarly, we often think that things that are convex are likely to be figure. So again, here's a case. This is like the foot of the bear. So that's the figure. That's the ground. This is the heel of the bear. So this is a convex contour. And so you can measure the convexity.

And so the question is whether those cues are predictive of whether something actually is figure or actually is ground. And so this is the result of the analysis. So these are histograms that are plotting the frequency of occurrence of these different measures. This is the size measure and the convexity measure for the figure side of the figure and the ground side of the figure.

And so the point is that, for this size cue, the figure side of the figure tends to be smaller than the ground side of the figure. The red curve is to the left of the blue curve. And for the convexity cue, the figure tends to be more convex than the ground side of the curve. So it's to the right.

OK. So this is essentially saying that these so-called cues that were often proposed by gestalt psychologists, that they actually are borne out by the statistics of the world. So the things that are figure do tend to be smaller and do tend to be more convex.

All right. So the question is, these local cues, if we actually measure these cues, can we actually use them to classify which side of a contour is the figure? And so that's what this graph is showing. And so there were three cues here that were measured. I told you about two of them. There's this other one, L, which we can skip over in the interest of time. The point is there are three cues.

And so you can build a classifier that has access to either all three, or pairs of them, or one of them. And then you can vary the radius of the window over which those cues are measured. And the graph here is plotting classification performance, so how accurately you can predict which side of the contour is the figure based on individual cues or combinations of cues.

And what you can see is that, as you make the window bigger, you do a little bit better. And then you reach some radius and things level off. And unsurprisingly, when you have all three cues-- that's the red curve-- you do better than when you have fewer cues. But what's interesting is that performance seems to level out at about 75%. So that's well above chance, but it's not perfect.

So let's just assume that humans viewing these images would be close to 100% correct. And in fact, I mean, humans were used to label the images. So in some sense, they have to be 100% correct. So this is showing that these local cues allow you to predict which side is figure, but certainly not at 100%, only at 75% So what do you think might account for the performance gap? Mm-hmm?

STUDENT: Global context of the scenes.

JOSH MCDERMOTT: Yeah, so the global context of a scene, that could be it-- so your knowledge of the world, that you recognize that something is a cow and cows are going to be figure, stuff like that. So yeah, that could be it. Anything else? Yeah?

STUDENT: Yeah, so is this just looking at one circle at a time and then being like, I think this is the figure versus this?

JOSH MCDERMOTT: Yeah.

STUDENT: Well, it's probably the fact that [INAUDIBLE] Well, then leaves one figure, even if I'm classifying 25% of them correctly.

JOSH MCDERMOTT: So you're saying that if you just combined information from all of those little windows, you could--

STUDENT: Yeah.

JOSH MCDERMOTT: Yeah, so that could also be true. Yeah, and that's a slightly different version of the global context. It's not necessarily taking into account knowledge of the world, but just combining cues, yeah. So yeah, so that's possible. Yeah?

STUDENT: I might be restating some things, but maybe it's related to grouping?

JOSH MCDERMOTT: What do you mean by that, maybe it's related to grouping?

STUDENT: So if a stimulus seems to appear to be an edge. And other things also seem to point to something bigger. They could be put together and considered one thing.

JOSH MCDERMOTT: Yeah, yeah, yeah. That's true. So the strategy of combining cues could definitely be related to grouping via good continuation. Yeah, those are all good suggestions. Another possibility is that these three cues are not exhaustive. These are just three things that human scientists thought up. It could be other stuff that you could measure.

But at any rate, it's one approach to showing how these perceptual phenomena can be related to the nature of the world-- in particular, like the probabilities of things being one way or the other in the world. Any questions about figure or ground?

OK. Last thing I want to tell you about-- this is also pretty cool. So there's also a pretty intriguing neural correlate of these figure ground relationships that also is found in area V2. And this was also from the lab of Rüdiger von der Heydt, who was very interested in these mid-level visual phenomena.

And these are the existence of neurons that seem to encode the direction of border ownership. And so what we mean by border ownership is it's really synonymous with figure and ground. So the idea is that you have some contour. And one side of that is figure. The other side of that is ground. The side that is figure, we say, owns the border, in the sense that that's the object that actually caused that contour.

And so here's essentially the phenomenon. So this is the same kind of display that we've been looking at. These are raster plots. So each row is a trial, so a presentation of the stimulus. The x-axis is time. Each little line segment here is an action potential. This is the visual display. And you can see the receptive field of the cell as that little oval that is centered on part of the display.

And so the idea is that the receptive field is positioned on an edge there. And we've got four stimuli here. And what is varied is both the sign of the contrast-- so here, we have a light square on a dark background. Here, we have a square in the same position. But now, the square is dark gray and the background is light.

And so now, the stimulus that's inside the receptive field has actually flipped the sign of the contrast. Here, we've again got a square on a background. And the stimulus that's inside the receptive field is identical to the one up here. So on the left side, we have light. And on the right side, we have dark.

But now, the square is on the opposite side. So the idea with these two displays is that, at this particular point on the contour, in this case, the left side owns the contour. And in this case, the right side owns the contour. The local stimulus is the same. So this is another example of this local ambiguity. But the rest of the image gives you a bunch of information indicating that either this side is figure or this side is figure.

And then here's the alternative version where, again, the square is on the right side, but the contrast is flipped. So now, the local stimulus is the same as the one up here. All right. And so what you see from this is that this neuron is giving you a big response whenever the figure is on the left side of the contour. So it's when the left side owns the contour. And it gives you a small response whenever the figure is on the other side.

So this seems to be coding the direction of border ownership. It's a representation of which side of the contour is the figure. So this doesn't tell us how the neuron figures out one side is figure and one is the other. It just says that this is evidence that quantity is being represented. Yeah?

STUDENT: What is the y-axis on these graphs?

JOSH MCDERMOTT: These are different trials, yeah. Yeah, so each one is just a different stimulus presentation. OK. So big picture here, talking about mid-level vision, these interesting aspects of perceptual organization, they're rooted in probabilities in the world. And we have various little bits of hints of how they're coded in the brain. We're very far from having complete explanations of these things, in terms of neural circuitry. But in terms of illusory contours and border ownership, there are some indications that V2 is representing some of the relevant quantities.

So summary of what we've talked about-- at the very start of the lecture, we talked about how the visual system consists of many regions that are distinguished by retinotopy and that are organized into hierarchical pathways. Mid-level vision loosely refers to a set of perceptual phenomena that involve inferences about the world. They're not always clearly linked to neural mechanisms or stages. Although, sometimes we have little glimpses of that.

We talked in particular about grouping, and popout, and figure and ground. In both cases, they seem to be linked to the statistics of the world, which nowadays can be measured and compared to human perception. And then we also talked about modal and amodal completion. These are processes that seem to represent perceptual inferences of object edges, even in cases where you don't have local evidence for those edges. What questions do you have about mid-level vision?

OK. Let's talk about lightness perception. I reorganized the course just a little bit for logistical reasons. So this used to come later. But we're doing it earlier. And we're going to talk about lightness perception, and then color perception, and then get into depth and motion.

All right, so here's the problem. So organisms need to estimate surface pigmentation. So pigmentation tells us a lot of important things about what things are made of. So for instance, it would tell you whether a piece of fruit is ripe or not. It might tell you whether your baby's got a fever or not. Lots of important things about the world are signaled by pigmentation.

And so lightness perception and color perception are really just about estimating surface pigmentation. Lightness perception really is the one-dimensional problem. So we're going to first consider the problem ignoring wavelength and just talk about the problem of estimating what we call reflectance, where we're going to define reflectance as the proportion of light that a surface reflects.

So this would be like if there was a world where everything was just on the continuum between black and white. And so the term lightness is used to refer to the perceptual correlate of this notion of reflectance. So it's like our estimate inside our heads of the extent to which something is white, black, gray, dark gray, et cetera.

So the whole reason that we can talk about this is because many surfaces in the world exhibit something that's well-approximated as Lambertian reflectance. So a surface that is Lambertian scatters light uniformly in all directions. So here's the technical details. So I is the amount of incident light. Theta is the angle of the illumination relative to the surface normal.

And so what this means is that, if the light is directed perpendicular to the surface, there will be a lot of light that is scattered. And if it is oblique, there will be less light that is scattered. But the key concept is that the light gets scattered uniformly in all directions. So that's the definition of a Lambertian surface or material.

So that turns out to be important because a Lambertian surface's reflectance can be characterized by just one number, which is the proportion of light that gets reflected. So this allows us to refer to the lightness of a surface and have that have some meaning. So in the world of lightness perception, which is what we're going to be talking about in this lecture-- so this is where we don't care about wavelength-- this means that the problem of lightness perception or reflectance estimation can be summarized by this equation.

So there's this quantity of an object in the world called reflectance. That's the proportion of light that the surface will reflect. There is illumination that's coming from a light source. And the luminance that is reflected by the surface is the product of the reflectance and the illumination.

So critically, this is what you measure. Light is reflected off of a surface. And it enters your eye. And your eye records that. But that light is caused by two different variables-- the illumination, so the light source in the world, and the reflectance of a particular object.

All right. And so all of this really, it makes sense if we think of surfaces as being Lambertian. So importantly, many surfaces are not fully Lambertian. So this is a case where you have something that's pretty Lambertian. Sometimes though, you get non-Lambertian reflectance, in particular specular reflections. So this is a case where the light comes in and gets reflected in some particular direction. So that's like what a mirror would do.

And oftentimes, the surfaces that we encounter in the world, they have a combination of diffuse and specular reflection. So you get something like this, where there's a whole bunch of light that's scattered uniformly. But then there'll be a special direction that emits a fair amount of light.

And then in addition, sometimes what you can have is this diffuse reflectance, and then some specular reflections that have some amount of blur, that get blurred out spatially. And this is a function of the material properties. And our visual systems are amazingly attuned to these properties. And you can get a sense of this just by looking at computer graphics.

So this is a case where there is a specularity on the surface. So that's what this is. So this is rendered with a model of a specular reflection. And that causes the ball to appear to be somewhat glossy. And one way to see that is to just selectively remove the specular reflection. And now, the ball looks like it's matte. So the material appearance is quite different. So the local presence of those specularities causes you to infer this property of the entire surface as being shiny.

At the start of the class, I showed you this photo that I like, where you look at this person's legs and they look shiny. So I think what actually happened is there's some suntan lotion that just got streaked along. But your visual system mistakes this for specular reflections. And it looks like the person's legs are shiny.

Here's some graphical rendering of different surface materials that vary in their roughness. So the surface roughness will determine the extent to which the specular reflections are blurred and the amount of specular reflectance. So it's how shiny the thing is. And so all of these look like spheres that you might encounter in the world. And they may look like familiar objects. And they just differ in the way that they reflect light. And your visual system is pretty attuned to that. So there's a whole field now of material perception that relates to stuff like this.

So we often, in the context of lightness perception, will just explore what happens if surfaces are purely Lambertian. But in the real world, things are more complicated. There's additional complications, which is translucency. So translucent objects are those where light enters an object, and then emerges a little bit later.

So you can have a light source. There can be some diffuse reflectance, some specular reflectance that will cause the perception of gloss. And then photons can enter the material, and then bounce around, and then later emerge. And that will cause something to appear translucent.

Here's another picture that gives the basic idea. So translucency is actually really important in perception, in part because skin has elements of translucency. So this is a diagram that shows you what happens when light interacts with skin. So your skin consists of these different levels of tissue. And there will be some amount of reflection off of the skin. But then photons enter the skin, and then eventually emerge. And so if you want to convincingly render the appearance of a person, you have to model what's called subsurface scattering, so this element of translucency.

And in inanimate objects, the rendering of subsurface scattering can really change the appearance of material. So this is a little figurine. The one on the left looks like it's ceramic. And the one on the right looks like maybe wax. So this is done with computer graphic rendering.

So that's just a brief little aside to say that surfaces in the world are rich and complicated in the way that they reflect light. And your visual system is highly attuned to the way that they reflect light. And we could have an entire class just about that. But it's nonetheless the case that many surfaces are approximately Lambertian. And you can learn a lot just from thinking about the Lambertian case, even though it's a little bit of a simplification.

All right. So the focus of this lecture will be on the phenomenon of what is called lightness constancy. So this refers to the fact that people are pretty good at correctly estimating reflectance. So I can show you this eraser. And you'll probably say this is pretty black. I can show you this piece of paper. And everybody would agree this is white.

But we're in a room that actually has relatively low light levels. So if we walk out there in the atrium, the amount of illumination will go up a fair bit. But if we take the eraser and the white piece of paper out there, this is still going to look black and this is still going to look white. And if we go out on the street, out on Vassar Street, there's a lot more illumination than there is in the atrium. This will still look black and this will still look white.

All right. So the point is that the illumination changes. And that means that the amount of luminance that's being reflected off of the surfaces will change dramatically from environment to environment. But our perception of what the surface is made of, our subjective sense of lightness, which, again, is an implicit estimate of the reflectance, that's relatively unaffected by that. So a gray piece of paper looks about the same shade of gray indoors as it does outdoors, despite reflecting dramatically more light outdoors. So this suggests, anecdotally, that we're fairly accurate in estimating the reflectance of everyday objects.

So I want to end by giving you an additional piece of evidence that we estimate reflectance. And that lies in the subjective observation that luminance changes that we interpret as being due to illumination look really different when we instead interpret them as reflectance. And so we're going to attempt to do this demo. And I want to explain to you how this is supposed to work. And then we'll try it out up here.

So here's what happens. We're going to turn shadows into paint in your head. So to do this, you place an object and a light source such that a shadow is cast on a piece of paper. And so we're going to do it up here. We're not going to do it right now because we have to turn out the lights. So we're going to do it using this lamp, which is on, good.

So the object is going to cast a shadow on the paper. What we're going to do then is we're going to take this black magic marker. And we're going to trace the outline of the shadow. And it will look like the region within the traced outline has a different color than the rest of the paper.

Now, in fact, the number of photons that's being reflected by the paper and the pattern inside here will be exactly the same. We're just going to add a little outline around it. But your brain will interpret those photons completely differently. And the intuition is just that, well, shadows never have these black outlines. So when I put the black outline around it, your brain refuses to interpret that as an illumination boundary.

And so the point here is that this is going to illustrate that your visual system codes reflectance and illumination as very different things, more evidence that you actually estimate reflectance. There is this observed image intensity-- this is luminance. This is registered by your photoreceptors-- that is the result of the incident light, the illumination, and the reflectance. That's a quality or property of the pigmentation of the surface.

So you can think of the generative process as this. So at every point in space, there is an amount of incident illumination. There's also a certain amount of reflectance. Those get multiplied. And that generates a luminance image. So at every pixel, you have one number. And you want to estimate two-- or at least one of the two. You can't unmultiply. So it's a classic example of an ill-posed problem.

But humans seem to do it. So we correctly perceive reflectance most of the time. And this is the phenomenon of lightness constancy. So how do we do it? And what we're going to talk about when we resume next time is a variety of illusions that provide some clues as to how we do it.

So this is one of the very earliest lightness illusions. It's called simultaneous contrast. So it's an illusion because these two squares are the same physical gray level. But when you look at them, it probably looks like the one on the left is lighter than the one on the right. Now, you'll also probably notice that the one on the left is embedded in a black-- or a dark surround. And the one on the right is embedded in a light surround. And indeed, that's what causes the one on the left to look lighter than the one on the right. But both of the squares have the same luminance, the same physical intensity in the image.

And when we come back, we will talk about some different classical explanations for this and what it tells us about lightness perception. And we will also get to much more striking lightness illusions. So to be continued.