[SQUEAKING]

[RUSTLING]

[CLICKING]

JOSH MCDERMOTT: Let me just kick things off by showing you a picture. So this is what is called a flat map of the macaque brain. So the cortex is this sheet which, in your head, is kind of folded up into the shape of your brain, basically. And a flat map consists of taking the cortex, making a few cuts, and then kind of flattening things out, such that you can see the entire cortical surface at once. All right.

And this is a flat map where what are considered to be different regions of the visual system are outlined, and different areas are in different colors. All of the regions here that are colored are considered to be part of the visual system. That means that they respond when people are looking at stuff.

And so this is the intact macaque brain. And so you can see that that the visual system is the back half of the brain, along with some stuff here in the frontal lobe. And this is kind of what it looks like when you flatten it out. And so there's a lot of different regions, and it's a big chunk of the brain.

All right. So what is it that distinguishes the different visual areas? Well, the main criteria by which you typically distinguish visual areas is retinotopy, where each visual area typically contains a retinotopic map. So a map of the entire visual field. And so the visual areas can be localized by measuring retinotopy. In this case, by measuring how the neural responses change with stimulus location.

So this is an example where a participant in an experiment is looking at this display. So they stare at the fixation point. And there are these annuluses that kind of gradually expand. So there are checkerboards, which are these high-contrast stimuli, which produce big visual responses. So the idea is that it expands. And so the location of the visual stimulus varies over eccentricity as a function of time.

And in this case, you have a pie wedge of a checkerboard that rotates around. So you map out polar angle. And so in each case, every point in the visual cortex is color-coded, in this case, as a function of the eccentricity at which the stimulus evokes the biggest response. And so you can see that you get this kind of gradient here. And in this case, as a function of polar angle. And you get another gradient that's kind of orthogonal to this.

So this is the retinotopic map that we have seen before. So we talked a lot about this in the context of primary visual cortex. So remember this diagram where this describes how the two visual fields get mapped onto your visual system. So each eye can see part of both visual fields, but then the contralateral visual field kind of gets-- or I should say, the left visual field gets mapped to the right hemisphere, and the right visual field gets mapped to the left hemisphere.

So this is an example of an experiment where you can see a whole bunch of different visual regions. So this is a one-- this is from one hemisphere. And here, the color is mapping out the polar angle at which the stimulus evoked the biggest response. And so you can see in what's considered to be area V1, primary visual cortex, the color kind of sweeps from green to red all the way up to dark blue. And so this is one hemisphere. And so you're only getting one hemifield, but you see the full range of visual angles.

So then what happens, once you get to what's considered to be the border of V2, is you start to see a lot of blue and green. So there are these gradients, but they're blue and green, and there's not a whole lot of red, whereas below V1 you see gradients from red to green. And so this is an indication that there's kind of a separation of the upper and the lower hemifields. But they're inverted.

So the lower hemifield is mapped onto the region above V1, and the upper hemifield is mapped onto the region below V1. And so these visual areas have been given names that are a little bit complicated sometimes because of the nonlinear historical trajectory over which they were discovered. But you've got V1, V2, V3, and then there's this other one called V3a, because they had already discovered V4, and then they realized there was this extra region and they had to squeeze it in. So there's stuff like that.

All right. So there's lots of these regions. The key idea is that we distinguish these different regions via retinotopy. And so, in particular, what happens when you get to the border of a visual region is that the sign of the retinotopy flips. So here it's going from green to red. And then it switches back to going from red to blue. And the same thing happens with tonotopy in the auditory system, where you see a reversal of the sign of the map.

And so in the early days, long before there was fMRI and you could look at this stuff in humans, people were doing this in non-human animals with electrodes. So they'd have electrodes in the brain. They'd be measuring receptive field locations and moving the electrode around and observing how the location of the receptive fields kind of varied across the cortical surface. But the same kind of principles applied. It's just with fMRI, you can look at everything at once. And so it's helpful for looking at these maps in this way.

So another kind of important concept that we've already alluded to is this idea that you can think about the visual system as consisting of pathways, and that these pathways kind of have their roots in the retina. So we've talked about how you've got midget cells and parasol cells and that they project to these different parts of the LGN.

Now, OK, this diagram has this other slightly annoying complexity to this, which is that occasionally people change the nomenclature with which retinal ganglion cells are described. And so sometimes the parasol cells are referred to as M cells. I guess they thought this would make it easier to remember that they project to the magnocellular layers. And sometimes the midget cells are referred to as P cells.

So it's just a gigantic mess. But you'll all keep it straight. But the essential idea is that you have these two different populations of neurons in the retina. They have different properties. And they project to different layers of the LGN. Those in turn project to different parts of primary visual cortex, these different sublayers. And then those in turn project to different parts of subsequent areas.

So area 18, which is often known as V2, has these different subcomponents to it. And so there's this degree of functional segregation that kind of persists. And it persists even up to areas that are fairly deep in the hierarchy. So inferior temporal cortex and the parietal lobe.

And so one way to think about this organization that remains very common and probably, to some extent, true, is this idea that you can think of two main pathways in the visual system-- one that extends ventrally, that mediates object recognition, culminating in inferotemporal cortex-- it's often called the "what" pathway-- and then one extending dorsally, culminating in the parietal lobe, that's involved in the localization of objects, often for the purpose of mediating actions. But to some extent, this division has its roots earlier in the visual system.

So this is a really famous picture. It kind of looks like a horrible subway map, but it's actually a diagram of the primate visual system. So each little box on this picture is a visual area. So distinguished by retinotopy. And then the lines between them represent connections, or I guess, fairly dense connections, dense enough that they considered it worth plotting on the graph.

So these are the retinal ganglion cells. You've got these two classes predominantly, the two different layers, types of layers of the LGN, V1, V2, and then an assortment of other areas. And so you can see that the things to take away from this are, one, there's a lot of areas. Two, there's a lot of connections. Three, these are organized and conceptualized in a hierarchy.

So we start out at the beginning of the system where the light enters. And then some of these regions are kind of situated more deeply into the system than others. So that's a very important idea, is that the sensory system is hierarchical. So some regions are getting input from others and then providing input to others.

And look at what happens when you move more deeply into the system. So if you record from neurons in these different areas and try to understand what they represent or what they're responding to, again, another common theme is that as you move more deep into the system, the responses get more complicated. All right.

And so there are very famous examples, so famous that you've probably encountered them in other classes. So this is an example of a neuron in inferotemporal cortex that's selective for faces. So it will respond a lot when you present it with an image of a face. Less so when you blur out the eyes. It can even be kind of a schematic of a face, and you get a pretty good size response.

So these neurons are much harder, much more difficult to describe mathematically than the neurons that you see in V1. But often selective to more complicated things that are more behaviorally relevant. And so there are some pretty famous examples at this point that really illustrate this point. So how many people have heard of the Jennifer Aniston neuron? Yeah. So that's one that made its way into pop culture.

So this is in inferotemporal cortex of a human who evidently watched some TV or movies. And you can see-- so the diagram here shows example images and then, underneath it, the response of this particular cell. And so you can see that there are these two different images of Jennifer Aniston. They both produce a pretty good sized response.

This is kind of non-trivial because the images, if you look at the pixel intensities, they don't really have a whole lot in common. Even in abstract terms, I mean, there's glasses being worn in one case and not the other. There's just lots of differences. And it doesn't respond to lots of other images. Looks like Brad Pitt decreases the response a little bit. You see stuff like this deep in the system.

This is my favorite example. This is a neuron in entorhinal cortex. I don't know how easy this is to see, but this is somebody who evidently liked *Star Wars* because the neuron responds to images of Luke Skywalker, but it also responds to the words "Luke Skywalker" in text. It also responds to someone saying the words "Luke Skywalker." And there's a moderate response to Yoda.

So you move deep into these sensory systems and things get more complicated. So that's a common theme. And we'll elaborate on that in much more detail and much more rigor in subsequent lectures. And a lot of times, one of the other things that you often see signs of as you move more deeply into sensory systems is that the responses become more invariant.

And remember, at the very start of the course, we talked about one of the challenges of, in particular, recognition tasks-- so one of the things our sensory systems are good for is helping us recognize things. And one of the challenges of recognition tasks is that different images or different sounds of the same thing in the world can often physically be really different.

And so these are all different images of a house. There's in fact an image of the same house taken from two different viewpoints. So you can both tell that these things are all houses. You can also pick out which two are the exact same house. But of course, the pixel intensities are completely different in all of these different cases. So somehow or another, you have to construct representations where those invariances exist.

Now, what we're going to talk about today is not this kind of stuff so much, but something that's kind of in between these very complicated recognition tasks and what we call early vision. And this normally is what is referred to as mid-level vision. So remember, the big picture here is that perception involves inferring the structure of the world from measurements of energy that are generated by the world. So in the case of vision, these are patterns of light.

And so far, the past three or four lectures, we've been talking about what is often referred to as early vision. So early vision we often think of as like a set of useful measurements. So you get this image as an input to the retina. And then there are a set of filters that are measuring different things about the image in different positions in space. So in V1, for instance, the various kinds of receptive fields give us local measurements of orientation, contrast, disparity, color, spatial frequency, and so forth, sort of like the ingredients of images.

Now there are also perceptual phenomena that are linked to these measurements that also fall under what's called early vision. And those tend to be phenomena that people classically think are explained by those measurements. So for instance, we talked about these adaptation effects, like the tilt aftereffect or spatial frequency adaptation and the effect on the contrast sensitivity function. So there are perceptual phenomena that fall under the rubric of early vision. They often are used to make inferences about the mechanisms of early vision.

So mid-level vision, by contrast, typically refers to processing stages that involve inferences that, in some way, are about the world, that are based on measurements that are made in early vision, eventually leading up to object recognition and scene perception, which typically people would call high-level vision. So we got early vision, mid-level vision, and high-level vision. And these are loose and sloppy terms, but they're nonetheless kind of used to indicate different parts of the field and different aspects of visual perception.

And the other thing that I should say is that unlike early vision, where there's often fairly reasonable linkages between some of the associated perceptual phenomena and neurons, typically retina, LGN, and V1, mid-level vision and the various perceptual phenomena that are associated with it is less well linked to specific anatomical stages of the visual system and to individual neurons. And so in some cases, those relationships exist, and we'll talk about them, but they're not as tight.

So one important theme that's central to mid-level vision is the idea that local measurements are ambiguous. So neurons in the visual system in some sense view the world through these little apertures, right? They have receptive fields. So that means that there's a region of visual space that drives their response. So they're looking at the world through these localized spatial receptive fields, measuring something that happens here or happens here or happens here.

And these local measurements are often quite ambiguous. So they can be ambiguous in terms of what is actually causing the image intensities that's within that local region that's being measured. So this is an image that was designed by my PhD advisor, Ted Adelson. And it depicts a simple thing that looks like it's painted different colors.

And so you have these different kinds of edges. So you have this edge here and this edge here. And inside those apertures, the image is exactly the same in the two cases. There's light gray on one side and dark gray on the other side. But you can just tell from looking at it that this is a change in pigmentation. It looks like maybe a change in the color of paint that was used to paint it. Whereas this is due to what's called shading. The fact that there's different image intensities there is because the surface orientation changes.

So of course, you look at this thing, and your visual system, acting in concert, is able to come up with the correct interpretation. But the point is that the local measurements that are being made early in your visual system are ambiguous. And an individual local measurement on its own is not enough to tell you what's actually happening in the world to cause that particular pattern.

You can see similar kinds of things in actual images. So these are some interesting cases where-- so each of these patches here-- this one, this one and this one-- are kind of close-ups of three regions of the edge of this log. So there's just a log on a background of stones. And again, when you look at the image, it's obvious that there's the edge of a log there.

But then, if you zoom in to what's actually evident at the level of a receptive field, you can see that it's really not very clear. So in particular, like this case, for instance, if you look closely, the edge of the log is kind of right here. But in fact, all the contrast in this particular local region is up here because there's a shadow being cast by a rock that's right next to that edge.

And so the point being that if all you had to analyze was this, and you were trying to figure out where the object boundary was, it'd be pretty hard to do. And the regions vary in terms of how much evidence there is for that edge, but they all have some degree of ambiguity. And that's actually a really useful exercise to do. If you really want to appreciate this more, you can take some images and view them through a little aperture. And if you look around with the aperture, it's really hard to make sense of what's going on.

So local measurements are ambiguous. In order to make inferences about the world, somehow they have to be combined in some way. And so one kind of phenomenon that has some relationship to that general idea and is typically associated with mid-level vision is perceptual grouping. So this refers to the fact that things that are similar tend to subjectively group together, in the sense that they appear to be part of the same thing.

And there was a movement in psychology a long time ago called gestalt psychology. And one of the main things they were interested in were these grouping rules. So what is this all about? So if you look at these two images, there's a sense in which you see these images as consisting of rows. You look at them, and the natural description is that you have rows of circles and squares or black dots and green dots.

And so that subjective sense, that these circles kind of belong together and these circles belong together, is called grouping. And in general, similarity is a major factor that kind of determines grouping. Common fate-- so the fact that things move together-- that has a really big effect. So that thing that kind of looks like a number 4, they all look like they're part of the same thing.

There's texture grouping. There's grouping by proximity. So again, you look at the thing on the left, and you see these horizontal rows, presumably because the circles are kind of closer together in the horizontal dimension than the vertical dimension. You look here, and you tend to see this as two groups of dots.

All right. So there's lots of different ways to think about these sorts of phenomena. And you remember this idea that there are these different levels of analysis. There's the computational level where we can talk about the problem that's being solved and the constraints that allow it to be solved. There's the algorithmic level. There's the implementation level where you talk about how you would describe things in terms of neural circuitry.

And it's possible to think about grouping at all of these different levels. So for instance, we can give an explanation of grouping in terms of neurons, for instance, by saying that there are neurons in the brain that respond more strongly when their neighbors, which will have nearby receptive fields, are also responding. And in fact, we'll see some evidence that there's probably some truth to that.

And so you get a strong response when many dots are in a line or in a clump. So that's one way to explain how the perceptual effect of grouping would come about from neurons, but it doesn't really tell us why that would happen. So alternatively, we could talk about grouping in terms of probability. So the Helmholtzian approach to thinking about this at the computational level would be the idea that what we see is our best guess as to what is in the world based on the input data and based on our prior experience.

So you might suppose that when things are close to each other in the image, there's a good chance that they're actually part of the same object in the world. And so we have a tendency to see them as part of the same object, because that sense that we see them as the same thing represents an inference that they're actually caused by a single object in the world. So the idea here is that there are these heuristics that are based on probabilities in the world.

And that's a perfectly useful way to think about this phenomena, but it doesn't tell us anything about how you would implement this with neurons. Do you have a question?

STUDENT: Yeah. So I saw a poster outside your office, something along the lines of testing these illusions with legally blind kids who just had surgery and gotten their vision back. So if this Helmholtz explanation actually is true, does this not work?

JOSH MCDERMOTT: Does what not work?

STUDENT: Would a person that just got their vision from a surgery see the right side, for example?

JOSH MCDERMOTT: Yeah. So the question is, would someone who's been visually impaired or blind from birth, who suddenly has their vision restored, would these effects hold for them? And I don't know. I mean, something that could be tested, but I think it really just depends on the extent to which all these factors are really something that result from evolution or result from development. And it could be some combination of both.

So one possibility is that your brain is wired up in a certain way in order to leverage these probabilistic relationships that exist in the world. So natural selection caused the brain to evolve to give rise to this kind of grouping. That's possible. And the prediction there would be that, well, from birth, or maybe largely independent of developmental experience, you would experience this more or less the same.

But the other possibility is that these are things that we learn online over the course of development, just by our interactions with the world, by looking out and then grabbing things and realizing what actually constitutes an object and stuff. And in general, we typically don't know a whole lot about the relative importance of development and evolution for a lot of things in perception. So a lot of those questions are still not very well answered.

And so in the case of grouping, I don't really know the answer to that. And another way to test that would be to ask, well, what would happen if somebody was reared in a world that had very different statistics? Would they end up having very different perception? That's another way you could, in principle, ask that question. Of course, that natural experiment doesn't typically occur. And there's no way to do that in people often. But sometimes you can do those experiments with animals in the lab. And you might in principle be able to do that. OK.

OK. So this is perceptual grouping, and there's these different levels of explanation. And there's lots of factors that affect grouping. So this is what's known as grouping by good continuation. So you see an image like this. And what you perceive is a circle and a square.

Now this is a perfectly valid description of how you would create that image. You got a Pac-Man and then this other funny shape that are just placed together in a certain way. But that's not what you see. So what you tend to see are things that are kind of continuous contours.

You also group by closure. So this image on the right could be described in a whole bunch of different ways. For instance, you could see it mostly as these closed shapes. You could see it as these kind of cross-like shapes. You can see it-- I don't what that-- what would you call that, a funny x-like thing? They're all valid descriptions of the image.

But people tend to see the circular organization. And again, then the naive idea is that, well, closed contours are pretty common in the world. Objects tend to be like that. And so your visual system kind of is wired up or programmed in some way to enable you to see that.

OK. I want to end there. When we come back, we will resume talking about grouping. Have a good weekend. And actually, it's spring break, right? So I hope everybody has a great spring break, that you get a chance to kick back a little bit. And I will see you in, like, 10 days. Bye.