BRAIN SCIENCE

with

Ginger Campbell, MD

Episode #139

Interview with Jeff Hawkins, Author of *On Intelligence*

aired 11/28/17

[music]

INTRODUCTION

Welcome to *Brain Science*, the show for everyone who has a brain. I'm your host, Dr. Ginger Campbell, and this is Episode 139. Ever since I launched *Brain Science* back in 2006, my goal has been to explore how recent discoveries in neuroscience are helping unravel the mystery of how our brains make us human. I'm really excited about today's interview because, in some ways, it takes us back to the beginning.

My guest today is Jeff Hawkins, author of *On Intelligence*, and founder of Numenta, a company that is dedicated to discovering how the human cortex works. Jeff's book actually inspired the first *Brain Science* podcast, and I interviewed him way back in Episode 38. Today he gives us an update on the last 15 years of his research.

As always, episode show notes and transcripts are available at brainsciencepodcast.com. You can send me feedback at brainsciencepodcast@gmail.com or audio feedback via SpeakPipe at speakpipe.com/docartemis. I will be back after the interview to review the key ideas and to share a few brief announcements, including a look forward to next month's episode.

[music]

INTERVIEW

Dr. Campbell: Jeff, it is great to have you back on *Brain Science*.

Mr. Hawkins: It's great to be back, Ginger. I always enjoy talking to you.

Dr. Campbell: It's actually been over nine years since we last talked, so I thought we would start by asking you to just give my audience a little bit of background, and I'd like you to start by telling us just a little about your career before Numenta.

Mr. Hawkins: Yes, sure. I have a sort of an A-B-A career, you might call it. When I got out of college, I fell in love with brains, and I decided that this was what I was going to do with my life, I wanted to be a theoretical neuroscientist. I wanted to understand theories of the neocortex and try to understand how the brain works.

And I became a graduate student at Berkeley in the mid-80s, and I found out at that time that you really couldn't be a theoretician, you had to be an experimentalist. And I didn't really want to be an experimentalist, I really wanted to be focusing on theory. So, I put my plans on hold (I thought it would be for about four years) to go back and work in the computing industry—which I had done work in before.

And so, I did all that. And I went back, and I went back for more than four years. Because, during that time, I started two mobile computing companies, Palm and

Handspring. And they became successful and large, and they occupied my time. But everybody I worked with knew I really wanted to be doing neuroscience research. Everybody knew this: my investors knew it, my employees knew it. It was like everyone knew this; this is what my goal was.

And so, I finally extracted myself from the mobile computing space about 15 years ago. And then, the first thing I did was I started the Redwood Neuroscience Institute, which is now at Berkeley, and I ran that for three years. And then, I started a company called Numenta, which I'm still at right now.

And Numenta is like private research lab. And we have a team of scientists and engineers, primarily, and we focus on cortical theory. We're doing what I always wanted to do. And the whole period in mobile computing—which I'm reasonably well known for—that, to me, was like an actor who really wanted to act but was waiting tables. So, I was doing mobile computing, when I really wanted to get back into brain theory. And I used to describe it that way at the time; so, it's not revisionist history.

But we've been doing great. And we've been making some great progress. I'm really excited to talk about some of the progress we've made in understanding how the neocortex works.

Dr. Campbell: Yes, and I'm going to give you plenty of time. But the reason that I'm asking you about this is because I have a lot of listeners with all kinds of backgrounds who are discovering their passion for neuroscience, and they think that maybe they can't follow it because they've already taken some other path. And so, I think your story could inspire people that might share the sidetracks.

And I'm really curious, given that I've interviewed guys from the Allen Brain Institute a couple of times, and I realize that was founded by Paul Allen from Microsoft, but I don't know much about Allen's story and whether you share anything in common, other than starting things that were going to explore the brain.

Mr. Hawkins: I would think this is a fair characterization that I'm a practicing neuroscientist. I would think people would say I am extremely knowledgeable about this field and I'm doing active research. I don't think Paul Allen is doing that. I think he is very interested in many things, including brain theory, and he has basically used his financial resources to create an institute, but I don't think he is practicing. And on the other hand, I have a very small group here—it's not a large group like the Allen Institute—but I'm hands-on and doing actual research.

So, I'm not sure there are a lot of parallels there, other than there are many people who are very interested in how the brain works, and we come at it from different directions, and we use our resources as best we can.

Dr. Campbell: Okay. So, you alluded to this a minute ago, but as I understand and remember reading in your book, *On Intelligence* (which actually inspired the very first episode of this show over 10 years ago) that your interest is in the cortex of the brain and trying to model it? It that basically the...?

Mr. Hawkins: Well, it's really trying to understand how it works. And I'll tell you what that means—"to understand how it works "—from my perspective.

But just to remind your listeners, the neocortex in a human brain is somewhere between 70% and 75% of the volume of the human brain. And only mammals have a neocortex—a proper neocortex. And it, of course, is the organ that we most associate with all of our intelligence functions. So, language—my neocortex is generating the language right now that I'm speaking, and you and your listeners' neocortex is interpreting that—and high-level vision, and planning, and so on, all this stuff occurs in the neocortex.

So, if you're interested in intelligence, it is the organ to study. It can't be understood completely on its own, but a lot of it can be. And so, that's the area of my focus. And I think there are many, many people is neuroscience who ultimately think that one of the most important goals in neuroscience is to understand the neocortex because it's so associated with us as humans, and our intelligence, and our ability to even talk about these problems.

So, that is my focus. And it includes other regions, too; we'll talk in a minute about, well, we have to understand many other parts of the brain to understand the neocortex, specifically the thalamus, and the hippocampus, and the entorhinal cortex. These things all interact with it, but our focus is primarily on the neocortex.

And I should remind people that one of the most amazing things about the neocortex is that, even though it's so big, if you could take it out of your head it would literally be a sheet about two and a half millimeters thick and about the size of a large dinner napkin. So, it's this large sheet.

And in everywhere you look in that sheet, there are regions that do different things—there are regions for vision, hearing, touch, language, and planning—but everywhere you look, there's this detailed microcircuitry. And that circuitry is remarkably the same everywhere—it's incredible. In fact, it's remarkably the same as a mouse neocortex and a cat neocortex.

And so, there's what they call a canonical circuitry. There is this circuitry that it seems exists everywhere—for all these different functions, for all these different animals—that seems to be repeated over and over again. So, to understand the neocortex, it's not like you have to understand each part separately; if you can understand what the core circuitry does anywhere, you basically can understand the whole thing.

So, it's very enticing from that point of view: *can we understand this complex circuitry that exists, literally in just two and a half millimeters of thickness of neural tissue?* And that's what I want to talk about today, because I think we've made some significant progress in understanding that.

Dr. Campbell: As I was reading your new paper, and also a couple of the papers from last year, it reminded me of some of the key ideas from *On Intelligence*—which I admit I haven't read in many years—but as I read your papers, I remembered that two of the ideas that seemed to be important were the role of prediction and the gap between how real neurons work and artificial intelligence approaches. And those two ideas came back to me as I was reading your papers. Do you feel that is a fair representation of some of what's going on?

Mr. Hawkins: Absolutely. It's actually quite good that you narrated on those two items, because those would be the two key ones, I would say.

Let me talk about prediction. I had this insight when I was a graduate student, back in the 80s actually, about the neocortex: your brain is making these predictions continuously about everything. And you're not aware of them; they're not conscious predictions. But your brain is predicting what it's going to feel when it touches things, it predicts what it's going to see when you move your eyes, it's going to predict what I'm going to say as I speak to you.

And this prediction is occurring at a very low level. The inputs coming into the brain are being predicted; they're saying, *okay, the brain has a model of the world*. And that was the genesis of the book, *On Intelligence*; it was about the brain as a prediction machine.

What has really changed is that, when I wrote that book I made the argument that understanding how the brain makes predictions is going to be key to understanding how it all works. What has happened in the last 10 years is we've made detailed progress about exactly how those predictions are made. And so, we sort of filled in the neural machinery, understanding this, as we went along.

We first asked *how can the brain make predictions about naturally-occurring changes in the world*. So, if someone is listening to this right now, they're not moving, and the patterns coming onto their ears are changing. So, that's like listening to a melody: how do you predict the next note in a melody, or what word I'm going to say, and things like that. And that's one type of prediction.

The other type of prediction is when the inputs to the brain change because you, yourself, are moving. Most of the reason that the inputs to the brain change, and most of the reason the brain is making predictions is that your eyes are constantly moving (you're moving them, your brain is moving them), or you're touching things by moving your hands over them, or you're walking forwards, and so on. And every time you move any part of your body, the inputs to the brain change, and the brain has to make a prediction about those, too.

So, there are these two classes of prediction: one is for naturally-changing things in the world—a bird flies by, or you're hearing a song—and the other is because you are moving. And we attacked those two problems, one after the other. And it turns out the same neural mechanism is used for both of them, with a very interesting twist for the one where you were moving.

That idea that prediction is important was the foundation of all of our research. It is, continually; we ask the question, *how does the brain do this?*

Now, when it comes to the neuron—I had this insight many years ago, as well—neurons are very complex; this is not my insight, but we know that neurons are quite complex. They have many thousands of synapses—the connections between cells. These connections are on the dendrites, which are complex processing elements themselves. And so, the neuron is a complex thing.

And almost all neural network models, still today, have a very, very simplistic and unrealistic idea, or model of what a neuron is. They have no explanation for all the thousands of synapses, they have no explanation for why they were distributed the way they are, there's no overall theory about how neurons really function in current neural network models—that is, used in like machine learning and things like that. There are, of course, neuroscientists who model neurons at a very detailed level.

But one[1] of the papers we published last year—the one that's called "Why Neurons Have Thousands of Synapses: A Theory of Sequence Memory in Neocortex"—we introduced a theory about why neurons look the way they do. And I can explain that, if you want. It has caught the attention of a few researchers who are really excited about it, and so, we're starting to do some collaborations about it, because it makes for some very interesting predictions.

So, the way to think about this is, most people think, *Oh, okay, a neuron has these inputs called synapses, and if you have enough of them active, then the neuron generates a spike, and then it projects to other neurons*. And that is, of course, true. But it's really only true for about 10% of the synapses on a neuron; the ones that are closest to the cell body. So, those are called proximal synapses.

And, so, when people think about a neuron—the simple neuron models—they think of it, *Oh, there's some number of synapses near the cell body, and you add up the input, it makes the cell fire*. But over 90% of the synapses are farther away from the cell body, and if you activate them, they don't really do much; they're not strong enough to make the cell become active and generate a spike.

Well, what the researchers have found out—and it's really fascinating—is that where the synapses are out on the dendrites (you can imagine they're like tree

---

[1] J Hawkins & Ahmad, S. "Why Neurons Have Thousands of Synapses, A Theory of Sequence Memory in Neocortex." *Frontiers in Neural Circuits Journal*•2016/03/30

branches), if you activate a small number of them (meaning some 15 or 20 of the synapses receive an input at the same time) and those synapses are close together (they have to be near each other, very close)—so, they receive an input close in time and they're also next to each other on the dendrite—the dendrite will generate its own spike; a different type of spike called a dendritic spike (most of them are called NMDA spikes due to the chemicals involved).

And that dendritic spike will travel to the cell body, but it, too, is not strong enough to make the cell fire. But what it is does is it raises the interior voltage of the cell significantly. So, it's like priming the cell to become active. It says, *I'm not going to make you active, but you're going to be ready; you're going to be primed to be active.*

And what happens—in our theory that's in that paper—is that, when a cell is in this state, it's in a predictive state. It's saying, *I think I'm likely to become active soon. I've seen a pattern out there that is usually predictive of me becoming active, and I'm going to be slightly depolarized—meaning I'm going to be more likely to fire than other cells nearby.*

And what happens now is, if an input comes in onto the proximal synapses—the ones that actually make the cells fire—I might have several cells that all respond to the same proximal input (they all have the same basic response properties), but one of them will be depolarized—its voltage will be raised internally—and that cell will fire first. It beats out the others.

So, I might have 10 cells that all respond to the same stimulus in the world, but one of them is going to be primed. So, one of them is going to be saying, *Ah, I'm expecting to become active.* And that one gets out its spike a little bit quicker and shuts down the other ones.

And so, what happens here is you can take an input—something like a visual line, or a sound, or a tactile touch on your finger—and it turns it into a very unique pattern in the brain that's related to the context at which it occurs. So, if I was listening to a note in a melody, instead of just representing the note in the brain, I would represent the note in the context of the previous notes. And so, I would have a unique representation of this note, or this interval, in the context of a song.

Anyway, this neuron model accounts for the thousands of synapses. We did the math behind it explaining how an individual neuron can recognize hundreds of unique patterns. They're very precise things. People don't realize that the neurons literally learn to recognize hundreds of unique patterns, and any one of those patterns could be predictive for the cell, and the cell will say, *Ah, that's a pattern that might indicate I'm going to become active next*.

And we illustrated the process of how inputs are modified in this way, and why the neurons look like this. And we showed that these depolarized states are really the predictive states of neurons. And this is why you're not aware of them, because it's an internal state to the neuron; it's not something you can perceive, it's an internal thing.

But we showed, also, that if the input comes in that wasn't expected, then the whole system can recognize that. And so, we notice when things are wrong, but we don't notice our predictions when they're right—at least, not most of them.

Dr. Campbell: So, in that paper, you add a section on testable predictions, and I'm just curious whether or not any of those testable predictions have been able to be tested yet?

Mr. Hawkins: Well, it's interesting, because some of these things can be tested, actually by just going through the literature. I mean, literally you can go back and find, *Oh, someone actually saw this*.

But the answer to your question is yes, some of these have been tested. We have been in collaborations with several labs who are excited about this theory. And there are people who have already contacted us saying, 'I've seen these results; I didn't have an explanation for them, but now I do.'

And we have other people who, hopefully starting the beginning of next year, are going to be designing experiments specifically to test these. So, we've had collaborations where people went back through old experiments and said, 'Yes, I can see some of these,' and then others where people said, 'I'm going to go further and test these.'

It is exciting, and there are almost no other theories about why neurons look like this—at a system level. We can explain what the neurons are doing, but we can also explain how 10,000 neurons work together to do something useful, using this. And so, it has generated a reasonable amount of interest.

Dr. Campbell: There were a couple of other ideas in the 2016 papers[2] that I thought we should touch on before we talk about your new paper: the ideas of sparseness of representation. Could you explain what that is and why it's important?

Mr. Hawkins: Yes, it is actually the key ingredient holding all this together. But it is kind of a conceptual idea, and some people have trouble with it. But I'll do my best.

---

[2] S Ahmad & Hawkins, J. "How do neurons operate on sparse distributed representations? A mathematical theory of sparsity, neurons and active dendrites." May 2016 arXiv:1601.00720 [q-bio.NC]

When we say ‘sparse,’ what we're meaning is that, if you look in the brain and you look at all the neurons (and there are billions of neurons in the neocortex), at any point in time, relatively few of them are active—anywhere between less than one percent or a few percent—and the vast majority are relatively inactive at any point in time. So, that's what the term ‘sparse’ means; it's a sparse activation.

And the question is, *Why is it like that*? That's very different, by the way, than computers. Computers use what we would call a ‘dense representation,’ meaning if I gave you a computer word, like 32 bits, you could have one that's all ones, or half ones, or quarter ones, or three-quarter ones—you know, ones and zeroes. But in the brain, it's always sparse.

And there are some really, really interesting properties, amazing properties, that come out of this. I'll try to touch on a few. It's a fascinating topic, but it's hard to get really into.

First of all, if you realize that if I have just a few thousand neurons (let's say I have 5,000 neurons) and I say two percent are active (so, that would be 100 neurons) and the rest are inactive, well, you could say, ‘How many different ways can I activate 100 neurons out of 5,000?’ And it turns out, it's an incredibly large number; it's more than the atoms in the universe. So, practically speaking, it's infinite. So, you can basically represent that scheme—this is just a property of having a large number of bits; a few thousand bits or several thousand neurons—there's an unlimited number of things you can represent.

Now, here's the next thing; it's very interesting: if I were to just randomly choose 100 neurons out of these 5,000, and then I just randomly choose another 100 neurons, and then I randomly choose another 100 neurons, and I do this all day long, almost none of the patterns I would choose would overlap much with any of the other patterns.

Yes, theoretically you could have two patterns that overlap by 99 neurons, but it's almost impossible. And what happens is you can pick patterns out of the 100 neurons at a time, and they'll almost all be, we might say orthogonal; they almost all overlap very little. That is, you would only have one or two neurons in common out of the 100, and all the others would be unique.

And so, this means I can choose patterns in this space all day long, and they'll all be very distinguishable; they'll all be very unique. I'm not going to worry about them looking similar. This is a very interesting property.

And then this tells us, if I want to recognize one these patterns—so, imagine I had a neuron, and this neuron wants to recognize the pattern of active cells—well, one way to recognize a pattern of active cells is to form connections to all hundred active neurons. So, there's 100 out 5,000, I would form a connection to all hundred, and if I see those hundred, I'd say my pattern is active.

Well, the math works out that you don't have to connect to hardly any; you can connect to maybe 10. And if you randomly choose 10 of the 100 neurons, or 15, or something like that, no matter how you choose them, if you see those 15, or those 10, active, you can almost be guaranteed that you found the pattern of 100; that is, it's very robust.

And so, when we think about neurons—if I'm a neuron and I have thousands of synapses, and as, a moment ago, I said I might only need 15 to be active at the same time to make the dendritic spike I talked about—well, I'm a neuron, I might be looking at a population of 50,000 cells someplace else, or 10,000 cells nearby, but if I want to recognize a pattern of activity in those cells, I only need to connect to 10 or 15 of them and I can be guaranteed that I'm going to detect the larger pattern.

So, a neuron can recognize hundreds of unique patterns in thousands of cells nearby and do so with a fairly small number of synapses. Which is surprising. But it's true, and the math works out, and it's very interesting.

Dr. Campbell: Would you explain the idea of the word 'robust' you used a minute ago, because it relates to this? I mean, you just described it, but would you explicitly explain it?

Mr. Hawkins: Sure. You can do it in various ways: One would be that, let's say there was some noise injected into the system so the neurons weren't very reliable—and neurons aren't very reliable. So, let's say five percent of the time neurons don't work. Let's say five percent of the time they just don't fire when they're supposed to fire; the system will work just fine, because I'm looking at 15 active cells maybe, and if five percent of them, a couple of them, don't fire, it makes no difference. The system continues working. So, it's robust to noise.

It's also robust to failure or decay. So, for example, if five percent of the neurons died and just never worked again, the system would still work—unlike a computer, but very much like brains, because we are talking about brains. And there's the point; the point is that brains are incredibly tolerant to noise, and dying cells, and things that don't work. The system is incredible. When we say 'robust,' it can't be exactly as good when those things happen, but it's almost as good; you could almost not detect it.

And so, in the paper we did in 2016, we did a lot of simulations to show this—to show how incredibly robust the system is to noise, and trauma, and death. The system just keeps working. At some point, if enough cells die, then the system starts falling apart.

But we as humans—this is a fact—after you turn 22, or something like that, your brain shrinks. There's cell death going on, and things are not quite as good as they used to be. But we survive pretty well—up until the point where you have really massive decay of some sort; then things start falling apart. But you can handle a lot of problems and not even notice it.

So, this is great for biological systems, of course, and it's great that our brains are built this way. But all of this comes about from the sparsity of activations. And it's a really cool way of computing. It's a very, very different way than computers do it.

Dr. Campbell: Is there anything else you think we need before we start talking about the new paper?

Mr. Hawkins: I only mentioned it since you brought it up. The old paper, the second part of the title was "The Theory of Sequence Memory." And so, that paper did detail a very biologically-precise, accurate mechanism by which like a layer of cells and then your cortex could learn sequences.

And the way it did that is that the cells connected to other cells nearby, and that was because the other cells nearby were a context, or the prior state, and so, it could learn, like, *Oh, given this state, I can now predict the next state; and given that state, I can predict the next state; and given that state, I can predict the next state.*

So, we showed how a layer of neurons can make a prediction about the next input in a robust, accurate way, based on just connecting to other neurons nearby. And —we think this is going on in several layers in the neocortex—we took that exact same mechanism, and then we applied it to the second problem I mentioned earlier, which is *how do we make predictions when we move*?

So, the only thing I wanted to say here is that the new paper starts with where we left off in the old paper and then adds to it and just says, *Hey, that same* *mechanism can work for learning how to predict things when you move*—what we call sensory motor inference. So, that would be just the connection between those two.

[music]

I love challenging myself in new ways and broadening my perspective. One of my favorite ways to do this is with The Great Courses Plus. There is unlimited access to thousands of lectures on everything from science, history, language, even playing chess, all presented by award-winning professors.

You can watch The Great Courses Plus from any TV, laptop, tablet, or smartphone. And now you can stream the audio, too, if you use The Great Courses Plus app. So, you can listen along as you go ahead with your day, with the flexibility to switch back to video whenever you want. I have tested this, and it works really well.

This month, I am recommending a wonderful course called Cognitive Behavioral Therapy Techniques for Retraining Your Brain, with Dr. Jason Satterfield. What I like about this particular course is that it gives you a practical toolbox for working on lots of the problems that we all face in our day-to-day life.

I know you are going to love The Great Courses Plus just as much as I do. You can sign up today and get a free month of unlimited access. To watch all of the videos, all you have to do is use my special URL, which is thegreatcoursesplus.com/ginger.

[music]

Dr. Campbell: The new paper[3] is called "A Theory of How Columns in the Neocortex Enable Learning the Structure of the World." Want to give us an overview?

Mr. Hawkins: Yes, sure. Funny, that wasn't our first choice for the title, but in the review process that was the last point we settled on. We wanted to parallel the first paper which was starting with a question, but for this time they didn't want us to do that.

So, the new paper, basically I already said what it's about; it's about how do we make predictions when we move? So, imagine you're touching something, and you move your fingers over it, and your brain is predicting what it's going to feel—if it's a familiar object. And we also believe that, when your eyes are moving, they're predicting this evidence that this is true—that your brain is predicting what it's going to see.

There are a lot of things in this new paper that I think are important, but there was one key discovery, which I want to start off with, which I think is going to unravel many of the mysteries of the neocortex. So, I talked earlier about, in the neocortex there's this common circuitry—'canonical circuitry,' sometimes I call it; you see it everywhere.

And let me just describe that a bit. In the two-and-a-half millimeter thickness of the cortex, there are about two dozen different cell types. Sometimes people say there are six layers, but that's not really true. There are really about two dozen different cell types—maybe more—that are unique and have different connections.

---

[3] J Hawkins, Ahmad, S & CuiA,Y. "Theory of How Columns in the Neocortex Enable Learning the Structure of the World" *Frontiers in Neural Circuits Journal*•2017/10/25

And so, there are all these different cells that are doing things. They tend to be in this laminar, or layered form; so, some are in the top of the two and a half millimeters, and some are in the middle, and some are in the bottom. And there is all this very complicated circuitry that has been detailed over decades of very difficult research, where people have been teasing about how these cells are connected, how these layers are connected, and which connections go this way and which connections go that way.

So, there is this tremendous literature on the circuitry of the neocortex, but there is very, very little theoretical understanding of what is going on; and most models are extremely simplistic. They say, *Oh, the input comes in to this region of the cortex or this area of the cortex, and it extracts a feature, and then it sends that feature to another region, and then that region builds more features and extracts it to another region.*

And it's sort of like, *Well, why do you have…?* In a square millimeter of neocortex, you have 100,000 neurons of 25 different cell types doing all this complicated processing, but what is it? No one really knows. And most of the neurons are sort of completely opaque. Why are the neurons here? We know where they're connected, but we don't know what they are.

We discovered (I say that carefully, because it's more like we deduced; we didn't go in and probe and find these neurons, we figured something out, and then we found evidence for it) that there's a second input to each cortical region, each cortical column—a column being like a millimeter square or half a millimeter square of the cortex.

We think of the input in the primary sensory regions as coming directly from the eyes, or from the ears, or from the skin. So, these senses go into these primary sensory regions in the neocortex and then they're processed. So, that's the primary input; everybody knows about that. But there is another signal that is generated in the cortex—which we are speculating in this new paper—and that other signal is surprising. And I'll try to describe what it is: it is a location; it is a representation of a location.

What do I mean by that? Imagine I'm looking at a cup; I have a coffee cup in front of me right now—there we go, I just put it on the table—and I can see it. And if my eyes are looking at it, there's an image on my retina. And the retina passes that image (in a very distorted form) to the brain—to the visual cortex. And then, we know that it extracts features, like, *Oh, there's an edge, and there's another edge,* or *there's a curve,* or something like that.

What we're saying is that, not only do we know that there's an edge, but every cortical column, or every region, is determining where that edge is relative to the object. It is saying *that is an edge* in a reference frame centered on the object itself. It's adding like a 3D element to it. It's saying, *Not only is it a feature that I'm detecting, but I know where that feature is on the object that is being sensed.*

And that location is independent of where the coffee cup is relative to me. That is, if I move the coffee cup farther away from me or I move it closer to me, the edge of the coffee cup is still the edge of the coffee cup; the location of the handle does not change.

Dr. Campbell: So, the location is relative to the object, not relative to me or you.

Mr. Hawkins: That's right. And the term neuroscientists use for this is sometimes called an allocentric location—'allo' meaning 'other'—so, it's the location centric to something else.

I'm sure you know this—and many of your listeners will know this, too—that there are other places in the brain where they've discovered location signals like this. And these are things called place cells and grid cells. These are very famous, and the people who discovered them got Nobel Prizes, and so on. And what they do is they represent where you and I are, relative to, say a room, or to some environment.

So, when I'm sitting in this room I'm in right now, there are cells that are encoding where I am right now. And if I move a little farther over to the next seat, those cells will update and say, *Oh, you're in a new location.* These things have been known for decades. And the grid cells are in the entorhinal cortex, and the place cells are in the hippocampus, and they've been found in other parts of the brain.

What we're proposing is something very similar is going on, but instead of locating my body relative to a room, which is what grid cells do, we're locating the sensory organs relative to the object that's being sensed. So, if I were touching my coffee cup with my finger, the tip of the finger is getting a sensation, but the sensory cortex—the somatosensory cortex, or the touch cortex—it knows both what it is feeling and where it is in the world.

And as soon as you add this location signal, then all kinds of things make sense, and all kinds of mysteries get resolved. And, all of a sudden, we can understand what all these layers are doing. And it tells us that the cortex, even a single column of the cortex, is much more powerful than people thought. It allows even a small section of the cortex to model complete objects and know the entire structure of objects.

So, in the paper, we develop this idea. And we talk about some of the networks, and how they work, and how you learn the structure of objects by movement.

Dr. Campbell: Right. And it was fascinating the way you showed that, even by modeling just one single column, the network could recognize the object. And then you had three columns, and the key difference was more columns meant you could recognize the object faster.

Mr. Hawkins: Yes, which is something we do, too. The theory says that to learn a new thing, whether you're touching it or looking at it, you have to move your senses; you can't learn something new without moving your eyes over it, or walking around it, or touching it in multiple locations.

Dr. Campbell: Which is the whole philosophy of embodied cognition.

Mr. Hawkins: Right. But what we're saying is happening when you're doing that, that there are thousands and thousands of these columns, and they are all basically learning models of the object. That is, there are many, many models of the coffee cup in my primary sensory cortex—well, we have to be careful, because there are some limits to what you can learn in each region—but the basic idea is that there are many, many models of the same thing being learned in each neighboring column.

So, I have to move over the object to train the system, but once I've trained the system, now I can show you an image, and you don't even have time to move your eyes and you can recognize the image. Or I can hold the coffee cup in my hand and I perceive the entire coffee cup, even though I'm only touching parts of it, but my perception is that the whole thing is there. Which is a fascinating thing to think about: you know, you're only touching parts of it, but you don't feel these funny sensations on your skin, you perceive the coffee cup.

So, what's going on is these individual columns, even though at any point in time they can only sense a part of the object, so, they don't know enough (my one finger, my index finger, is touching the rim of this coffee cup right now; on its own, that's not enough to know that it's a coffee cup), but the other fingers are touching other areas, and what they do is, they vote.

They are essentially sending these horizontal connections in the cortical regions—which are well-documented—those are like votes. It's like one column is saying, *Well, I'm feeling this edge here but I'm not sure exactly what it is, it could be this, this, or this*, and the other guy says, *I'm feeling this bump over here, I'm not sure what it could be, it could be this, or this, or this*. And, within a matter tens of milliseconds, they quickly settle on the only answer that makes sense for all of them.

And so, this is why you can do this flash inference. I can just flash an image in front of you, and you can say, 'Oh, I know what that is.' And it could be a very, very short, brief presentation. And it is why you can perceive things—you know, you have this perception of an entire object even though you're only touching parts of it. So, it explains a lot of things like that.

But yes, it totally flips around the way we think about the neocortex. Today's theory, what most people think about it is like each region in the cortex is extracting some small feature, and it sends it to the next region which extracts bigger features, which sends it to the next region which extracts bigger features, and somewhere up in the hierarchy of neocortical regions, you recognize the coffee cup.

Dr. Campbell: But that doesn't make sense, because it doesn't fit ... I mean, why are all those feedback signals existing? That doesn't fit.

Mr. Hawkins: None of this fits, Ginger; it kind of fits, but it doesn't fit. Look, almost all neuroscientists will tell you they don't really understand this.

It's not that there's something wrong, but I believe that once you understand that each column is determining this location—which is a tricky thing to do; it's not easy to do this, it's tricky; and a lot of the neural machinery in the brain is dedicated to doing this, we believe—but once you've done that, now it flips everything around. And you start thinking about the brain completely differently, you start thinking about regions differently, you start thinking about columns differently, you start thinking about the hierarchy differently. And it kind of comes together, and like, *Oh, my gosh, that's what's going on*; *it's not what we thought.*

But you're right, it didn't really make sense before. But there were no better ideas. So, you know, you've got to go with… until you know answer, it's not easy.

Dr. Campbell: I assume that you know the work of Olaf Sporns, even though he is working in a different area than you. I've interviewed him several times, and one of the things that he keeps saying is that we need theories; we can't just keep acquiring more and more neuroscientific facts without theories. And that's why we need people like you, who want to do the theory. I mean, because it doesn't seem like theory is what attracts people to neuroscience—as a general rule.

Mr. Hawkins: That's a very interesting question. I could talk at length about this. Remember I told you I started as a graduate student at Berkeley? That was in '86; I started in January of '86. And after two years, I left.

And the reason I left was because I was told—and it became obvious to me that this was true—that you couldn't be a theorist; that that was not acceptable, it was not considered proper science for neuroscientists. And I have speculations of how this came about and the history behind this—it's a fascinating question—and many, many people had recognized this problem.

When I started the Redwood Neuroscience Institute, I had a gentleman on our board who was very well connected in Washington, D.C. His name was Steve Zornetzer. He used to run the Office of Naval Research, and he was a neuroscientist. And so, we went and took a trip to Washington, D.C., and we visited the NIH, and the NSF, and DARPA. Those are the three government agencies that were funding neuroscience research.

And it was fascinating. When I went to the NIH—which is where the bulk of the funding for neuroscience research comes from—I met with about 20 or 30 of the program directors there, and I explained what I was doing. I said, 'I'm creating this new institute; it's a theory institute. We're not going to do experimental work, we're going to do just pure theory, and we're going to partner with experimentalists.'

And they were like, *This is great. This is what we need. We really need this, and so, we're so excited.* And I said, 'Look, I'm not even coming in asking for funding, I'm just letting you know and maybe we can work together.' And then they said, 'We're so glad you're doing this, because we can't do it.'

And I said, 'What do you mean you can't do it?' And they started telling me, *You can't believe, if anyone proposes a theory, then someone else shoots it down, and everything has to be by unanimous consent, and so, any kind of new theoretical ideas never get funded, and we can't do anything about it.* And they went on and on and on about this.

And so, it was a real eye-opener for me. Now, things have changed; it's better now. It's still not great, but it's definitely better now. There are a lot of people trying to do theory, and it's more accepted, and we are publishing theory articles. But even 15 years ago, that was really kind of considered out of the norm.

But I think it's true. I mean, neuroscience is a field with more data than theory. And it needs more theory. I view that as an opportunity; there's an opportunity to come in and make a big difference.

Dr. Campbell: So, what's next?

Mr. Hawkins: Given these insights we've had and that are described in the paper that just came out in *Frontiers* (I should mention, by the way, you can find these papers, and links to them, and videos, and explanations for them, and so on, on our company's website, numenta.com/papers), we see this as a tremendous opportunity to just make rapid progress on understanding what all the layers in the neocortex are doing.

It is like we're coming out of the woods. It was like almost nothing was understood, and now we're starting to understand a lot. And we have a long way to go, but I can tell you some of the kind of problems we're working on, but it's all in the context of, like, *okay, now we can start explaining what all these other cell types are doing and why the connections look the way they are.*

So, one of the big questions is actually how exactly is this location signal generated? How does a patch in V1—the primary visual cortex—or a patch in S1—the primary sensory cortex—how does that know where it is on some object in the world (that's a really weird question), and how would it know that? We think we know the answer to that question, but we're working through the details. And in our work, we want to make sure it matches to the biology. We don't want to just come up with theoretical, it's got to really be in the biology.

And one of the big clues was, fortunately, all that work that was done in grid cells over the last several decades. Because grid cells are amazing; these are the cells in the entorhinal cortex that tell you where you are in a room. And they have to solve some incredibly difficult problems. And we kind of know—not we, but the general community knows how they solve those problems.

So, we think the same mechanisms, the clever mechanisms that were evolved to do this in grid cells are being used in the neocortex. And so, we're trying to map those circuitries onto what are called the infragranular layers—the lower layers of the neocortex, layer V and VI—where we think this is occurring. So, that's one area of research.

Another area of research which is fascinating, too, I said earlier, and you pointed this out, Ginger, that a single column can basically learn entire objects. And we've shown that even a single column of maybe a half-a-millimeter square of cortex can learn hundreds of complex objects. Well, it's not just learning the shape of objects, we also learn how objects behave.

So, I have a pen in front of me, and the pen has a clip on it, and that clip bends up and down. And I know that, and I can push on it, and the clip slips down. And it has a cap that comes on and off. And I know how it behaves when I act upon it. That knowledge of how objects behave has to be stored with the objects themselves.

So, a pen is not just this thing that looks like this, and over here, somewhere else in the brain, I know what it does. No, I know what the pen does because the representation of the pen incorporates that. And so, we believe that even single columns in the cortex will encode all the behaviors that are associated with objects.

And this is a fascinating idea, and there is a lot of evidence to suggest it's true. And so, we're just beginning to think about how actually that is done: how could we encode that in the same column that's encoding where the definition of the shape and the morphology of an object is?

So, those are two big areas for us. Another big area is—and this is a little bit harder to describe—we didn't mention this in the paper, but we are able to share learning between our senses. I can learn an object by touching it. And now I could say, 'Ginger, here are four objects; I want you to reach in the box and feel them and learn what they feel like.' And then I could show you those four objects later, and not let you touch them, and you would be able to identify which was which.

Now, how does that happen? Well, that is something that is going on between... there is shared sort of learning between different regions of the cortex. And we have a weird term we use for that called 'disjoint pooling.' And it's hard to describe. But this idea that you can learn with one set of senses and recognize with another one actually turned out to be a very tricky and difficult problem. And we're working on a solution on that, too.

So, those are three areas, that I consider big problems, that we are currently occupied with: determining the location signal; understanding behaviors of objects and how they react; and then, how to do we learn across modalities.

Dr. Campbell: Well, Jeff, I have a question that I have taken to asking almost all of my guests, and that is, give advice for students.

Mr. Hawkins: Well, gosh, it's such a hard question to answer. There are several ways you can answer that question.

First of all, I think it is important to get as broad a neuroscience background as you can. Because if you start doing research right away and you work in someone's lab, you're just going to be so focused on one thing that you will not have a bigger picture.

I was fortunate in my life that, when I first got interested in this, I actually didn't start doing research right away, I avoided it because I didn't want to be an experimentalist. And so, I ended up spending two years basically reading everything. I read Kandel's big book, *Principles of Neural Science*. I read that twice, cover to cover.

And, it was like you got this background, because I find so often that neuroscientists don't know about other stuff. They don't know other things about the brain. They're just so focused. And having that broad background, that's one piece of advice: get that broad background. I don't care how you do it, but just go to classes, read books, learn a lot of stuff, study other animals.

The other piece of advice is (and this would be true for almost anybody in any kind of career) you're going to work for somebody in their lab—so, she or he is doing something, and you're going to be spending a lot of time as a graduate student doing post-doc work in someone else's lab—pick carefully.

You want to pick carefully: First of all, pick someone who's really going to be interested in teaching you, and advancing your career and not theirs. And second, pick a field that's an open field, one where it looks like no one knows what they're doing.

Dr. Campbell: Yes, where they don't think they will have all the answers already.

Mr. Hawkins: Exactly! Because so often what you hear is like, *You don't want to study that, because no one knows anything about it*. But, that's the beauty of it, right?

Because you can spend a lot of time (and a lot of neuroscientists do this) where you're working on refinements. And there's nothing wrong about this—refinements of ideas—but it's exciting to work on greenfield areas where there are things that people haven't done, and people don't know what's going on.

And there is lots of that in neuroscience. And people just say it's risky because you might not make progress, but that's where greatness comes from: you go and attack problems that people didn't think were solvable. And that's the only way to really make a big difference.

Dr. Campbell: Absolutely! Well, is there anything else you'd like to share that we've missed or is coming up for you?

Mr. Hawkins: I'm so happy to speak to you, Ginger, because I know you reach a good audience and an educated audience. I'm not doing this from a self-centered point of view. I really think we're making some significant progress in neocortical theory.

And if you're interested in that, I would suggest reading the two papers you mentioned and some of the other ones. If nothing else, there are some really interesting ideas in there. I think they're right, but even... So, some of them are wrong and there's some ... There are very few places you can read about ideas about the neocortex that really kind of deal with large-scale theoretical ideas.

And I'd love to have feedback, and people discussing these ideas, and saying what's right about them and what's wrong about them. I'm trying to elevate the conversation, too, in the general neuroscience community to discuss more theoretical ideas like this.

Dr. Campbell: Well, I'm glad we were able to talk.

I do have a question that is going to reveal my ignorance. I just, very recently, became aware of the idea of deep learning. And you made a reference in the paper to the difference between the HTM approach in deep learning. And I was wondering, is it just because deep learning uses the classical artificial intelligence model of neurons, or is there something else that separates the approaches?

Mr. Hawkins: So, deep learning is a model. It's a hierarchal processing model which was inspired by neuroscience; I mean, they're using artificial neurons in a hierarchy arrangement. And it's been extremely successful doing certain things in the AI world these days. So, it's a very hot topic right now. And all these engineers, and in fact, a lot of neuroscientists are leaving neuroscience to study it, because there's more money there.

So, there are two things I can say about it. First of all, it's extremely successful—so, good, we don't want to take anything away from that—for doing the things it does. It is not a biological model. It is very, very far from a biological model.

Not only are the neurons simplistic, they're completely wrong, and they have negative and positive synapses, they rely on very precise synaptic weights, they don't have any of the dendritic processing, there's no complexity that we see in the brain. It's a very, very overly simplistic interpretation, so it doesn't really inform us much.

And the way they train is biologically impossible. And they don't work like brains. When they get to train a deep-learning network, you have train it on millions of images; and in humans, we don't do that. I can pick something up and look at it for a few seconds, and I get it. So, they're really very different. So, they don't really inform neuroscience, and neuroscience, today, is not informing them. So, that's one issue.

The other issue is they're starting find real limits to them, that there are some fundamental problems with them that ... And I don't have to say this, because the leaders of deep learning are saying this themselves. Geoff Hinton, who is one of the founders of this whole field, has been saying lately, we have to start over from scratch. We have people like Demis Hassabis, who is the founder of DeepMind, the big Google research group now, and he, too, is saying, 'We're reaching our limits. We have to go look at the brain; the brain is going to tell us how to do these things.'

So, that community is starting to realize that they need more inspiration from the brain, because they're running up against limits. What they're doing is not a theory of intelligence, it was not built on a theory of intelligence, it's a clever pattern-recognition technique, but it's not deep.

So, I think these two fields are going to come together. And it's starting to happen. And I think the insights we've had have already started to influence that field. For example, we have been working with this group over in Europe who are building these neuromorphic hardware chips. These are chips that accelerate artificial neurons. And they had very simple neurons in the past, and now they're adding active dendrites. And so, this is based partly, or largely on our work in the paper that we wrote last year.

And so, there's this merging, and people starting to say, 'Oh, yes, we need these things.' So, I think this is going to happen. And I think people who are really interested in getting beyond the limits of today's AI technology—deep learning—will find our papers fascinating, because I think they lay out a road map for how we're going to get there.

Even if you don't emulate the brain *per se,* this idea of this location signal we've been talking about, I am totally convinced that the future AI machines are going to be built on this principle. It is so important; it is so fundamental. They're going to be built on sparsity, which you asked about earlier, and they're going to be built on neuron models that have this complexity we've talked about.

So, I'm very confident that it's going to happen. And I'm excited about it. I think this is a way the engineering world of machine intelligence is going to be interacting and becoming an important part of neuroscience and vice versa.

Dr. Campbell: Well, I'm fascinated to see what's going to happen in the future. And I hope that we don't go nine years between interviews next time.

Mr. Hawkins: Yes. You know, last year when our paper came out, I said, 'This is something Ginger might be really interested in;' but I wanted to wait until this year's paper, because I really wanted to have something important to talk about with you. And I feel we do.

And I think the progress is going to be much faster. I think we're going to really enter an accelerated period of cortical theory right now. And so, we shouldn't have to wait that long; we'll talk again in a year. I'll send you our papers next year, and we'll see how much progress we've made. Okay?

Dr. Campbell: That sounds like a plan.

[music]

It was great to talk to Jeff Hawkins again. I hope you got a sense of his passion for neuroscience.

Before I review a few key ideas, I want to mention that all the papers we talked about are freely available on the Numenta website, and I'm going to be putting links to these in the show notes at brainsciencepodcast.com.

The main reason that I continue to follow the work that Hawkins is doing at Numenta is that his team is actually trying to build a model of cortical function that matches the brain's anatomy and physiology, which, of course, also means generating testable hypotheses.

In 2016, they published a new model that incorporates the fact that not all synapses are identical; in particular, their function seems to change depending on where they are located on the dendrites. Only synapses on dendrites near the cell body can cause an action potential, but in this new model, Hawkins proposes that the other synapses prepare the cell to fire, which, in a sense, is a form of prediction.

The other paper that they published in 2016 focused on how sparseness leads to robustness. As Hawkins explained, having a sparse representation reduces overlap, which means that the signal is robust even in the noisy environment of real neurons.

Then this year, 2017, they expanded their model to represent a very simple column. This is a key step toward making the model realistic, because we know that the real cortex has a columnar structure, but exactly how the columns work is not yet known.

Hawkins proposes that the key information that is needed for a column to identify an object is positional, but the positional information is allocentric, which means it's based on the object itself, not on the person or animal doing the sensing. For example, if you are touching a coffee cup, the location of the handle relative to the cup helps you recognize what it is.

The other key idea is that, although a single column can theoretically recognize hundreds of objects, having additional columns allows recognition to occur more rapidly; just like you can recognize a coffee cup more quickly if you use several fingers compared to just one finger.

These ideas are actually easier to visualize with the help of a video, and I'm going to provide the link to this video in the show notes.

In 2016, their model included active dendrites and demonstrated how this can allow a single neuron to learn and remember a large number of objects. The model also included the importance of sparse representations, which improve robustness of learning and recognition in noisy environments.

In 2017, the model is now expanded to represent a simple column and demonstrate how object recognition could occur by combining feature recognition with a location signal. This appears to be an important step toward creating a model that more accurately reflects cortical anatomy and function.

We did talk briefly about how Hawkins' and Numenta's work differs from traditional AI approaches and from newer approaches such as deep learning. He emphasized again that the key difference is Numenta's commitment to a truly brain-based model. I will talk some more about this next month when I share some highlights from this year's Society for Neuroscience meeting, which I attended in Washington, D.C. One of the talks I went to was by the founder of the DeepMind Project.

I want to start my closing announcements with some advice for new listeners. If you're like me, when you find a new show, you like to go back and listen to earlier episodes. Unfortunately, the current version of the Apple podcast app makes this rather difficult. One easy way to get all the episodes of *Brain Science* is via the free *Brain Science* mobile app, which is available for IOS, Android, and Windows phones. You can access over 50 episodes free, and I will explain how to get all the episodes shortly.

Last month, I promised to share some highlights from my trip to the Society for Neuroscience meeting in Washington, D.C. Obviously, I decided to air Jeff's interview instead. So, next month, which will be our 11th Annual Review Episode, I will include some highlights from Neuroscience 2017 into that Review Episode.

I would also like to thank everyone who reached out to me at SFN and apologize to anyone that I didn't get to meet. I know there was at least one person whose email I could not seem to find once I got to D.C., and we never were able to find each other.

On the other hand, I really want to thank Rebecca Resnik and her husband, Philip, for taking me to dinner at Busboys and Poets. We had a great meal, and I highly recommend the restaurant. I also want to mention Rebecca's book, *A Family's First Guide to ADHD*.

Next, as I mentioned last month, I'm trying to plan a trip to Australia in 2018, and I've decided it would be fun to do this with a group of *Brain Science* fans. If you are interested, please write to me at brainsciencepodcast@gmail.com. I'm not going to post this on the website, because I really want to limit the group to regular listeners who actually take the time to listen all the way to the announcements at the end. The group will be limited to 20 people, so if you're interested, don't delay.

Over the last few weeks, some Premium subscribers have had problems accessing episode transcripts. This problem has been resolved, but it made me aware that quite a few people are confused about how to get these transcripts. If you're on the main *Brain Science* podcast website and you're looking at show notes for a particular episode, you will see a link that allows you to buy an individual transcript. Unfortunately, you can't get to the Premium transcript on the main website, because the Premium content is hosted by libsyn.com. Premium subscribers have unlimited access to older episodes, as well as access to all transcripts for only $5 a month.

There are two ways to get this content. One is via the free *Brain Science* mobile app, and the second, if you're using your web browser, is to go to brainsciencepodcast.libsyn.com. If you're on the *Brain Science* podcast website, brainsciencepodcast.com, then look for login buttons that will take you directly to the My Libsyn login page.

I want to thank everyone who helps support my work. *Brain Science* is independently produced, and it relies on your support. The Premium subscription is very easy and a great option for new listeners who want to access the episodes posted before the most recent 50, which, right now, means those before 2013.

However, there are also two other ways you can support the show. Patreon allows you to pick a monthly amount, and I do make episode transcripts available on Patreon. That's patreon.com/docartemis. Another option is just to do a direct donation via PayPal or even by check. To learn about all these options, go to brainsciencepodcast.com/donations.

Finally, don't forget to check out our new sponsor, thegreatcoursesplus.com/ginger.

But, as always, I want to remind you that even if you can't afford to support the show financially, you can help by sharing it with others. And don't forget to subscribe in iTunes or in the Apple podcast app, even if you listen in a different app. The reason for this is that it helps us to go up in the rankings and helps new listeners find the show.

I would really love to hear from you. You can write to me at brainsciencepodcast@gmail.com. There's a *Brain Science* podcast Facebook page. I am DocArtemis on Twitter. And you can leave audio feedback at speakpipe.com/docartemis.

Thanks again for listening. I will be back with you next month.

[music]

[music]