[SQUEAKING]

[RUSTLING]

[CLICKING]

NANCY KANWISHER: All right, let's get started. So today, we're going to talk at some length about what I mean by this idea of Marr's computational theory level of analysis. It's a way of asking questions about mind and brain, and we're going to talk about that in the case of color vision.

And that's going to take a while. We'll go down and do the demo. We'll come back and talk about color vision and how we think about it at the level of computational theory and why that matters for mind and brain.

And then we're going to start, in the second half, a whole session, which is going to roll into next class, on the methods we can use in cognitive neuroscience to understand the human brain, and we'll illustrate those with a case of face perception. And we'll talk about computational theory, light very briefly, a face perception, what you can learn from behavioral studies, and what you can learn from functional MRI. And then we'll go on and do other methods next time. Everybody with the program? All right.

So to back up a little, the biggest theme addressed in this course, the big question we're trying to understand in this field is, how does the brain give rise to the mind? That's really what we're in it for. That's why there's lots of cognitive science. We're trying to understand how the mind emerges from this physical object.

And so for the last few lectures, you've been learning some stuff about the physical basis of the brain, what it actually looks like. Some of you guys got to touch it. I hope you thought that was half as awesome as I did. And we got a sense of the basic physicality of the brain and some of its major parts.

But now the agenda is, how are we going to explain how this physical object gives rise to something like the mind? And the first problem you encounter is, what is the mind anyway? I drew it as a weird, big, amorphous cloud because it's just not obvious how you think about minds, right? It feels like one of those things like, we could even have a science of the mind. What is mind? All kind of nervous making, right?

And so our field of cognitive science, over the last few decades, has come up with this framework for how we can think about minds, and this isn't even a theory. It's more meta than that. It's a framework for thinking about what a mind is, and the framework is the idea that the mind is a set of computations that extract representations.

OK, now that's pretty abstract. You can think of a representation in your mind as anything from a percept, like I see motion right now, or I see color. And as you learned before, you might see motion even if there isn't actually motion in the stimulus. But that representation of motion in your head, that percept, that's a kind of mental representation.

Or if you're thinking, why is Nancy going through this really basic stuff? She's insulting our intelligence. If something like that is going on in the background as I'm lecturing, that's a thought. That's a mental representation of a sort.

Or if you're thinking, oh my god, it's after 11:00, and I'm not going to get to eat until 12:30. I'm going to starve. Whatever thoughts are going through your head, those are mental representations, too, right? And so the question is, how do we think about those?

And so this idea that mental processes are computations and mental contents are representations implies that ideally, in the long run, if we really understood minds, we'd be able to write the code to do everything that minds do, right? And that code would work, in some sense, in the same way. Now, that's a tall order.

Mostly, we can't do that yet, like not even close, a few little cases in perception, kind of sort of maybe, but mostly, we can't do that yet. But that's the goal. That's the aspiration. And so the question is, how do we even get off the ground trying to launch this enterprise of coming up with an actual precise computational theory of what minds do?

And the first step to that is by thinking about what is computed and why, and so that is the crux of David Marr's big idea, the brief reading assignment that I gave you guys from Marr. And he's talking about, how do we think about minds and brains? Step number one, what is computed and why? So we're going to focus on that for a bit here.

And let's take vision, for example. You start with a world out there that sends light into your eyes. That's my icon of a retina, that blue thing in the back, the back of your eyes-- sends an image onto your eye, and then some magic happens. And then you know what you're looking at, OK?

So that's what we're trying to understand. What goes on in there? In a sense, what is the code that goes on in here that takes this as an input and delivers that as an output, OK?

More specifically, we can ask, as we did in the last couple of lectures-- let's take the case of visual motion. So suppose you're seeing a display like this, like something in front of you. Somebody jumps on a beach like that, and there's visual motion information.

What are the kinds of things-- so that's your input. What are the kinds of outputs you might get from that? Well, to understand that, we need to know what is computed and why.

So what is computed? Well, lots of things. You might see the presence of motion. You might see the presence of a person.

Actually, you can detect people just from their pattern of motion. We should have done this at the demo. Write me a note to think about that next time.

If we stuck a little tiny LEDs on each of my joints and we're in a totally black room and I jumped around and all you could see was those dots moving, you would see that it was a person. It would be trivially obvious. So motion can give you lots of information aside from "something's moving" and "what direction is it moving?"

You can see someone's jumping. That also comes from the information about motion. You can infer something about the health of this person or even their mood, so there's a huge range of kinds of information we glean from even a pretty simple stimulus attribute like motion.

And so if we're going to understand how do we perceive motion, we first need to get organized about, what's the input, and which of those outputs are we talking about? And probably, the code that goes on in between in your head or in a computer program, if you ever figured out how to do that, will be quite different for each of those things, but that's the way you need to be thinking about minds. OK, what are the inputs? What are the outputs?

And then as soon as you pose that challenge-- OK, let's say it's just moving dots, and you're trying to tell if that's a person. Think about, what is the code you'd write? Just these moving dots. How the hell are you going to go from that to detecting if those dots are on the joints of a person who's moving around versus on something else?

That's how you think, what are the computational challenges involved, OK? And I'm not going to ever ask you guys to write that code. We're just going to consider it as a thought enterprise to kind of see what the problem is that the brain is facing, that it's solving.

OK, and so Marr's big idea is this whole business of thinking about what is computed and why, what the inputs and outputs are, and what the computational challenges are getting from those inputs to those outputs, that all of that is a prerequisite for thinking about minds or brains, OK? So we can't understand what brains are doing until we first think about this. That's why I'm carrying on about this at some length.

And Marr writes so beautifully that I'm just going to read some of my favorite paragraphs because paraphrasing beautiful prose is a sin. So Marr says, "Trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers. It just can't be done.

To understand bird flight, you need to understand aerodynamics. Only then can one make sense of the structure of feathers and the shape of wings. Similarly, you can't reach an understanding of why neurons in the visual system behave the way they do just by studying their anatomy and physiology," OK? You have to understand the problem that's being solved, OK?

Further, he says, "The nature of the computations that underlie perception depends more on the computational problems that have to be solved than on the particular hardware in which their solutions are implemented." So he's basically saying we could have a theory of any aspect of perception that would be essentially the same theory whether you write it in code and put it in a computer or whether it's being implemented in a brain. Yeah.

AUDIENCE: Was Marr an engineer?

NANCY KANWISHER: Marr was many things. He was a visionary, a visionary who studied vision, a truly brilliant guy with a very strong engineering background. And this is now pervading the whole field of cognitive science, that people take an engineering approach to understanding minds and brains, to try to really understand how they work.

OK, so to better understand this, we're going to now consider the case of color vision. And so in this case, we start with color in the world that sends images onto the back of your retina, so magic happens. And we get a bunch of information out.

So the question we're going to consider is, what do we use color for, OK? And we're going to use the same strategy we used in the Edgerton Center. We're trying to understand some of the things that we use color for by experiencing perception without color, OK? What are the outputs?

NANCY KANWISHER: OK, so to do that, we're going to head over right now to the imaging center, and we're going to have a cool demo by Rosa Lafer-Sousa. So if it's going to be faster to leave your stuff here-- I don't know. Maybe we should-- yeah?

AUDIENCE: [INAUDIBLE]

NANCY KANWISHER: Yeah, we'll lock the room, OK? Yeah?

AUDIENCE: How long are we going to be there?

NANCY KANWISHER: 10 minutes, something like that, and I need everyone to boogie because there's a lot of stuff I want to get through today. So let's go. All right, so what do we use color for when we have it? It's not a trick question. It's supposed to be really obvious now. Yeah, what's your name?

AUDIENCE: Chardon.

NANCY KANWISHER: Chardon, hi.

AUDIENCE: Like choosing which food to eat.

NANCY KANWISHER: Yeah, yeah. Choosing which. What else related to that but different? Yeah.

AUDIENCE: Check-in procedure.

NANCY KANWISHER: Yeah, yeah, like what? What did you notice that you could identify better?

AUDIENCE: Different types of things.

NANCY KANWISHER: Mm-hmm, but besides identifying and choosing, what else?

AUDIENCE: More generally, bringing things into our awareness with the reds in particular with the strawberries.

NANCY KANWISHER: Yeah, do you find them easier to find?

AUDIENCE: No, much harder.

NANCY KANWISHER: Oh, yeah, right, harder without the light, right. Exactly. What else. Yeah.

AUDIENCE: Like driving. You need to have color to know the traffic lights.

NANCY KANWISHER: Totally, totally. That's a modern invention but a really important one. What else?

AUDIENCE: Are we general? Are we very general or like--

NANCY KANWISHER: Whatever. What do we use color for?

AUDIENCE: I mean, we used to figure out what to eat because one of the strawberries wasn't actually a strawberry. So yeah, I used color to [INAUDIBLE].

NANCY KANWISHER: Uh-huh, and the bananas. Did anybody notice? Sometimes, it's hard to tell. Yeah, boy in the back?

AUDIENCE: For assessing health risks.

NANCY KANWISHER: Say more.

AUDIENCE: If someone's face doesn't have color in it, you tend to assume that they're sickly.

NANCY KANWISHER: Totally. Did you feel like people's faces looked a little sickly? Absolutely, absolutely.

OK, so this is just to show you that a lot of computation theory starts with common sense of just reasoning, what do we use this stuff for? It helps to not have it to reveal what we use it for, but you guys have just reinvented the key insights in early field of color vision. OK, so standard story is to find fruit.

If you ask yourself how many berries are here, take a moment. Get a mental tally. How many berries?

OK, ready? Now how many berries, OK? You see more.

And in fact, there's a long literature showing that primates who have three cone colors-- we're not going to go through all the physiological basis of cones and stuff like that, but they have a richer color vision because the number of different color receptors in their retina-- they're better at finding berries.

And in fact, a paper came out a couple of years ago where they studied wild macaques on an island off of Puerto Rico called Cayo Santiago, and the macaques there have a natural variation genetically where some of them have two color photoreceptors instead of three, OK? And in fact, they followed them around, and the monkeys that have three photoreceptor types are better at finding fruit than the ones that have only two, OK? So that story that's just been a story for a long time turns out it's true.

And also, as you guys have already said, to not just find things but identify properties-- you can probably tell whether you'd want to eat those bananas on the bottom. Maybe not, but it's hard to tell on the top which ones you like. And yet that's all you need to know, OK? So these are just a few of the ways that we use color and why it's important.

But there is a very big problem now that we try to figure out, OK, what is the code that goes between the wavelength of light hitting your retina and trying to figure out, what color is that thing? So here's the problem. We want to determine a property of the object, of its surface properties, its color, right? That's a material property of that thing.

But all we have-- so here's a thing. We'll call that reflectance. It's a property of wavelength, but you can think of it as, for now, just a single number.

It's a property of that surface, but all we have is the light coming from that object to our eyes. That's called luminance. I'm not going to test you on these particular words, but you should get the idea, OK? So that's what we have. That's our input.

But here's the problem. The light that's coming off the object is a function not just of the object but of the nature of the light that's shining on the object. That's called the illuminant. So the problem is we have this equation. This light coming from the object to our eyes is the product of the properties of the surface and the incident light, OK?

And our problem is, we have to solve for L. I'm sorry. We have to solve for R, the property of the object. Given L, what is R? That's a problem, OK?

That's kind of like if I said, A times B is 48. Please solve for A and B, OK? That's known in the field as an ill-posed or underdetermined problem. We don't have enough information to uniquely solve this, OK?

That's a very, very deep problem in perception and a lot of cognition. We are often, in fact, most of the time in this boat, OK? So the implications are, when we want to infer reflectance of the property of the object from L, we must bring in other information, right? We must have some way to make guesses about I, about the light shining on that object, OK?

So the big point is many, many inferences in perception and cognition are ill posed in exactly this way, all right? And so here are two other examples of ill-posed problems in perception. In shape perception, you have a similar situation. You have stuff in the world that's making an image on the back of your eyes, OK? That's optics.

What we're trying to do as perceivers is reason backwards from that image. What object in the world caused that image on my retina? That's sometimes called inverse optics because you're trying to reason the opposite way. That's basically what we're doing in vision.

So here's a problem. It's a crappy diagram, but if you can see here, there's three very different surface shapes here that are all casting the same image, for example, on a retina. You could do this with cardboard and cast it with a shadow. So everybody get what this shows here?

What that means is if you start with this and you have to reason backwards to the shape that caused it, that's an ill posed problem, big time. It could be any of those things. This information doesn't constrain it. Does everybody see that problem? OK, so that's another ill-posed problem.

Here's a totally different example of an ill-posed problem that's big in cognition. When you learn the meaning of a word, especially as an infant trying to learn language, the classic example the philosophers like-- God knows why. Philosophers like weird stuff, but never mind. Somebody points to that and says "gavagai," and your job is to figure out, what does "gavagai" mean, OK?

So "gavagai" could mean all kinds of different things. It could just mean rabbit if you already have a concept of a rabbit. It could mean fur. It could mean ears. It could mean motion if the rabbit is jumping around, or in the example of the philosophers love, it could mean undetached rabbit parts. Weird, but anyway, philosophers like that kind of thing.

Anyway, the point is, it's ill posed. We don't know from this what is the correct meaning of the word. Does everybody see how this underdetermines the correct meaning of the word? We don't have enough information to solve it, OK?

So yeah-- blah, blah. So there's a whole literature on the extra assumptions that infants bring to bear to constrain that problem, so they can make a damn good guess about what the actual meaning of the word is, OK? It's a whole big literature, quite fascinating. OK, but for now, I just want you to understand what an ill-posed problem is and why it's central to understanding perception and cognition.

OK, so back to the case of color. As I said, the big point is that lots of inferences, including determining the reflectance of an object, are ill posed, and so we have to bring in assumptions and knowledge from other places, from our knowledge of the statistics and the physics of the world, our knowledge of particular objects. All kinds of other things must be brought to bear, OK?

So all of that, again, is considering the problem of color vision at the level of Marr's computational theory. Notice we haven't made any measurements yet. We've just thought about light and optics and what the problem is and what we use it for, OK? All this stuff.

What is extracted and why? R, the reflectance of an object, useful for characterizing objects and finding them. What cues are available? Only L, and that's a problem because it's ill posed.

OK, next question-- so obviously, we get around, and we can figure out what colors are which. What are the other sources of information that we might use in principle and that humans do use in practice, OK? And so all of that kind of stuff has been done without making any measurements. We're just thinking about the problem itself, OK?

All right, so next, Marr's other levels of analysis, algorithm and representation in hardware, are more standard ones you will have encountered, which is why I'm making a big deal of computational theory. It's really his major novel contribution, but it's better understood by contrast with these.

So at the level of algorithm and representation, this is like, what is the code that you would write to solve that problem, right? And so we could ask, how does the system do what it does? Can we write the code to do it, and what assumptions and computations and representations would be entailed?

So how would we find out how humans do this? Well, one of the ways is a slightly more organized version of what you guys just did, and that's called psychophysics. Psychophysics just means showing people stuff and asking them what they see or playing them sounds and asking them what they hear. And you can do it in very sophisticated, formalized ways, or you can do it like we just did. Talk to us about what the world looks like, OK? Usually, psychophysics means a slightly more organized version.

OK, so here's an example. In fact, it's a cool demo, also from Rosa. And so what I'm going to do is I'm going to show you a bunch of pictures of cars, and your task is going to be to shout out loud as fast as you can the color of the car, OK? They're going to appear on the screen. Everyone ready? As fast as you can, shout it out loud. Here we go. What color?

AUDIENCE: Red.

AUDIENCE: Green.

AUDIENCE: Red.

AUDIENCE: Red.

AUDIENCE: Red.

AUDIENCE: It's green.

NANCY KANWISHER: Interesting. OK, here's another one.

AUDIENCE: Blue.

AUDIENCE: Yellow.

AUDIENCE: Yellow.

NANCY KANWISHER: Uh-huh, interesting. Ready? Here we go. Here's another one.

AUDIENCE: Green.

NANCY KANWISHER: OK, here's another one.

AUDIENCE: Blue.

NANCY KANWISHER: Ah, you guys cottoned on to that pretty fast. OK, so good job. Nice consensus, although I noticed a little bit of transition there, which is very interesting.

But here's the thing. All of those cars are the exact same color. The body of the car is the exact same in all of them, and if you don't believe it, I'm going to include everything except for a patch, OK? Here we go.

Boom, they're all gray. I know. It's awesome. It's Rosa that's awesome, not me. I just had to bum this because it's so awesome.

OK, so Rosa spent months designing these stimuli to test particular ideas about vision, but the basic demo is simple and straightforward. You can get the point here, OK? So what's going on here?

What's going on here is that you guys, the algorithm running in your head that's trying to figure out what is the color of that car, is trying to solve the ill-posed problem, and it's using other information than just the luminance of light coming from the object. It's using information from the rest of the object. It's making inferences about L, the luminance, the light hitting the object, OK? And in particular, when you look at that picture up there, what is the color of light shining on that car?

AUDIENCE: Green.

NANCY KANWISHER: Yeah, right, officially known as teal in the field, but some of you shouted out "green" first because that's what you saw first, is the color of light, OK? What's the color of light hitting that car? Yeah, purple, magenta. Here?

AUDIENCE: Blue.

NANCY KANWISHER: Yeah, and over there?

AUDIENCE: Yellow.

NANCY KANWISHER: Yeah, yellow-orange. Yeah, OK, so basically what your visual system did is look quickly, figure out the color of the incident light, L, and use that to solve the otherwise ill-posed problem of solving for R, the color of the car. And in this case, this demo shows that if you just change the color of the illuminant light and hold constant the actual wavelengths coming from that patch, you can radically change the perceived color of the car. Everyone got that? OK.

AUDIENCE: So wait, wait, wait--

NANCY KANWISHER: Yeah.

AUDIENCE: If you ran this through a computer, asked it to get the intensity of the pixel on the hood of the car there, it would not correspond to yellow, it would correspond to green?

NANCY KANWISHER: Well, it depends what you're asking the computer exactly. If you hold up a spectrophotometer that's just going to measure the wavelength of light, they're all gray. Right there on top of those cars, they're all the exact same neutral gray. That's just the raw physical light coming from that patch, OK? But if you coded up the computer to do something smart and you coded it up to take other cues from the image, try to figure out what L is, and therefore solve for R, you might be able to get it to do the right thing.

AUDIENCE: Yeah, I mean just like you looked at the pixels, like in the matrix, I mean, the color on the car, would it be what yellow is, or would it be more grays?

NANCY KANWISHER: Gray. They're all gray. So that's what I was trying to show you here is that, in fact, they are actually gray, right?

The cars are underneath there, and you can see they're all exactly the same. And they're gray, and there's no color in it, OK? Everyone got that?

All right, OK, so all of that is a little baby example of psychophysics, what we do at the level of trying to understand the algorithms and representations extracted by the mind to try to figure out, what are the strategies that we use to solve problems about the visual world?

OK, and so behavior or psychophysics or seeing, as you just did, can reveal those assumptions and reveal some of the tricks that we're using in the human visual system to solve those ill-posed problems, OK? So in this case, it was assumptions about the illuminant that enabled us to infer the reflectance from the luminance.

OK, the third level Marr talks about is the level of hardware implementation. In the case of brains, that's neurons and brains, and so we won't cover this in any detail here. But there's lots and lots of work on the brain basis of color vision. We'll mention it briefly next time.

So this is some of Rosa's work showing those little blue patches on the side of the monkey brain that are involved in color vision, and some work that Rosa did in my lab showing the bottom surface of the human brain with a very similar organization with those little blue patches in there that are particularly sensitive to color. So you can study brain regions that do that. If it's a monkey, you can stick electrodes in there and record from individual neurons and see what they code for, and you can really tackle at multiple levels the hardware neural basis of color vision in brains as well.

OK, so the big general point is we need lots of levels of analysis to understand a problem like color vision, OK? And so accordingly, we need lots of methods to understand those things, all right? So what I want to do next is now launch into this whole thing about the different methods that we can use in the field, and this part of the lecture will go on to next time. But let's get going. Everybody good with this so far?

All right, so we're going to use the case of face perception to think about the different kinds of questions and different levels of analysis in face perception. So let me start by saying, why face perception? Not just that I've worked on it for 20 years, although I'll admit that's relevant. There's lots of other good reasons beyond that, why we should care about face perception.

So I don't have a demo that enables me to kind of put you in a situation where you can see everything but faces. That would be cool and informative if we could do that, but failing that, I can tell you about somebody who's in that situation. And this is a guy named Jacob Hodes.

So this is a picture of him recently. I met him around a decade ago when he was a freshman at Swarthmore. And he sent me an email, and he said, I've just learned about face perception and the phenomenon of prosopagnosia, the fact that some people have a specific deficit in face recognition. And it explains everything in my life, and I want to meet you.

And I said-- because he knew I worked on face perception. I said, that's awesome. I would love to meet you, but I got to tell you, I'm not going to be able to help. So if you're interested in chatting, please, please come by, but I don't want you to feel like I'm going to be able to do anything useful. He said, no, I don't care. I want to understand the science.

So he comes by, and by the way, one of the things that people have wondered for a while is, are people who have particular problems with face recognition-- are they just socially weird? Are they just bizarre, maybe a little bit on the spectrum? They don't pay attention to faces, and so they don't get them very well and so forth. Or can they be totally normal in every other respect except for just face perception?

And so I was very interested. I'd only emailed with this guy, and when he showed up in my office, within about 15 seconds, it's like, this is like the nicest, normalest kid you could ever meet. Such a nice guy, so normal, socially adept, smart, thoughtful-- lovely, lovely person.

So I chatted with him for a long time, and he told me he was then halfway through-- he grew up in Lynn, Massachusetts, and he went off to Swarthmore his freshman year. And he had been having a really rough time of it because in his hometown, he was with the same group of kids all the way from first grade through high school. And so in fact, he just can't recognize faces at all, never could.

When he was a little kid his mom used to drive him to the practice field, and they would sit there and come up with cues about, this is how you tell that's Johnny. He's got this weird thing about his hair, and this is how you tell that's Bobby. And they would practice and practice. And so he developed these clues to be able to figure out who was who in his small little cohort of kids that he knew all the way through high school.

Then he goes off to college, and it's all these new people. And he's screwed. And he said to me that he was just devastated because he would go to a party, and he would meet someone and think, wow, this is a really nice person. I would really like to be this person's friend, but he would realize he would have no way to find that person again.

And he's like, it's kind of oversharing to say when you've met somebody for 10 minutes, by the way, I'm not going be able to find you. You have to find me. It's like, you just don't want to have to go there yet, right?

So there's all kinds of things that would make it a real drag to not be able to recognize other faces, and now having said all of that, I'll say that a surprisingly large of the population is in Jacob's situation, about 2% of the population. It would be unsurprising if there were one or two of you in here, and if there is, you can tell me later. I'd love to scan you.

But about 2% of the population routinely fails to recognize family members, people they know really well, right? And interestingly, this is completely uncorrelated with IQ or with any other perceptual ability or ability to read or recognize scenes or anything else. Yeah.

AUDIENCE: Is this the kind of thing where you either have it or don't have it?

NANCY KANWISHER: Oh, good question. No, it's a gradation. So the 2% at the bottom are not this 2% who are really screwed, and everyone else is up here. It's a hugely wide distribution, and the point is that the bottom end of that distribution is really, really bad. They just can't do it at all.

Similarly, the top end of that distribution is weirdly good. They are so good at face recognition that they have to hide it socially because otherwise people feel creeped out. For example, as one of those people-- they're called super recognizers. A bunch of them have been hired by investigation services in London recently as part of their kind of crime-solving unit. Those people are so good that one of them said to me-- we scanned a few of these people.

One of them said, if I-- she recounted this event where she's standing in line waiting for movie tickets, and she realizes that the person in front of her in line was sitting at the next table over at a cafe four years before. She says, if I share this information with that person, they'll be creeped out, so I've just learned to keep it to myself. But I know that was the same person, right? So there's a huge spread. You had a question a while back.

AUDIENCE: It's not a problem with recognizing facial images, is it? So for example, Jacob, looking at a person, could he describe them?

NANCY KANWISHER: Absolutely. He knows that it's a face. He can tell if they're male or female. He can tell if they're happy or sad. It looks like a face to him. It just doesn't look different than anyone else. Yeah.

AUDIENCE: Is there any difference-- OK, for example, my father-- he can tell faces in person just fine, but when he watches videos of people, he just cannot. He cannot recognize faces at all, so is there any difference?

NANCY KANWISHER: There are lots of cues. I mean, that's a very interesting exercise to think about. What are the cues that you have in person, right? You have all kinds of other things.

So first of all, there's lots of constraining information. The person you're looking at-- there are all kinds of things you know about where you are and who that might be that help, right? So yeah, there's many different cues to face recognition that might be engaged here.

So my point is just that face recognition matters. You can get by if you can't do it, but it sucks. It's really hard, OK? OK, so yes, question?

AUDIENCE: Do you have any idea what people at [INAUDIBLE] university? Are there any sort of models that suggest that when there are-- like, when they see faces, do they just see a sort of conglomerate of shapes?

NANCY KANWISHER: No, they see the structure of the face. They see a proper face. If the eye was in the wrong place, they would know. They absolutely know the structure of the face. They all look kind of the same.

By the way, we won't have time to talk about this in any detail, but there's a well-known fact that probably many of you guys have experienced, which is called the other-race effect. And that is the fact that they all look the same. Whoever they are, if you have less experience looking at that group of people, you're less well able to tell them apart, OK?

I have this problem teaching all the time. I grew up in a rural lily-white community. My face recognition is not so good to begin with, and it's really not good for non-Caucasian faces.

It's embarrassing as hell. It feels disrespectful. I hate it. I fault myself, but actually, it's just a fact of the perceptual system. Your perceptual system is tuned to the statistics of its input, and it's not so plastic later in life.

And so a way to simulate a version that some of you may have experienced is whatever race of faces you have less experience with, if you find those people hard to distinguish, it's not that you can't tell it's a face. It's not that you wouldn't be able to tell if the nose was in the wrong place. It's just hard to tell one person from another, so it's a lot like that. I really need to get going, so I'll take one more question and go.

AUDIENCE: Wait, could you kind of use an analogy? It's like being able to tell people apart by their hands or something to the point that you just-- you can't really tell people apart by their hands, usually, so is that kind of how people with face blindness feel? It's just looking at [INAUDIBLE].

NANCY KANWISHER: That's all you had, yeah. Yeah, probably, probably. Yeah, and there is, by the way, an interesting literature. You show people photographs of their own hand and a bunch of other hands. People can't pick out their own hand from-- so yeah, you're right. We're not so good at that.

OK, I'm going to go ahead. If you guys are interested, I could post-- there's a whole fascinating literature here, but actually, I got dinged last year for talking about face recognition too much and prosopagnosia. We all heard about it in 900. So I took most of that out, and now you guys are asking me. So I don't know what the right thing is. But I'm going to go on, and I will put some optional readings online, especially if you send me an email and tell me to do that.

OK, so point is, faces matter a lot. They matter for the quality of life. They're important because they convey a huge amount of information, not just the identity of the person but also their age, sex, mood, race, direction of attention.

So if I'm lecturing like this right now and I start doing that, you guys are going to wonder, what the hell's going on over there? Yeah, I saw a few heads turn. I'm just doing a little demo here, right?

We're very attuned to where other people are looking, OK? So this is just one of many different social cues we get from faces. There's just an incredibly rich bunch of information in a face.

We read in aspects of people's personality from the shape of their face even though it's been shown with some interesting recent studies there's absolutely nothing you can infer about a person's personality from the shape of their face. We all do it, and we do it in systematic ways.

Another reason this is important-- and faces are some of the most common stimuli that we see in daily life, starting from infancy where, I think, about 40% of waking time, there's a face right in front of an infant's eyes. And probably, these abilities to extract all this information have been important throughout our primate ancestry. So that's just to say, there's a big space of face perception, and now we're going to focus in on just face recognition, telling who that person is, all right?

So what questions do we want to answer about face recognition? Well, a whole bunch of them, and what methods do we want to use? So let's start with some basic questions about face recognition.

Well, first, as usual, we want to know, what is the structure of the problem in face recognition? What are the inputs? What are the outputs? Why is it hard, right? Just as we've been doing for motion and color-- that's Marr's computational theory level.

We want to know, how does face recognition actually work in humans? What computations go on? What representations are extracted?

And is that answer different? Are we running different code in our heads when we recognize faces from when we recognize toasters and apples and dogs, OK? Another facet of that-- do we have a totally different system for face recognition from the recognition of all those other things? If so, then we might want different theories of how face recognition works from our theories of how object recognition works.

How quickly do we detect and recognize faces? That will help constrain what kinds of computations might be going on. And of course, how was face recognition actually implemented in neurons in brains? So this is just some of the big, wide-open questions we want to answer.

So let's consider, what are our tools for considering these things? And you guys should all know what tools are available for thinking at the level of Marr's computational theory-- basically just thinking, right? You can collect some images, too, but basically to understand this, we just think.

So for example, as I keep saying, at the level of Marr's computational theory, we want to know, what is the problem to be solved? What is the input? What is the output? How might you go from that input to that output, OK?

So for example, here's a stimulus that might hit a retina, and then some magic happens. And then you just say, Julia, OK? So we want to know, what's going on in that magic, OK?

And if a different image hits your retina, you go, oh, Brad. That is, I wouldn't. I live in a cave. I barely get out of the lab, but I understand that these are people most people recognize. That's why I use them.

That's the question. What goes on here in the middle? And your first thought is, well, duh, easy. We could just make a template, a kind of store the pixels that match that image and take the incoming image and see if it exactly matches.

And that's going to work great, right? No. Why not?

AUDIENCE: Different angles [INAUDIBLE].

NANCY KANWISHER: Louder.

AUDIENCE: Angles [INAUDIBLE].

NANCY KANWISHER: Yeah, yeah, absolutely. That's not going to work at all. And the problem is that we don't just have one picture of Julia that we can match. There are loads of loads of totally different kinds of pictures of Julia, all of which we look at and immediately go, Julia, no problem, OK?

And so that means, what is it that we're doing in our heads? If we're storing templates, we have to store a lot of them, OK? So all those differences in the images-- so we could memorize lots of templates.

Well, that has long been taken as like the reductio ad absurdum. That's the ridiculous hypothesis. How could that be? How could there be room in here to store lots of templates of each person, and furthermore, how would that work for people we don't know?

The other idea, which is very vague right now, is that, well, maybe we extract something that's common across all of those, maybe something like the distance between the eyes, something about the shape of the mouth, other kinds of properties that might be invariant across those images, that is, that you could pull out that information from any of those images, OK? It's sounding very vague because it is vague. Nobody knows what those would be, but the idea is, maybe there's some image invariant properties of a face you can get from here that you can then store and use to recognize faces, OK?

So now to think about this, we can step back and say, OK, how is this done in machines? So machine face recognition didn't work well at all until very recently, OK? And then all of a sudden, a couple of years ago-- here's another paper from the different one that I showed you before. This one is VGGFace, one of the major deep net systems for face recognition.

It's widely used. There was another one the year before. All of this since 2014, 2015-- hugely cited, widely influential. They're on all your smartphones. Boom, it all just happened like nearly overnight with the availability of lots of images to train deep nets.

So now these things are extremely effective and accurate, and so in some sense, those networks are possible models of what we're doing in our heads when we recognize faces. It doesn't mean we do it in the same way, but it's a possibility. It's a hypothesis we could test, OK? Yeah.

AUDIENCE: What is the current state of the literature surrounding getting other information from people's faces like moods or what they're--

NANCY KANWISHER: Lots, lots. There's conferences and machine vision competitions on extracting personality properties, mood properties, every possible thing you could imagine. This is a huge-- a lot of people care about this. It's a huge field in computer vision, and it's also a huge field in cognitive science, asking what humans pull from faces.

AUDIENCE: How much success that you've seen?

NANCY KANWISHER: Oh god, others would know that better than me. I bet it's pretty damn good, a lot of it. Yeah, yeah, I mean, these things are suddenly extremely effective. Yeah, OK, and there will be, by the way, later in the course-- my postdoc, Katharina Dobs, who knows that literature much better than I do, will talk about deep nets and their application in human cognitive neuroscience. And she knows a lot about the various networks that process face information.

OK, so this is progress. Now we have some kind of computational model. Trouble is, nobody really has an intuitive understanding of what VGGFace is actually doing. You know how to train one up. There it is, but we don't really understand what it's doing.

And further, we have no idea if what it's doing is anything like what humans are doing, OK? So it's progress that we have a model now that we didn't have like five years ago, but we still have all these questions open. OK, so on this first question, what do we want to know? What we discover at the level of Marr computational theory is a, if not the, central challenge in face recognition is the huge variation across images, which you know just by thinking about it or trying to write the code.

OK, so ooh, I'm just barely able. I'm going to race along, and Anya's going to tell me in five minutes to switch. OK, so I want to talk just a little bit about behavioral data. I'll run out of time, and we'll roll this in less time, because I want to include functional MRI because you guys need it for the assignment.

OK, so how are we going to figure out what humans represent about faces? OK, so here we are. We consider this possibility that one way to solve this problem is by essentially memorizing lots of templates for each person. Another possibility is this kind of vague, inchoate idea that maybe there's some abstract representation that'll be the same across all of those.

How are we going to figure out which humans do? Well, if we're really memorizing lots of templates for each person and that's how we recognize them in all their different guises, that wouldn't work for people we didn't know. That is, you wouldn't be able to take two different photographs of the same person and know if it's the same person or not, right? Because you could only do this by memorizing.

Does everybody get that idea? Whereas whatever this other idea is, it should work somewhat for novel individuals you don't already know. Here are two photographs. Same person or different person?

So now let's ask, can humans do this? Do we store lots of templates for individuals, or can we do something more abstract? Well, if we simply deal with this problem by storing lots of templates for each individual, maybe not literally pixel templates but some kind of snapshot, then the key test is, we shouldn't be able to do this matching task if we don't know that person. Everybody get the logic here?

OK, so let's try it. So this paper a few years ago-- Jenkins et al. Asked that question. So here's what they did.

They collected a whole bunch of photographs of Dutch politicians with multiple images of each politician, OK? Then they gave them to people on cards, and they said, there are multiple images of each person. And I'm not going to tell you how many different politicians are in this deck. Just sort them in piles, so there's a different pile for each person, OK?

I'm going to show you a low-tech version of this. I'm going to show you a whole bunch of pictures, all in one array, and you guys are going to try to figure out how many people are there, OK? Everybody ready?

I'm just going to leave it up for a few seconds. There's going to be lots of pictures. Your task is how many different individuals are depicted here. Here we go.

OK, write down your best guess. Just kind of look around. OK, everybody got a guess? OK, write down your guess.

OK, how many people think there were over 10 different individuals there? One. OK, how many people think over five? Yeah, probably half of you.

How many people think over three? Most of you. There are two.

What does that mean? That means you can't do it. That means you can't match different images of the same person if you don't know that person. Pretty surprising, isn't it?

We think we're so awesome at face recognition because most of the time, what we're doing is recognizing people we know, people we've seen in all different viewpoints and hair arrangements and stuff. If you don't have lots of opportunity to store all those things and it's a novel face, we're really bad at that, OK? Yeah.

AUDIENCE: But there's a constraint of time?

NANCY
KANWISHER: Yeah, yeah, I was trying to make the demo work. But OK, so the way they do this task, people have unlimited time, and they're just kind of sorting them. The mean number of piles that people made in this experiment was 7 and 1/2. Correct answer's two, OK?

OK, now you might say, well, maybe those are shitty photographs, right? OK, so here's the control. Those are Dutch politicians. They then did the same experiment on Dutch people who look at that photograph and in about two seconds say two, duh, OK? So if you know there's nothing wrong with those photographs, it's just a matter of whether you know those people or not.

OK, so the point of all of this is that this crazy story that, in fact, what a lot of what we're doing-- I'm sort of simplifying here, but a lot of what we're doing in face recognition, a lot of the way we deal with all this image variability is not that we have some very abstract, fancy, high-level representation of each individual face. We just have lots of experience with faces, and we use that so that if we have a novel face that we don't have all that experience with, we're not so good at it. I'm going to run out of time, so I'll take one question and go on.

AUDIENCE: How do they control for-- you know the issue you said about if you don't have experience with similar races.

NANCY KANWISHER: Yeah. I'm sure whenever you do face recognition experiments, you make sure that if your dominant subject pool is Caucasian, you have Caucasian faces or whatever. Yeah. Unless it's something you don't understand-- I'm going to hang around after class. You can ask me questions there, or if you have to go, you can email me because I really want to get through this next bit, OK?

OK, so there we are with that. So what this suggests, kind of, sort of, is that whatever we're doing, it's something that benefits enormously from lots and lots of experience with that individual. Maybe it's not literal memorization of actual pixel-like snapshots, but it's something more like that than anybody would have guessed before this experiment, OK?

OK, all right, I'm going to skip this awesome stuff here. OK. OK, so the benefits of-- actually, I'm going to come back and do that slide next time, too.

And we're going to cut straight to functional MRI. I'm sorry about this, but I just really want you guys to have this background in case you don't. You probably do.

So functional MRI-- another cool method in cognitive neuroscience, and how would it be useful here? OK, so first, what is it? Functional MRI is the same as regular MRI that's in probably tens of thousands of hospitals around the world.

The big advances in functional MRI were when some physicists in the early '90s figured out how to take those images really fast and how to make images that reflect not just the density of tissue but the activity of neurons at each point in the brain, OK? That was big stuff, OK? Early 1990s.

And so the reason it's a big deal it is the best, highest spatial resolution method for making pictures of human brain function noninvasively. That means without opening up the head, all right? So that's an important thing. That's why there's lots and lots of papers on it. That's why we're going to spend a lot of time on it.

The bare basics are that the functional MRI signal that's used is called the BOLD signal. That stands for blood oxygenation level dependent signal, OK? And what that means is this.

Basic signal is blood flow. And so the way it works is if a bunch of neurons someplace in your brain start firing a lot, that's metabolically expensive to make all those neurons fire, and so you have to send more blood to that part of the brain. So it's just like if you go for a run, the muscles in your legs need more blood delivered to them to supply them metabolically for that increased activity, and so the blood flow increase to your leg muscles will increase, OK? Well, similarly, the blood flow increases to active parts of the brain.

Now, the weird part of it is that for reasons nobody completely understands, the blood flow increase more than compensates for the oxygen use, so the signal is actually backwards. Active parts of the brain have less, not more, deoxygenated hemoglobin compared to oxygenated hemoglobin, and the relevance of that is that oxygenated hemoglobin and deoxygenated hemoglobin are magnetically different in the way that the MRI signal can see. So the basic signal you're looking at is, how much oxygen is there in the blood in that part of the brain, and hence, how much blood flow went there? And hence, how much neural activity was there?

Did that sort of make sense? I'm not going to test you on which is paramagnetic and which is diamagnetic. I never remember. I couldn't care less, but you should know what the basic signal is, right? It's a magnetic difference that results from oxygenation differences that result from blood flow differences that result from neural activity. Yeah.

AUDIENCE: So when there's more blood flow, there's more oxygenated?

NANCY KANWISHER: More oxygenated, and because it overcompensates for the metabolic use of the neurons, the active parts that you see with an MRI signal have more oxygenated hemoglobin, right? OK, all right, so that's the basic signal. And because that's the basic signal, there's a bunch of things we can tell already.

So first of all-- I'm just going to-- am I going to do this? Yeah, I'm going to skip over this. It doesn't really matter.

Because it's all based on blood flow, one, it's extremely indirect-- neural activity, blood flow change, overcompensation, different magnetic response, MRI image, right? So you would think with all those different steps that you would get a really weird, nonlinear, messy, crappy signal out the other end. And it is one of the major challenges of my personal atheism, but actually, you get a damn good signal out the other end. And it's pretty linear with neural activity, which seems like kind of a freaking miracle, given how indirect it is, OK? But that has empowered this whole huge field to discover cool things about the organization of the brain.

OK, nonetheless, there are many caveats. Because it's blood flow, the signal is limited in spatial resolution down to-- people fight about this, but around a millimeter. There are cowboys in the field who think that they can get less than a millimeter.

Maybe. I don't know. It's debated.

And the temporal resolution is terrible. Blood flow changes take a long time. Think about it.

You start running. How long does it take before the blood flow increases to your calves? Well, if you're really fit, it's probably fast, but still going to take a few seconds.

It takes about six seconds for those blood flow changes in the brain after neural activity. And it happens over a big sloppy chunk of time, and so you don't have much temporal resolution with functional MRI. Does that make sense? OK.

OK, because it's this very indirect signal, that also means that when we get a change in the MRI signal, we don't exactly know what's causing it. Is it synaptic activity? Is it actual neural firing?

Is it one cell inhibiting another? Is it a cell making protein? I mean, it could be any of these things, right? So we don't know, and that's a problem.

And another problem is the number you get out is just the intensity of the detection of deoxyhemoglobin. It doesn't translate directly into an absolute amount of neural activity. The consequence of that is all you can do is compare two conditions.

You can never say, there was this exact amount of metabolic activity right there. You can only say it was more in this condition than that condition, OK? All right, so those are the major caveats. Nonetheless, we can discover some cool stuff.

OK, so let's suppose, to get back to face recognition, you wanted to know, is face recognition a different problem in the brain from object recognition, right? If it was, you might want to write different code to try to understand it from the code you're writing for object recognition. It's something you'd kind of want to know, OK?

So here's an experiment I did-- god-- 20 years ago. Anyway, simplest possible thing-- so it's the easiest way I can explain to you the bare bones of a simple MRI experiment. You pop the subject in the scanner. You scan their head continuously for about five minutes while they look at a bunch of faces.

For 20 seconds, they stare at a dot. They look at a bunch of objects. They stare at a dot, OK? It's a five-minute experiment. You're scanning them that whole time.

And then you ask, of each three-dimensional pixel, or voxel, in their brain, whether the signal was higher in that voxel while the subject was looking at faces than while they were looking at objects, OK? And when you do that, you get a blob. I've outlined it in green here, but there's a little blob there. This is a slice through the brain like this. That blob is right in here on the bottom of the brain.

And the statistics are telling us that the MRI signal is higher during the face epochs than the object epochs-- everybody with me here-- which implies very indirectly that the neural activity of that region was higher when this person was looking at faces than when they were looking at objects. OK, now whenever you see a blob like that, really, you want to see the data that went into it, so here's mine. This is now the raw average MRI signal intensity in that bit of brain over the five minutes of the scan.

You can see the signal's higher in that region when the person is looking at faces-- these bars here-- than when they're looking at objects there. Everyone get that? That's what the stats are telling us. This is just the reality check of the data that produced those stats.

OK, so now, in fact, you can see something like that in pretty much every normal person. I could pop any of you in the scanner, and in 10 minutes, we'd find yours, OK? Now, here's the key question.

Does this so far-- let's suppose you find this in anyone. You do all the stats you like. It's as robust as you could possibly want.

Do these data alone tell us that that region is specifically responsive to faces? No. Why not?

AUDIENCE: Because it could-- just like that certain arrangement of features, or it could be reacting to the variable light intensities that's reflecting off people's skin-- coloration of the skin.

NANCY KANWISHER: Good. Keep going. What else? Yes, you.

AUDIENCE: Different faces of anything, like even animals.

NANCY KANWISHER: Yeah, then it might still be faces, but it would be different if it's human faces versus any faces. We kind of want to know, right? The code would be different. Yeah.

AUDIENCE: It might be because of the face is part of the bigger whole and the object's a [INAUDIBLE].

NANCY KANWISHER: Uh-huh, the face is a part of something, absolutely, where the object is the whole thing. What else? Yeah.

AUDIENCE: Maybe the orientation of the object is so much simpler than the human face is.

NANCY KANWISHER: Just objects are simpler or maybe just easier. Maybe it's just hard to distinguish one face from another, and so you need more blood flow. Really, what that thing is-- that's a generic object recognition system, but it has a harder time distinguishing faces from each other because they're so similar. So there's more activity. Everybody get that?

OK, what else? I'm going to go two minutes over, so if people have to leave, that's OK. I'll try not to go more than two minutes over. What else? Yeah.

AUDIENCE: It could just as easily be responsible for seeing maybe functional cues rather than face recognition.

NANCY KANWISHER: Yeah, yeah. As I just said, there's all this stuff we get from a face, not just who is it, but are they healthy? What mood are they in? Where are they looking? All that stuff.

OK, so what you guys just did-- this is just basic common sense, but it's also the essence of scientific reasoning. And we'll do a lot of that in this class. And the crux of the matter is, here's some data. Here's an inference.

And so your job is to think, is there any way that inference might not follow from those data? How else might we account for those data, OK? And you guys just did that beautifully, OK?

So the essence of good science is whenever you see some data and an inference, ask yourself, how might that inference be wrong? How else might we account for those data? OK, so that's what you guys just did. I had previously made a list of other things that might mean. It could respond to anything human. You had said any kind of face, but it could also be just anything human, may be a response to hands, any body part, anything we pay more attention to, anything that has curves in it, or any of the suggestions you guys made, OK?

So the crux of the matter in how you do a good functional MRI experiment or make a strong claim about a part of the brain based on functional MRI is to take all these alternative accounts seriously. And so as just one example, what we did in our very first paper is say, OK, there's lots of alternative accounts. Let's try to tackle a bunch of them.

So we scan people looking at now 3/4 views of faces and hands, and we made them press a button whenever two consecutive hands were the same-- that's called a one-back task-- or whenever two consecutive faces are the same. By design, that task is harder on the hands than the faces, so we were forcing our subjects to pay more attention to the hands than the faces, OK? And what we found is you get the same blob, still responding more to faces than hands.

And so the idea is by showing that, we've ruled out every one of those things. It's not just any human body part. It doesn't go to hands-- oh, sorry, anything human.

It's not just any body part. It's not anything we paid attention to because we made them pay more attention to the hands. It's not anything with curvy outline, OK?

And so that's just a little tiny example of how you can proceed in a systematic way to try to analyze what is actually driving this region of the brain. You come up with a hypothesis, and then you think of alternatives to the data. And you come up with more hypotheses, and then you think of ways to test them. And we'll do a lot of that in here.

OK, so that's what I just said, and I'll just say that there's lots of data since then. That region of the brain actually extremely very much prefers faces, and it's present in everyone. And next time, we will talk about the fact that that looks like it's suggesting that we have a different system for face recognition than object recognition. But we haven't yet nailed the case, and you guys should all think about what remains.

OK, thank you. Sorry I was racing there. I will hang out if you guys have questions.