GABRIEL KREIMAN: What I'd like to do today is give a very brief introduction to neural circuits, why we study them, how we study them, and the possibilities that come out of understanding biological codes, and trying to translate those ideas into computational codes. Then I will be a bit more specific, and discuss some initial attempts at studying the computational role of feedback signals.

And then I'll switch gears and talk for a few minutes about a couple of things that are not necessarily related to things that we've made any real work on, but I'm particularly excited about in the context of open question challenges, and opportunities, and what I think will happen over the next several years in the field. In the hope of inspiring several of you to actually solve some of these open questions in the field.

So one of the reasons why I'm very excited about studying biology and studying brains is that our brains are the product of millions of years of evolution. And through evolution, we have discovered how to do things that are interesting, fast, efficient. And so if we can understand the biological cause, if we can understand the machinery by which we do all of these amazing feats, that in principle, we should be able to take some of these biological codes, and write computer code that will do all of those things in similar ways.

In similar ways that we can write algorithms to compute the square root of 2, there could be algorithms that dictate how we see, how we can recognize objects, how we can recognize auditory events. In short, the answer to all of these Turing questions, in some sense, is hidden somewhere here inside our brain. So the question is, how can we listen to neurons and circuits, decode their activity, and maybe even write in information in the brain, and then trying to translate all of these ideas into computational codes.

So there's a lot of fascinating properties that biological codes cover. Needless to say, we're not quite there yet in terms of computers and robots. So our hardware and software worked for many decades. I think it's very unlikely that your amazing iPhone 6 or 5 or 7 whatever it is, will last four, five, six, seven, eight, nine decades. None of our computers will last that long. Our hardware does.

There's amazing parallel computation going on in our brains. This is quite distinct from the way we think about algorithms and computation in other domains now. Our brains have a reprogrammable architecture. The same chunk of tissue can be used for several different purposes. Through learning and through our experiences, we can modify those architectures.

A thing that has been quite interesting, and that maybe we'll come back to, is the notion of being able to do single shot learning, as opposed to some machine learning algorithms that require lots and lots of data to train. We can easily discover a structure in data.

The notion of fault tolerance and robustness to transformations is an essential one. Robustness is arguably a fundamental property of biology and one that has been very, very hard to implement in computational circuitry. And for engineers, the whole issue about how to have different systems integrate information, and interact with each other, has been and continues to be a fundamental challenge. And our brains do that all the time. We're walking down the street, we can integrate visual information, with auditory information with our targets, our plans, what we're interested in doing, on social interactions, and so on.

So why do we want to study neural circuits. So I think we are in the golden era right now, because we can begin to explore the answers to some of these Turing questions in brains at the biological level. So we can study high level cognitive phenomena at the level of neurons, and circuits of neurons. And I'll give you a few examples of that later on.

More recently, and I'll come back to this towards the end, we've had the opportunity to begin to manipulate, and disrupt, and interact with neural circuits at unprecedented resolution. So we can begin to turn on and off specific subsets of neurons. And that has tremendously accelerated our possibility to test theories at the neural level.

And then again, the notion being that empirical findings can be translated into computational algorithms-- that is, if we really understand how biology solves the problem, in principle, we should be able to write mathematical equations, and then write code that mimics some of those computations. And some of the examples of that, we talk about in the visual system in my presentation, but also in Jim DiCarlo's presentation.

These are just advertising for a couple of books that I find interesting and relevant in computational neuroscience. I'm not going to have time to do any justice to the entire field of computation neuroscience at all. So all these slides will be in Dropbox, so if anyone wants to learn more about computational neuroscience. These are lot of tremendous books. Larry Abbott is the author of this one, and he'll be talking tonight.

So how do we study biological circuitry. And I realize that this is deja vu and very well known for many of you. But in general, we have a variety of techniques to probe the function of brain circuits. And this is showing the temporal resolution of different techniques, and the spatial resolution of different techniques used to study neural circuits. All the way from techniques that have limited spatial and temporal resolution, such as PET and fMRI-- techniques that have very high temporal resolution, but relatively poor spatial resolution-- all the way to techniques that allow us to interrogate the function of individual channels with neurons.

So most of what I'm going to talk about today is what we refer to as the neural circuit level, somewhere in between single neurons and then ensembles of neurons recording the local field potential, which give us the resolution of milliseconds, where we think a lot of the computations in the cortex are happening, and where we think we can begin to elucidate how neurons interact with each other.

So to start from the very beginning, we need to understand what a neuron does. And again, many of you are quite familiar with this. But the basic fundamental understanding of what a neuron does is to integrate information-- receive information through its dendrites, integrates that information, and decides whether to fire a spike or not.

Interestingly, some of the basic intuitions of our neuron function were essentially conceived by a Spaniard, Ramón y Cajal. He wanted to be an artist. His parents told him that he could not become an artist, he had to become a clinician, a medical doctor. So he followed the tradition. He became a medical doctor. But then he said, well, what I really like doing is drawing. And so he bought a microscope, he put it in his kitchen, and he spent a good chunk of his life drawing, essentially. So he would look at neurons, and he would draw their shapes. And that's essentially how neuroscience started.

Just from these beautiful and amazing array of drawings of neurons, he conjectured the basic flow of information. This notion that this integration of information through dendrites, all of this integration happens in the soma. And from there, neurons decide whether to fire a spike or not. Nothing more, nothing less. That's essentially the fundamental unit of computation in our brains.

How do we think about and model those processes? There's a family of different types of models that people have used to describe what a neuron does. These models differ in terms of their biological accuracy, and their computational complexity. One of the most used ones is perhaps an integrate and fire neuron. This is a very simple RC circuit. It basically integrates current, and then through a threshold, the neuron decides when to fire or not to fire a spike.

This is essentially treating neurons as point masses. There are people out there who have argued that you need more and more detail. You need to know exactly how many dendrites you have, and the position of each dendrite, and on and on and on and on.

What's the exact resolution at which we should study neuron systems is a fundamental open question. We don't know what's the right level of abstraction. There are people who think about brains in the context of blood flow, and millions and millions of neurons averaged together. There are people who think that we actually need to pay attention to the exact details of how every single dendrite integrates information, and so on.

For many of us, this is a sufficient level of abstraction. The notion that there's a neuron that can integrate information. So we would like to push this notion that we can think about models with single neurons, and see how far we can go, understanding that we are ignoring a lot of the inner complexity of what's happening inside a neuron itself.

So very, very briefly just to push the notion that this is not rocket science. It's very, very easy to build these integrate-and-fire model simulations. I know many of you do this on a daily basis. This is the equation of the RC circuit. There's current that flows through a capacitance. There's current that flows through the resistance, which, this RC circuit, we think of as composed of the ion channels in the membranes of the neurons. And this is all there is to it in terms of a lot of the simulation that we use to understand the function of neurons.

And again, just to tell you that there's nothing scary or fundamentally difficult about this, here's just a couple of lines in MATLAB that you can take a look at if you've never done these kind of simulations. This is a very simple and perhaps even somewhat wrong simulation of an integrate-and-fire neuron. But just to tell you that it's relatively simple to build models of individual neurons that have these fundamental properties of being able to integrate information, and decide when to fire a spike.

The fundamental questions that we really want to tackle in CBMM have to do with putting together lots of neurons, and understanding the function of circuits. It's not enough to understand individual neurons. We need to understand how they interact together. We want to understand what is there, who's there, what are they doing to whom, and when, and why. We really need to understand the activity of multiple neurons together in the form of circuitry.

So just a handful of basic definitions. If we have a circuitry like this, where we start connecting multiple neurons together, information flows here in this circuitry in this direction. We refer to the connections between neurons that go in this direction as feed forward. We refer to the connections that flow in the opposite direction as feedback and I use the word recurrent connections for the horizontal connections within a particular layer. So this is just to fix the nomenclature for the discussion that will come next, and also today in the afternoon with Jim DiCarlo's presentation.

Throughout a lot of anatomical work, we have begun to elucidate some of the basic connectivity between neurons in the cortex. And this is the primary example that has been cited extremely often of what we understand about the connectivity between different areas in the macaque monkey. We don't have a diagram like this for the human brain. Most of the detailed anatomical work has been done in macaque monkeys. So each of these boxes here represents a brain area, and this encapsulates our understanding of who talks to whom, or which area talks to which other area in terms of visual cortex. There's a lot of different parts of cortex that represent visual information.

Here at the bottom, we have the retina. Information from the retina flows through to the LGN. From the LGN, information goes to primary visual cortex, sitting right here. And from there, there's a cascade that is largely parallel, and at the same time, hierarchical, of a conglomerate of multiple areas that are fundamental in processing visual information. We'll talk about some of these areas next. And we'll also talk about some of these areas today in the afternoon when Jim discusses what are the fundamental computations involved in visual object recognition.

One of the fundamental clues as to how do we understand, how do we know that this is a particular visual area, how do we know that this is important for our vision, has come from anatomical lesions. Mostly in monkeys, but in some cases, in humans as well. So if you make lesions in some of these areas, depending on exactly where you make that lesion, people either become completely blind, or they have a particular scotoma, a particular chunk of the visual field where they cannot see. Or they have more high order types of deficits in terms of visual recognition.

As an example, the primary visual cortex was discovered by people who were of the [INAUDIBLE] they were studying, the trajectory of bullets in soldiers during World War I. And by discovering that some of those peoples had a blind part to their visual field, and that was a topographically organized depending on the particular trajectory of the bullet through their occipital cortex. And that's how we became to think about V1 as fundamental in visual processing.

It is not a perfect hierarchy. It's not there is A, B, C, D. Right? For a number of reasons. One is that there are lots of parallel connections. There are lots of different stages that are connected to each other. And one of the ways to define a hierarchy is by looking at the timing of the responses in different areas.

So if you look at the average latency of the response in each of these areas, you'll find that there's an approximate hierarchy. Information gets out of the retina approximately at 50 milliseconds. About 60 or so milliseconds in LGN, and so on. So it's approximately a 10 millisecond cost per step in terms of the average latency. However, if you start looking at the distribution, you'll see that it's not a strict hierarchy. For example, there are neurons in area V4 that are the early neurons in V4 may fire before the late neurons in V1. And that shows you that the circuitry is far more complex than just a simple hierarchy.

One way to put some order into this seemingly complex and chaotic circuitry, one simplification is that there are two main pathways. One is the so-called what pathway. The other one is the so-called where pathway. The what pathway essentially is the ventral pathway. It's mostly involved in object recognition, trying to understand what is there. The dorsal pathway, the where pathway, is most involved in motion, and being able to detect where objects are, stereo, and so on. Again, this is not a strict division, but it's a pretty good approximation that many of us have used in terms of thinking about the fundamental computations in these areas.

Now we often think about these boxes, but of course, there's a huge amount of complexity within each of these boxes. So if we zoom in one of these areas, we discover that there's a complex hierarchy of computations. There are multiple different layers. The cortex is essentially a six layer structure. And there are specific rules. People have referred to this as a canonical micro circuitry. There's a specific set of rules in terms of how information flows from one layer to another in terms of each of these cortical structures.

To a first approximation, this canonical circuitry is common to most of these areas. There are these rules about which layer receives information first, and sends information to areas are more or less constant throughout the cortical circuitry. This doesn't mean that we understand this circuitry well, or what each of these connections is doing. We certainly don't. But these are initial steps to sort of decipher some of these basic biological connectivity that has fundamental computational properties for vision processing.

So our lab has been very interested in what we call the first order approximation or immediate approximation to visual object recognition. The notion that we can recognize objects very fast, and that this can be explained, essentially, as the bottom-up hierarchical process. Jim DiCarlo is going to talk about this extensively this afternoon, so I'm going to essentially skip that, and jump into more recent work that we've done trying to think about top-down connections.

But just let me briefly say why we think that the first pass of visual information can be semi-seriously approximated by these purely bottom-up processing. One is that at the behavioral level, we can recognize objects very, very fast. There's a series of psychophysical experiments that demonstrate that if I show you an object, recognition can happen within about 150 milliseconds or so.

We know that the physiological signals underlying visual object recognition also happen very fast. Within about 100 to 150 milliseconds, we can find neurons that show very selective responses to complex objects, and again, you'll see examples of that this afternoon.

The behavior and the physiology have inspired generations of computational models that are purely bottom-up, where there is no recurrency, and that can be quite successful in terms of visual recognition. To our first approximation, the recent excitement with deep convolutional networks can be traced back to some of these ideas, and some of these basic biologically inspired computations that are purely bottom-up. So to summarize-- and I'm not going to give any more details-- we think that the first 100 milliseconds or so of visual processing can be approximated by these purely bottom-up, semi hierarchical sequence of computations.

And this leaves open a fundamental question, which is, why we have all these massive feedback connections? We know that in cortex, there are actually more recurrent and feedback connections than feed-forward ones. And what I'd like to talk about today is a couple of ideas of what all of those feedback connections may be doing.

So this is an anatomical study looking at a lot of the boxes that I showed you before, and showing how many of the connections to any given area come from one of these other variants. For example, if we take just primary visual cortex, this is saying that a good fraction of the connections to primary visual cortex actually come from V2. That's from the next stage of processing, rather than from V1 itself.

All in all, if you quantify for a given neuron in V1, how many signals are coming from a bottom-up source that is for LGN versus how many signals are coming from other V1 neurons or from higher visual areas, it turns out that there are more horizontal and top-down projections than bottom-up ones. So what are they doing? If we can approximate the first 100 milliseconds or so of vision so well with bottom-up hierarchies, what are all these feedback signals doing?

So this brings me to three examples that I'd like to discuss today of recent work that we've done to take some initial principles in thinking about what this feedback connections could be doing in terms of visual recognition. So I'll start by giving you an example of trying to understand the basic fundamental unit of feedback. That is these canonical computations, and by looking at the feedback that happens from V2 to V1 in the visual system.

Next, I'm going to give you an example of what happens during a visual search, where we also think that feedback signals may be playing a fundamental role, if you have to do or Where's Waldo kind of task, where you have to search for objects and in the environment. And finally, I will talk about pattern completion, how you can recognize objects that are heavily occluded, where we also think that feedback signals may be playing an important role.

So before I go on to describe what we're seeing the feedback from V2 to V1 maybe doing, let me describe very quickly classical work that Hubel and Wiesel did that got them the Nobel Prize by recording the activity of neurons in primary visual cortex. They started working in kittens, and then subsequently in monkeys, and discovered that there are neurons that show orientation tuning, meaning that they respond very vigorously.

These are spikes, each of these marks corresponds to an action potential, the fundamental language of computation in cortex. And this neuron responds quite vigorously when the cat was seeing a bar of this orientation. And essentially, there's no firing at all with this type of stimulus in the receptive field.

This was fundamental because it transformed our understanding of the essential computations in primary visual cortex in terms of filtering the initial stimulus. This is what we now describe by Gabor functions. And if you look at deep convolutional networks, many of them, if not perhaps all of them, start with some sort of filtering operation that is either Gabor filters or resembles this type of orientation that we think is a fundamental aspect of how we start to process information in the visual field.

One of the beautiful things that Hubel and Wiesel did is not only to make these discoveries, but also to come up with very simple graphical models of how they thought this could come about. And this remains today one of the fundamental ways in which we think about how our orientation tuning may come about.

If you recall the activity of neurons in the retina or in the LGN, you'll find what's called center surround receptive fields. These are circularly symmetric receptive fields, with an area in the center that excites the neuron, and an area in the surround that inhibits the neuron. What they conjecture is that if you put together multiple LGN cells, whose receptive fields are aligned along a certain orientation, and you simply combine all of them, you simply add the responses of all of those neurons, you can get a neuron in the primary visual cortex that has orientation tuning. This

is a problem that's far from solved, despite the fact that we have four or five decades. There are many, many models of how orientation tuning comes about. But this remains one of the basic bottom-up feed-forward ideas of how you can actually build orientation tuning from very simple receptive fields.

This has informed a lot of our thinking about how basic computations can give rise to orientation tuning in a purely bottom-up fashion.

In primary visual cortex, in addition to the so-called simple cells, are complex cells that show invariance to the exact position or the exact phase of the oriented bar within the receptive field. And that's illustrated here. So this is a simple cell. So this simple cell has orientation tuning, meaning that it responds more vigorously to this orientation than to this orientation.

However, if you change the phase or the position of the oriented bar within the receptive field, the response decreases significantly. In contrast to this complex cell that not only has orientation tuning, meaning that it fires more vigorously to this orientation than to this one, but also has phase invariance, meaning that the response is more or less the same way, regardless of the exact phase or the exact position of the stimulus within the receptive field.

And again, the notion that they postulated is that we can build these complex cells by a summation of activity or multiple simple cells. So again, if you imagine now that you have multiple simple cells with different receptive fields that are centered at these different positions, you can add them up, and create complex cells.

These fundamental operations of simple and complex cells and primary visual cortex can be somehow traced to the root of a lot of the bottom-up hierarchical models. A lot of the deep convolutional networks today essentially have variations on these kind of themes, of filtering steps, nonlinear computations that give you invariance, and a concatenation of these filtering and invariance steps along the visual hierarchy.

So in following up with this idea, I would like to understand the basics of what's the kind of information that's provided when you have signals from V2 to V1. To do that, we have been collaborating with Richard Born at Harvard Medical School, who has a way of implanting cryo loops. This is a device that can be implanted in monkeys in areas V2, and V3, lower the temperature, and thus reduce or essentially eliminate activity from areas V2 and V3. So that means that we can study V1 without activity in area V2 and V3. We can study V1 sans feedback.

So this is an example of recordings of a neuron in this area. This is the normal activity that you get from the neuron. Here is when they present a visual stimulus. This is a spontaneous activity. Each of these dots corresponds to a spike. Each of these lines correspond to a repetition of the stimulus. This is a traditional way of showing raster plots for neuron responses. So you see that this is a spontaneous activity. You present the stimulus. There's an increase in the response of this neuron, as you might expect.

Actually, I'm sorry. This actually starts here. So this is the spontaneous activity, this is the response. Now here, they turn on their pump. They start lowering the temperature. And you see within a couple of minutes, they essentially significantly reduce the responses. The largely silence-- not completely-- but largely silence activity in areas V2 and V3. And these are reversible, so when they turn the pumps off, activity comes back in. So the question is, what happens in primary visual cortex when you don't have feedback from V2 and V3.

So the first thing they have characterized is that some of the basic properties of V1 do not change. It's consistent with the simple models that I just told you, where the orientation tuning in the primary visual cortex is largely dictated by the bottom-up inputs, by the signals from the LGN. The conjecture from that would be that if you silence V2 and V3, nothing would happen with orientation tuning in primary visual cortex. And that's essentially what they're showing here.

These are example neurons. This is showing orientation selectivity. This is showing direction selectivity, what happens when you move an oriented bar within the receptive field. So this is showing the direction. This is showing the mean normalized response of a neuron. This is the preferred direction, and direction orientation that gives a maximum response.

The blue curve corresponds to when you don't have activity in V2 and V3. Red corresponds to their control data. And essentially, the tuning of the neuron was not altered. The orientation preferred by this neuron was not altered. The same thing goes for direction selectivity.

So the basic problems of orientation tuning and direction selectivity did not change. Let me say a few words about the dynamics of the responses. So here, what I'm showing you is the mean normalized responses as a function of time. Time 0 is when the stimulus is turned on. As I told you already, by about 50 milliseconds or so, you get a vigorous response in primary visual cortex. And if we compare the orange and the blue curves, we see that this initial response is largely identical. So the initial response of these V1 neurons is not affected by the absence of feedback from V2.

We start to see effects, we start to see a change in the firing rate here. Largely at about 60 milliseconds or so after presentation. So in a highly oversimplified cartoon, I think of this as a bottom-up Hubel and Wiesel like response, driven by LGN. And signals from V2 to V1 coming back about 10 milliseconds later. And that's when we started seeing some of these feedback related effects.

I told you that some of the basic properties do not change. We interpret this as being dictated largely by bottom-up signals. The dynamics do change. The initial response is unaffected. The later part of the response is affected. I want to say one thing that does change. And for that, I need to explain what an area summation curve is.

So if you present the stimulus within the receptive field of a neuron of this size, you get a certain response. As you start increasing the size of this stimulus, you get a more vigorous response. Size matters. The larger, the better-- to a point. There comes a point where it turns out that the response of the neurons starts decreasing again.

So larger is not always better. A little bit larger is better. This size has an inhibitory effect overall on the response of the neuron. This is called surround suppression. And these curves have been characterized in areas like primary visual cortex. Also in earlier areas for a very long time.

It turns out that when you do these type of experiments in the absence of feedback, the effect of surround suppression does not disappear. That is, you still have a peak in the response as a function of a stimulus size. But there is a reduced amount of surround suppression. That is, when you don't have feedback, there's less suppression. You have a larger response for bigger stimulus.

So we think that one of the fundamental computations that feedback is providing here is this integration from multiple neurons in V1 that happens in V2. And then inhibition to activity of neurons in area V1 to provide some of the suppression. This is partly the reason why our neurons are not very excited about a uniform stimulus, like a blank wall. Our neurons are interested in changes, and part of that, we think, is dictated by this feedback from V2 to V1.

We can model these center surround interactions as a ratio of two Gaussian curves, two forces. One is the one that increases the response. The other one is a normalization term that suppresses the response when the stimulus is too large. There's a number of parameters here. Essentially, you can think of this as a ratio of Gaussians, ROGs. There's a ratio of two Gaussian curves. One dictating the center that responds. The other one, the surround response.

And to make a long story short, we can feed the data from the monkey with this extremely simple ratio of Gaussian's model. And we can show that the main parameter that feedback seems to be acting upon is what we call Wn-- that is this normalization factor here. So that the tuning factor that dictates the strength of the surrounding division from V2 to V1-- we think that's one of the fundamental things that's being affected by feedback.

So we would think of this as the gain. We think of this as the spatial extent over which the V2 can exert its action on primary visual cortex. We think that's the main thing that's affected here.

This type of spatial effect may be important in other role that has been ascribed to feedback, which is the ability to direct attention to specific locations in the environment. I want to come back to this question here, and ask, under what conditions, and how can a feedback also provide important features specific signals from one area to another. And for that, I'm going to switch to another task, another completely different prep, which is the Where's Waldo task--

the task of visual search. How do we search for particular objects in the environment.

And here, it's not sufficient to focus on a specific location, but we need to be able to search for specific features. We need to be able to bias our visual responses for specific features of the stimulus that we're searching for.

So this is a famous sort of Where's Waldo task. You need to be able to search for specific features. It's not enough to be able to send feedback from V2 to V1, and direct attention, or change the sizes of the receptive fields, or the direct attention to a specific location.

Another version that I'm not going to talk about of visual that has a related theme that relates to visual search is feature based attention, when you're actually paying attention to a particular face, to a particular color, to a particular feature that is not necessarily located, and to space, as our friend here has studied quite significantly. People always like to know the answer of where he is at.

OK. So let me tell you about a computational model and some behavioral data that we have collected to try to get at this question of how feedback signals can be relevant for visual search. This initial part of this computational model is essentially the HMAX type of architecture that has been pioneered by Tommy Poggio and several people in his lab, most notably, people like Max Riesenhuber and Thomas Serre. I was thinking that by this time, people would have described this in more detail. I'm going to go through these very quickly. Again, today in the afternoon, we'll have more discussion about this family of models.

So these family of models essentially goes through a series of linear and non-linear computations in a hierarchical way, inspired by the basic definition of simple and complex cells that I described in the work of Hubel and Wiesel. So basically, what these models do is they take an image. These are pixels. There's a filtering step. This filtering step involves Gabor filtering of the image. In this particular case, there are four different orientations. And what do you get here is a map of the visual input after this linear filtering process.

The next step in this model is a local max operation. This is pooling neurons that have similar identical feature preferences, but slightly different scale in the receptive fields. Or slightly different positions in their receptive fields. And this max operation, this non-linear operation is giving you invariance to the specific feature. So now you can get a response to the same feature, irrespective of the exact scale or the exact position within the receptive field.

These were labeled S1 and C1, initially in models by Fukushima. And this type of nomenclature was carried on later by Tommy and many others. And this is directly inspired by the simple and complex cells that I very briefly showed you previously in the recordings of Hubel and Wiesel.

These filtering and max operations are repeated throughout the hierarchy again and again. So here's another layer that has a filtering step and a nonlinear max step. In this case, this filtering here is not a Gabor filter. We don't really understand very well what neurons in V2 and V4 are doing. One of the types of filters that have been used and that we are using here is a radial basis function, where the properties of a neuron in this case are dictated by patches taking randomly from natural images.

All of this is purely feed-forward. All of this is essentially the basic ingredient of the type of convolutional networks that had been used for object recognition. You can have more layers. You can have different types of computations. The basic properties are essentially the ones that are described briefly here.

What I really want to talk about is not the former part, but this part of the model. Now I ask you, where's Waldo, you need to do something, you need be able to somehow look at this information, and be able to bias your responses or bias the model towards regions of the visual space that have features that resemble what you're looking for. Your car, your keys, Waldo.

So the way we do that is first, in this case, I'm going to show you what happens if you're looking for the top hat here. So first, we have a representation in the model of the top hat. This is the hat here. And we have a representation in our vocabulary of how units in the highest echelons of this model represent this hat. So we have a representation of the features that compose this object at a high level in this model.

We use that representation to modulate, in a multiplicative fashion, the entire image. Essentially, we bias the responses in the entire image based on the particular features that we are searching for. This is inspired by many physiological experiments that have shown that to a good approximation, this type of modulation in feature based attention has been observed across different parts of the visual field. That is, if you're searching for red objects, neurons that like red will enhance their response throughout the entire visual field. So have the entire visual field modulated by the pattern of features that we're searching for in here.

After that, we have a normalization step. This normalization step is critical in order to discount purely bottom-up effects. We don't want the competition between different objects to be purely dictated by which object is brighter, for example. So we normalize that after modulating that with the features that we are searching.

That gives us a map of the image, where each area has been essentially compared to this feature set that we're looking for. And then we have a winner take all mechanism that dictates where the model will pay attention to, or where the model will fixate on first. Where the model thinks that a particular object is located.

OK so what happens when we have this feedback that's feature specific, and that modulates the responses based on the targets object that we're searching for. In these two images, either in objects arrays or when objects are embedded in complex scenes, we're searching for this top object. And the largest response in the model is indeed in the location of where the object is. In these other two images, the model is searching for this accordion here. And again, the model was able to find that by this comparison of the features with the stimulus.

More generally, these are object array images. This is the number of fixations required to find the object in this object array images. So one would correspond to the first fixation. If the model does not find the object in the first location, there's what's called inhibition of return. So we make sure the model does not come back to the same location, and the model will look at the second best possible location in the image. And it will keep on searching until it finds the object. So the model performs in the first fixation at 60% correct. And eventually, after five fixations, it can find the object almost always right in here.

This is what you would expect by random search. If you were to randomly fixate on different objects, so the model is doing much better than that. And then for the aficionados, there's a whole plethora of purely bottom-up models that don't have feedback whatsoever. This is a family of models that was pioneered by people like Laurent Itti and Christof Koch. These are saliency based models. Although you cannot see, there are a couple of other points in here. All of those models cannot find the object either. It's not that these objects that we're searching for are more salient, and therefore, that's why the model is finding them. We really need something more than just bottom-up pure saliency.

We did a psychophysical experiment. We asked, well, this is how the model searches for Waldo. How will humans search for objects under the same conditions. So we had multiple objects. Subjects have to make a saccade to a target object. To make a long story short, this is the cumulative performance of the model and the number of fixations under these conditions, and the model that's reasonable in terms of how well humans do. This is data from every single individual subject in the task.

I'm going to skip some of the details. You can compare the errors that the model is making. How consistent people are with themselves with respect to other subjects. How good it is with respect to humans. The long story is the model is far from perfect. We don't think that we have captured everything we need to understand about visual search. Some people alluded to before, for example, the notion that the model doesn't have these major changes with eccentricity, and the fovea, and so on. A long way to go, but we think that we've captured some of the essential initial ingredients of visual search. And that this is one example of how visual feedback signals can influence this bottom-up hierarchy for recognition.

I want to very quickly move on to a third example that I wanted to give you of how feedback can help in terms of visual recognition. What are other functions that feedback could be playing. And for that, I'd like to discuss the work that Hanlin did here, and also, Bill Lotter in the lab, in terms of how we can recognize objects that are partially occluded.

This happens all the time. So you walk around and see objects in the world. You can also encounter objects where you can only find partial information, and you have to make pattern completion. Pattern completion is a fundamental aspect of intelligence. We do that in all sorts of scenarios. It's not just restricted to vision. All of you can probably complete all of these patterns.

We use pattern completion in social scenarios as well, right? You make inferences from partial knowledge about their intentions, and what they're doing, and what they're trying to do, OK? So we want to study this problem of how you complete pattern, how you extrapolate from partial limited information in the context of visual recognition.

There are a lot of different ways in which one can present partially occluded objects. Here are just a few of them. What Hanlin did was use a paradigm called bubbles that's shown here. Essentially, it's like looking at the world like these. You only have small windows through which you can see the object. Performance can be titrated to make the task harder or easier. So if you have a lot of bubbles, it's relatively easy to recognize that this is a toy school bus. If you have only four bubbles, it's actually pretty challenging. So we can titrate performance on the difficulty of this task.

Very quickly, let me start by showing you psychophysics performance here. This is how subjects perform as a function of the amount of occlusion in the image as a function of how many pixels you're showing for these images. And what you see here is that with 60% occlusion, performance is extremely high. Performance essentially drops to chance level when the object is more and more occluded. There is a significant amount of robustness in human performance. For example, you have a little bit more than 10% of the pixels in the object, and people can still recognize them reasonably well. So this is all behavioral data.

Let me show you very quickly what Hanlin discovered by doing invasive recordings in human patients while the subjects were performing this recognition of objects that are partially occluded. It's illegal to put electrodes in the human brain in normal people, so we work with subjects that have pharmacological intractable epilepsy. So inside of subjects that have seizures, the neurosurgeons need to implant electrodes in order to localize the seizures. And B, in order to ensure that when they do a resection, and they take out the part of the brain that's responsible for seizures, that they're not going to interfere with other functions, such as language.

These patients stay in the hospital for about one week. And during this one week, we have a unique opportunity to go inside a human brain, and record physiological data. Depending on the type of patient, we've used the different types of electrodes. This is what some people refer to as ECoG electrodes. Electrocorticographic signals. These are field potential signals, very different from the ones that I was showing you in the little spikes before. These are aggregate measures, probably of tens of thousands, if not millions of neurons, where we have very, very high temporal resolution at the millisecond level, but very poor spatial resolution, only being able to localize things at the millimeter level or so.

With these, we can pinpoint specific locations within about approximately one millimeter, but have very high signal to noise ratio signals that are dictated by the visual input. An example of those signals is shown here. These are intracranial field potentials as a function of time. This is the onset of the stimulus. And these 39 different repetitions, when Hanlin is showing this unoccluded face, we see a very vigorous change, quite systematic from one trial to another. All of those gray traces are single trials, similar to the raster plot that I was showing you before.

So now I'm going to show you a couple of single trials. We're showing individual images where objects are partially occluded. In this case, there's only about 15% of the pixels of the face that are being shown. And we see that despite the fact that we're covering 85%, more or less, of that image, we still see a pretty consistent physiological signal. The signals are clearly not identical. For example, this one looks somewhat different. There's a lot of our ability from one to another. But again, these are just single trials showing that there still is selectivity for these shape, despite the fact that we are only showing a small fraction of this thing.

These are all the trials in which these five different faces were presented. Each line corresponds to trial. These are raster plots. As you can see, the data are extremely clear. There's no processing here. This is raw data single trials. These are single trials with the partial images. You again can see there's a vigorous response here. The responses are not as nicely and neatly aligned here, in part because all of these images are different. All of the locations on the models are different. As I just showed you, there's a lot of variability here.

If you actually fix the bubble locations-- that is, you repeatedly present the same image multiple times still in pseudorandom order, but the same image, you see that the signals are more consistent. Not as consistent as this one, but certainly more consistent. Again, very clear selective response tolerant to a tremendous amount of occlusion in the image.

Interestingly, the latency of the response is significantly later compared to the whole images. So if you look at, for example, 200 milliseconds, you see that the responses started significantly before 200 milliseconds for the whole images. All of the responses here start after 200 milliseconds. We spent a significant amount of time trying to characterize this and showing that pattern completion, the ability to recognize objects that are occluded, involves a significant delay at the physiological level.

If you use the purely bottom-up architecture and tried to do this in silico-- this bottom-up model does not perform very well. The performance deteriorates quite rapidly when you start having significant occlusion.

I'm going to skip this and just very quickly argue about some of the initial steps that Bill Lotter has been doing, trying to add recurrency to the models. Trying to have both feedback connections as well as recurrent connections within each layer to try to get a model that will be able to perform pattern completion, and therefore, use these feedback signals to allow us to extrapolate from previous information about these objects. Bill will be here Friday or Monday, I'm not sure. So you should talk to him more about these models.

Essentially, they belong to the family of HMAX. They belong to a family of convolutional networks, where you have filter operations, threshold, and saturation pooling on normalization. Jim will say about this family of models today in the afternoon. These are purely bottom-up models. And what Bill has been doing is other than recurrent and feedback connections, retraining these models based on these recurrent and feedback connections, and then comparing their performance with human psychophysics.

So this is the behavioral data that I showed you before. This is the performance of the feedforward model. This is the recurrent model that was able to train.

Another way to try to get out whether feedback is relevant for pattern completion is to use with backward masking. Backward masking means that you present an image, and immediately after that image, within a few milliseconds, you present noise. You present a mask. And people have argued that masking essentially interrupts feedback processing. Essentially, it allows you to have a bottom-up flow of information-- stops feedback.

I don't think this is quite extremely rigorous. I think that the story is probably far more complicated than that. But to a first approximation, you present a picture, you have a bottom-up stream, you put a mask, and you interrupt all the subsequent feedback processing.

So if you do that at the behavioral level, you can show that when stimuli are masked, particularly if the interval is very short, you can significantly impair pattern completion performance. So if the mask comes within 25 milliseconds of the actual stimulus performance in recognizing these heavily occluded objects is significantly impaired. We interpreted this to indicate that feedback may be needed for pattern completion.

This is Bill's instantiation of that recurrent model. Because he has recurrency now, he also has time in this models. So he can also present the image, present the mask to the model, and compare the performance of the computational model as a function of the occlusion in unmasked and the masked conditions.

So to summarize this-- and there's still two or three more slides that I want to show-- I've given you three examples of potential ways in which feedback signals can be important. The first one has to do with the effects of feedback on surround suppression, going from V2 to V1. We think that by doing this type of experiments combined with the computational models to understand what are the fundamental computations, we can begin to elucidate some of the steps by which feedback can exert its role. We hoped to come up with the essential alphabet of computations similar to the filtering and normalization operations that are implemented by feedback.

The second example was feedback as being able to have features that dictate what we do in visual search tasks and the last example, in both our preliminary work, trying to use feedback, as well as recurrent connections to perform pattern completion and extrapolate from prior information.

So the last thing I wanted to do is just flash a few more slides about a couple of things that are happening in neuroscience and computational neuroscience that I think are tremendously exciting for people. If I were young again, these are some of the things that I would definitely be very, very excited to follow up on.

So the notion that we'll be able to go inside brains and read our biological code, and eventually write down computer code, and build amazing machines is, I think, very appealing and sexy. But at the same time, it's a far cry, right? We're a long way from being able to take biological codes and translate that into computational codes. It's really extremely tragic.

So here are three reasons why I think there's optimism that this may not be as crazy as it sounds. We're beginning to have tremendous information about wiring diagrams at exquisite resolution. There are a lot of people who are seriously thinking about providing us with maps about which neuron talks to which other neuron. And this was not present ever before. So we are now beginning to have detailed information that it's much higher resolution connectivity than ever before.

The second one is the strength in numbers. For decades, we've been recording the activity of one neuron at a time, maybe a few neurons at a time. Now there are many different ideas and techniques out there by which we can listen to and monitor the activity of multiple neurons simultaneously. And I think this is going to be game changing for neurophysiology, but also for the possibility of reputational models that are inspired by biology.

And the third one is a series of techniques mostly developed by people like Ed Boyden and Karl Deisseroth to do optogenetics, and to manipulate these circuits with unprecedented resolution. So let me expand on that for one second. This is the C. elegans. This is an intramicroscopy image of how one can categorize the circuitry. So it turns out that this pioneering work of Sydney Brenner a couple of decades ago has led to mapping the connectivity of each one of the 302 neurons. How exactly for each neuron, who it's connected with. And this is represented in that rather complex way in this diagram here.

Well, it turns out that people are beginning to do these type of heroic type of experiments in cortex. So we're beginning to have initial insights about connectivity about how neurons are wired with each other at this resolution in cortex. We're nowhere near being able to have these for humans. Not even other species, mice, and so on. Not even Drosophila yet.

There's a huge amount of [INAUDIBLE] and interest in the community of having a very detailed map. So the question for you for the young and next generation, what are we going to do with these maps. If I give you a fantastic detailed wiring diagram of a chunk of cortex, how is that going to transform our ability to make inferences, and build new computational models.

The second one has to do with our ability to start the recording for more and more neurons. This is that other I didn't have time to talk about. This is work also that Hanlin did with Matias Ison and Itzhak Fried. These are recordings of spikes from human cortex, again, in patients that have epilepsy. I'm just flashing this slide because I had it handy. These are 300 neurons. This is not a simultaneously recorded population.

These are cases where we can record from a few neurons at a time using micro wires now. This is different from the type of recording that I showed you before. These are actual spikes that we can record. And these 380 neurons is in a different task. So recording from these 318 neurons took us about three to four years of time.

There are more and more people that are using either two photon imaging and/or massive multielectrode arrays that are beginning to be able to record the activity of hundreds of neurons simultaneously. My good friend and crazy inventor, Ed Boyden, believes that we will be able to recover from 100,000 neurons simultaneously. Of course, he is far more grandiose than I am, and he can think big at this kind of scale. But even to think about the possibility of recording from 1,000 or 5,000 neurons simultaneously so that in a week or a month, one may be able to have a tremendous amount from a very large population. This is going to be transformative.

Three decades ago in the field of molecular biology, people would sequence a single gene, and they would publish the entire sequence-- ACCGG-- and so on. That was the whole paper. A grad student would spend five years just sequencing a single gene. Now we have the possibility of downloading genome by advances in technology.

I suspect that a lot of our recordings will become obsolete. We'll be able to listen to the activity of thousands of neurons simultaneously. And again, it's for your generation to think about how this will transform our understanding of how quick we can read biological codes.

In the unlikely event that you think that that's not enough, here's one more thing that I think is transforming how we can decipher biological codes. And that's again, Ed Boyden using techniques that are referred to as optogenetics, where you can manipulate the activity of specific types of neurons.

I flashed a lot of computational models today. A lot of hypotheses about what different connections may be doing. At some point, we will be able to test some of those hypotheses with unprecedented resolution. So if somebody wanted to know what is this neuron V2, what kind of feedback its providing, we may be able to silence only neurons in V2 that provide feedback to V1 in a clean manner without affecting, for example, all of the other feed-forward processes, and so on. So the amount of specificity that can be derived from these type of techniques is enormous.

So that's all I wanted to say. So because we have very high specificity in our ability to manipulate circuits, because we'll be able to record the activity of many, many more neurons simultaneously, and because we'll have more and more detailed diagrams, I think that the dream of being able to read out and decode biological codes, and translate those into competition codes is less crazy than it may sound. We think that in the next several years and decades, smart people like you will be able to make this tremendous transformation and discover specific algorithms about intelligence by taking direct inspiration from biology.

So that's what's illustrated here. We'll be happy to keep on fighting. Andrei and I will fight. We will be happy to keep on fighting about Eva and how amazing she is and she isn't. What I try to describe is that by really understanding biological codes, we'll be able to write amazing computational code. I put a lot of arrows here. I'm not claiming QED. I'm not saying that we solve the problem. There's a huge amount of work that we need in here.