Episode 6: Interview with a Neuroscientist – Dr. Blake Richards

Blake:	00:00	And actually that's a really good way of phrasing it because I think what's so fascinating about human learning and this is really what marks not just human, but I would say broadly sort of generalist species is that we seem to be able to actually define our own cost functions.
Matt:	00:23	You just heard a little sound bite from my interview today with Blake Richards. Thank you Paul Middlebrooks for giving me that idea. I've been watching his podcasts called Brain inspired. If you like our podcast, give him a shot at braininspired.co. Welcome to the show. This is the Numenta on Intelligence podcast and today we're going to have another Interview with a Neuroscientist. So stay tuned and we will get right into it.
Matt:	00:51	All right welcome to another episode of Interview with a Neuroscientist. I'm Matt Taylor with Numenta and today I'm really excited to have Dr. Blake Richards here with us. He's an associate fellow at the Canadian Institute for Advanced Research Hello Blake.
Blake:	01:06	Hi Matt.
Matt:	01:07	Great to have you here. I've been following your work for a while and I'm interested in the ideas you are bringing to the field as an observer of like Twitter and the neuroscience community for the past couple of years. I feel like you're part of this sort of new, new wave of neuroscientists coming up with some new ideas and not just about the science but also about

processes and protocols. How do you think the field is changing right now?

Blake:	01:33	Yeah that's a good question because it definitely feels like it's changing and it's not always easy to put one finger on exactly what is changing. I think the way that I would articulate what's happening right now is that we are actually seeing neuroscience more, at least parts of neuroscience, morph into something that's almost more akin to what cognitive science was back in the day. That is a truly interdisciplinary field of research that incorporates not only the components of biology that are relevant to understanding how brain cells communicate with one another but also components of computer science and philosophy and psychology in order to try to get a grasp of what we might call sort of general principles of intelligence and general principles of behavior that are important for understanding the ways in which any agent whether an animal or in fact an artificial agent works. And that's quite different from what neuroscience was when I started as a Ph.D. student you know a little over a decade ago where it was really more kind of the Sub-Branch of biology and with a bit of psychology thrown in occasionally.
Matt:	03:00	So it definitely is broadening a lot it seems. It's one point. And you think that's because like to understand the general principles of the brain you have to you have to think broader than just the molecular biology level, right?.
Blake:	03:13	That's right exactly. I think that's part of it. And I think it's also a result of the realization more broadly in biology altogether that it's - biological systems are so complex and their operations are so non-trivial. You really have to bring to bear any tool that you can to understand them. And it's not really viable to simply -- look, I think with the practice was in neuroscience for many years and what some people still do to some extent is what I call you know neuro stamp collecting where you basically just try to get as many facts about the brain and its operations on a biological level as possible. And there this this hope that you know the more facts we accumulate at some point we have something like an understanding of the brain. But you know Daniel Wolpert a researcher who lives at Cambridge I think he's moved to Columbia now. He had a great bit about this that he gives in his talks sometimes so there's a very famous neuroscience textbook by Kandel and a few others called Principles of Neural Science. And it's the textbook that many of us receive when we first start in the field and Principles of Neural Science - Daniel Wolpert has this plot where he shows the number of pages of the Principles of Neural Science keeps

increasing every year after year according to linear function and he, he points out that like if we were actually uncovering Principles of Neural Science, presumably the book wouldn't have to keep growing and growing because all it is at this point in time is an accumulation of potentially unrelated facts about the brain. So what people are starting to desire and why we're seeing this shift towards broader ways of thinking about the brain is something more like true principles and the way that Daniel Wolpert puts it is you know, we know we'll be successful in neuroscience in the coming decades if we can start to actually shrink the number of pages in the Principles of Neural Science textbook.

Matt: 05:32   Right, that make sense. When I'm reading a neuroscience paper, because I sometimes read neuroscience papers as a drive to try and understand all of the real biology behind the theory, and there's so many ways you can just go down rabbit holes and be lost forever, you know you can spend your whole career studying this one particular aspect of intelligence.

Blake: 05:52   That's right.

Matt: 05:53   It's amazing.

Blake: 05:54   Yes. Exactly and that's what many people have done in the past and historically you'd kind of pick your specialization and your particular circuit and you would study the hell out of it. So you would be the expert on you know the synaptic physiology of the Shaefer collaterals in the hippocampus or something like that. And that you know, that made sense in some ways in terms of like that was a good way to like, I think the impulse behind it was a good one. The idea being that you really want to fully understand the systems and you know these are complicated systems so why not take decades to study this one little circuit. But yeah if, if you don't actually end up bringing that to to unite with other things that we're learning about the brain and with broader principles that we might derive from artificial intelligence or psychology, then you know, how can you actually say that you've gained an understanding of the brain beyond just the stamp collecting as I say

Matt: 07:02   Right. We've got to put the thing, the facts together in some cohesive story about how it all works. All the things work. And that's in some ways of saying you know at the end. It involves imagination involves theorizing.

Blake: 07:17   That's right. Exactly. And I think it's something which many neuroscientists are uncomfortable with and it's why sometimes

we see some pushback against this slightly new direction in neuroscience because some people are uncomfortable with the idea that we are going to, basically, because part of what is required to develop the kind of cohesive broader picture of how the brain is working is occasionally not incorporating certain biological facts into the ways that you're thinking about something because there's just too many to to wrap your head around trying to make it all work. And I think that makes some people uncomfortable because it means that occasionally we're ignoring some components of the biology that we know exist even though, you know, we know it's true we were kind of like, well we're not going to think about that for our broader model right now. And that's something not everyone is comfortable with.

Matt:	08:13	Maybe we can explain it. We know something like this is happening and we might know why it needs to happen but not how.
Blake:	08:20	Right. Yes.
Matt:	08:22	So I'm afraid of getting too deep here. But you're a Doctor of Philosophy so, why not? I like to talk about reality, especially how it applies to artificial intelligence as you know the world perceives A.I. right now.
Blake:	08:38	Yeah.
Matt:	08:38	And so I love this idea that Max Tegmark introduced me to, this external reality that just exists. It's sort of like the ground truth. It's what's out there and all of us intelligent beings have an internal reality which is really just a model based on our experience with reality what we think it's like. And they're all sort of wrong and distorted. You know, it's just our it's our sensory perception over time of what we think is out there and in order for us to communicate with each other we have to establish sort of a consensus a reality where we can share ideas and we can say red and you know what I mean and I can say two plus two equals four and we know what that means. You know this sort of accumulated knowledge is in this consensus reality. And when you talk about AI, I mean if we're going to create intelligence sort of in our image if we're if we're trying to learn how the brain works and we think we can turn around and reverse engineer it and create something like that. It goes against this idea that some people want to make explainable AI. They want to know you know exactly why AI made a decision and it always bothers me because from that from the perspective of biology we can't do that with biology. So how can

		we expect to do that with, you know, machine intelligence in the same way?
Blake:	09:54	Quite. Yes I agree. That's a really good point and I think this the complaint that current Deep Learning Systems in AI are interpretable or unexplainable is certainly a funny one whenever it comes from neuroscientists because I am personally completely convinced that the brain is probably equally uninterpretable and unexplainable. Certainly you know I think Conrad Kording a neuroscientist at UPenn articulates this well you know when you when you actually go looking for you know oh does the brain respond to this stimulus does the brain respond to this stimulus, etc. Basically you can find almost anything you want in almost any brain region if you look hard enough and interpreting that is almost impossible and arguably the only way to interpret it is to come back to principles of optimization in the same way that you know we can. You know it's always happening when people say that we can't understand deep nets. We do understand them. We understand that they're optimizing on particular loss functions. We understand learning algorithms that enable them to optimize in that way. And so we can say very clearly why they developed the representations they developed. We just can't articulate exactly what their solution is to the problem in human readable format and it's entirely possible that the brain is the same way either as a result of the evolutionary optimization or learning during an individual's lifetime the specific wiring of our neural circuits that lets us do the things that we do, may or may not be human interpretable and there's no reason to expect that it would be really. So why would we expect the same for deep neural networks?
Matt:	11:48	Something Jonathan Michael said - he was on the program awhile back and I asked him what is it to grab a cup because he studies motor commands and monkeys and he's like What is that. Is that motion to grab a representation of grabbing a cup. How do you come up with that? His answer is basically you could bring together every time you grabbed a cup and every joint experience you've ever had in your entire life and that's what it is. How do you convey that to another person? That's sort of the level of information we're trying to capture.
Blake:	12:15	Right. Quite right. Yeah. And I think that you know the difficulty with all this stuff is that there aren't actually simple need to verbalize ways of you know describing what it is to pick up a cup or what it is to successfully navigate somewhere or what it is to suggest we perceive an object, you know - very, very abstract mathematical descriptions that we can give but that's not what

many people who are complaining about the lack of controllability are looking for. What they want is a simple few sentence description of what's going on and that just might not exist.

Matt:		Maybe it will exist in a consensus reality that we create with these intelligent systems over time.
Blake:		Yeah possibly. And so I think that what's interesting, what you are you're saying that way which is interesting is that arguably you know part of what happens with human beings is that we make some of our actions interpretable quote unquote by virtue of the stories that we tell each other about why we did something or other. Right?
Matt:		Right.
Blake:		And I think often the funny thing is if these things are false One of the things that we know that there's some evidence for research-wise is that you know we will kind of generate post hoc explanations for our actions even though the experimentalist knows that they've manipulated you in such and such a way. And the fact is that I suspect that's happening constantly. I think that you know we are often engaged in various behaviors, the ultimate reasons for why we do the things we do might be almost completely unexplainable. But we tell each other these post stories and then that becomes our shared reality. So you know I went to the store and whatever bought some ice cream because I was stress eating quote unquote like the exact computations behind that are surely far more far less interpretable rather than I was stress eating.
Matt:		Which is an interesting segue into an essay I wanted to talk about which was cost functions or loss functions. I don't know why stress eating, but you know, that's sort of that's the feeling in need in your brain somewhere.
Blake:		Yes quite.
Matt:		We've had discussions on Twitter about this but I think there's some in my audience that may not be familiar with that term. Could you maybe give it a 30000 foot definition what is a loss function.
Blake:		So a loss function is just a way of quantifying learning. So when we talk about learning, necessarily learning implies some kind of normative improvement. Right. If you are learning you're

getting better at something and if you want to quantify you're getting better at something then you need to identify some number, some function. That is a measurement of how good you currently are at whatever it is you're trying to learn. And the word we use in machine learning to describe these functions are loss functions or cost functions. And so then learning can be defined as anything which reduces a loss function.

Matt:	15:49	Right. So I have a background of software engineering so I think I can think of this as a function that takes input and gives an output. So in this sample what would a sample of that input be and the out the output would be you know how good it is right?
Blake:	16:04	That's right. Exactly. So the input would be the current setting for the agent. So in the case of a neural network it would be the current synaptic weight for the agent and the output is this measurement of how good it is now.
Matt:	16:23	Now can we abstract that even further? I like to think about video games. Obviously I was playing a lot of video games. If you think about a loss a cost function for Pong like for an AI player. Could I think of that as like all the input being the location of the ball as it moves and then the loss function judging how well the panel whether the paddle prevents the ball from going past it or not?
Blake:	16:50	Roughly but I think the way we would probably approach it in an actual like AI system is one step more abstraction. So the input would be the current policy as it were. That would work. That is to say the current set of actions that you would select as a pong player based upon the screen that you provided.
Matt:	17:16	Oh, so like all possible things you could you might do
Blake:	17:20	Exactly. All possible things you might do in response to all possible inputs and then the output would be a measurement of the average score that you would get in the game. And so in this case it's it's what we'd call rather than a loss function it's the inverse of our loss function. We want to increase our score so you want to see that improve over time.
Matt:	17:50	It's like an optimization function or something.
Blake:	17:52	That's right - an optimization function - precisely right.

Matt:	17:56	So more - again I keep thinking about video games. Could I also think of this in terms of behavior? Like if I'm playing Mario Kart or Pole Position depending on how old you are and I'm controlling a car can I even define that environment with a loss function? If I'm if I want to say I want to stay on the road I want to go around the track as fast as possible and I don't want to hit things. Is that - is that - is this working in that scene too?
Blake:	18:21	Yup exactly. So again you know the way that we approach it in machine learning is in this very sort of high level you where you say Okay so for all possible situations in this car game, what actions would you take at this point in time? And then you would get some score based upon that such that your score would go down if you ever drove off the road and it would go up for you know how rapidly you were able to go around the track or whatever. And that is your loss function then.
Matt:	18:58	In your brain though, I can imagine evolution provides loss functions over a long period of time you know like behaviors that expose themselves in order to help the animals survive right. Those are coded in genes and those are going to be stored, well I mean they're going to be expressed in older parts of the brain is that right?
Blake:	19:17	Well so when we talk about the loss functions that govern evolution, what's interesting there is effectively what we're talking about- the central loss function for evolution of course is the likelihood that your genes will propagate to the next generation and the input to that loss function. So that's the output of the loss function is what's the likelihood that your genes will propagate to the next generation. The input to that loss function is effectively your current physiological state and evolution is about shaping your physiology in order to maximize the probability that you're going to propagate your genes to the next generation. So that specific loss function itself isn't encoded in your DNA but your DNA has ultimately been shaped by this process of optimization on this loss function over time.
Matt:	20:17	So the example I'm thinking of, trying not to be crude, but all biological systems have to excrete waste and there is behavior in animals to excrete waste where you're not collecting food. That's something I think - Is that something that is at those low levels of the brain or is that something that you think is learned?
Blake:	20:38	Right. Well so OK so then we when we start talking about the intersection between the sort of learning that is evolution

because you can view evolution as a type of learning because it is this optimization.

| Matt: | 20:54 | A a very slow type of learning. |

| Blake: | 20:56 | That's right. And a type of learning that doesn't occur in an individual but instead occurs in a population. |

| Matt: | 21:02 | Exactly. |

| Blake: | 21:03 | So evolution is this very slow learning that occurs over a population and then within all of our brains we also have learning algorithms that help us as individuals to learn. And what I think is interesting is that part of what has probably happened over the course of evolution is that one of the things that that came out of our evolution was that it was beneficial from the evolutionary cost function for our brains to also optimize on some other cost functions. And sometimes you know our behaviors can seem a little bit weird. With respect to our survival. Because even though it might have been beneficial in the long run for us to be optimizing on these other cost functions, internally at the end of the day they might not always agree with the evolutionary cost function. And so the example I always give that way is with drug addiction. So in all likelihood we think that you know the brain seems to have a cost function that is some kind of reward maximization cost function. Right? You as an animal are going to do stuff that helps you to maximize the probability of obtaining rewards and the difficulty then of course is that if you take something that's very basically intrinsically rewarding like heroin, that cost function might go into a new behavior to just do whatever you can to get as much heroin as possible even though that's not beneficial for the evolutionary cost function of you propagating the genes to the next generation. |

| Matt: | 23:00 | It's sort of like shoring a circuit. |

| Blake: | 23:01 | Yeah that's right exactly a sort of short circuit. Exactly. And you know that's not to say that you didn't evolve that reward maximization cost function for evolutionary purposes. Because the African savanna that was probably a pretty good core function to be optimizing on but for a modern human maybe not so much. |

| Matt: | 23:27 | There are certainly examples of us humans enhancing our evolved cost functions for instance if we're using example you know don't excrete where you eat, at some point we decided it |

would be good if we started washing that process which increased our lifespan considerably. We learned that behavior. I mean it's almost like these cost functions, once they emerge, they're memes. They turn into memes.

Blake: 23:56 Yes right. Right. And actually that's a really good with phrasing it because I think what's so fascinating about human learning and this is really what marks not just human but I would say broadly sort of generalist species is that we seem to be able to actually define our own cost functions. So you know for example you know some people will just get obsessed about getting really good at that particular random tasks right like they will they will decide that they really want to be an expert on, I don't know, different types of lager or something like that and it's not immediately clear what cost function they're optimizing on besides this arbitrary one that being able to distinguish different types of lager. But they do it, right? And so we seem to have this ability to define our own cost functions in a way that makes us incredibly flexible as an animal and which again can sometimes seem to go against our evolution but probably in its origins was beneficial for our ancestors somehow.

Matt: 25:10 We're pretty much making it up as we go along at this point-defining our own functions, doing whatever we want. I mean performance art is a beautiful thing to behold when it is done right. And it's a cost function. And if it's appealing to the general public they get accolades for it. I mean they're basically defining beauty with one of these cost functions. It's amazing.

Blake: 25:35 Yes that's right. And yeah. And so actually to come back to your memetics point I suppose I got off track. What I think is interesting about your point that way is that we also and this ties back to your point about shared reality, arguably what happens in human society is that we develop joint shared cost functions. So you know we all decide that what we really want is, you know, whatever like particular B.C. house music or particular as you say like performance art with certain characteristics that are hard to find

Matt: 26:12 Certain type of politics or whatever.

Blake: 26:14 Yes that's right exactly. And so that then becomes the thing that we're all optimizing on because we were obsessed with these sorts of shared memetic goals that we develop.

Matt: 26:28 Wow. All right. I didn't know how they were going to go but we were pretty deep. That's awesome. OK. Well let's talk about deep learning. We haven't touched on really a whole lot yet.

You've done a lot of work and learning. My audience may not be the most proficient in the subject. I think, I think you know the HTM audience is more towards the neuroscience than the hobbyists and the engineers. So maybe you could talk about back propagation in a simple term. Can you define back propagation for us and why doesn't it work biologically. Because that's one question it'd be great to explain.

Blake: 27:03 Sure yeah. Ok so I'll start by just defining deep learning. So deep learning is a particular approach in machine learning that has two basic tenets. The first is that you should try to have minimal intervention from the programmer meaning you should hardwire as little as possible and have the system learn as much as possible. So this is in contrast to more traditional approaches to artificial intelligence which are sometimes referred to as good old fashioned AI or GOFAI

Matt: 27:47 Like expert systems or very finely tuned applications.

Blake: 27:51 That's right. That's right. Where you as the programmer say OK Computer here's the way I want you to act, here's the logical chain of reasoning that I want you to engage in, here's your understanding of the world as programmed by me. Go behave intelligently please. The deep learning philosophy says no you as the programmer should do as little hardwiring as possible and you should basically just focus on the development of learning algorithms that allow your agent to use the data that you provide it to figure out for itself how to behave exactly.

Matt: 28:30 A noble endeavor for sure yeah.

Blake: 28:33 So then the second tenet of deep learning which distinguishes it from quote unquote shallow learning is the idea that what you want to do is not only to learn as much as possible but to also have a what we call a hierarchical system where you process data in a series of stages or modules and you also ensure that your learning algorithm is adopting every one of those stages. So the analogy that deep learning people were ultimately building off of that they were ultimately inspired by was how our own brains work. So even though it's an oversimplification of what goes on in our brains to some extent you can say that when a data arrives at our retina it then gets processed by a series of stages where each stage of the processing in our brains identify ever more complex kind of abstract features of the image that we're looking at. So in the early stages of processing your brain identifies various lines and edges it then assembles that into an understanding of various joints and shapes and then that gets fed into areas that identify more abstract object

categories etc. etc. And so the deep learning approach was inspired by this and said we're going to have that same kind of like multiple stages of processing and per the first part of the philosophy, we're going to learn every part of that and that's what distinguished deep learning from some of the other quote unquote shallow approaches that were popular at the time that deep learning really took off such as support vector machines and chrome machines and related stuff. Those systems would have multiple stages of processing but typically only the final stage of processing was where any learning occurred and all the early stages of processing were hardwired by the programmer or followed some pre-determined mathematical formula and only the final stage was learned.

Matt:	30:57	Sort of a mash up of the old way and the new way.
Blake:	31:00	That's right. Yeah. So really what distinguished the learning was we're going to have these hierarchical processing stages and we're going to learn it all
Matt:	31:09	Right. So what is the back propagation? Is that the learning it all part?
Blake:	31:14	You got it. So the back propagation of error algorithm is a learning algorithm which provides you with to date probably the best guarantee that anyone's ever been able to develop that every stage of your processing hierarchy is going to be optimized based on a given cost function.
Matt:	31:38	So that's just not biologically feasible right? There just couldn't possibly be that many connections, is that the argument?
Blake:	31:45	Well no actually so in fact a big part of my research is that I believe that the brain also does this. I believe strongly that our brains optimize every stage of the hierarchy and they do so in a way that guarantees that the cost functions that we're optimizing are reduced by virtue of the changes that happen in every part of our brains where we say that back propagation is biologically infeasible is that back propagation is really just a specific way of implementing something known as gradient descent. So gradient descent is the following idea. Let's say, so we've got this cost function that we've discussed where the input is the current state of our system and the output is a measure of how well we're doing according to our learning goal.
Matt:	32:42	So the complete state, all of the neurons in the system.

Blake:	32:46	All of the neurons and all of the synapses. That's right. So we get we take all the neurons and all the synapses, that feeds into our loss function. We get out this number that measures how well we're doing on our learning task. The gradient descent approach says OK the way we're going to learn this system is we're going to try to estimate the slope of our loss function. So you can think of it if you can in kind of abstract terms: think of the loss function as representing the height of a hill and your position in kind of you know GPS coordinates as representing the current state of your network that you're feeding into your loss function.
Matt:	33:39	This is, of course is a very high dimensional thing.
Blake:	33:42	Very high dimensional. That's right. So you're not moving in a two dimensional space which is what you're doing when you're looking at GPS coordinates but instead you are moving in a say 10 million dimensional space. So you've got this hill effectively in a 10 million dimensional space and in the same way like let's say you're let's say you were a blind person who was trying to descend a hill, how could you do it? Well one potential way of doing it would be to basically try to figure out just by feeling around a bit which direction the slope was going and always just walk downhill.
Matt:	34:24	So local analysis sort of.
Blake:	34:26	Yes exactly a local analysis. So if you always just look at the slope to where you are and you go downhill eventually you're guaranteed to converge to a local minima. At some point you're going to reach something that is the bottom of some valley. Now it might not be the absolute bottom of the hill but you're guaranteed to approach a local minima anyway.
Matt:	34:52	That's a great explanation by the way.
Blake:	34:56	Good. Now what's interesting is that this is the reason that gradient descent is such a powerful approach is that if you consider it you know two dimensional or three dimensional landscape it's very easy to get trapped in local minima that are very far from the global minima and it's something that concerned many people when gradient descent approaches were first developed in artificial intelligence. Now imagine doing the same you know hill descending in a ten million dimensional environment. In order for something to be a true local minima you have to have it be a minima in ten million directions and the probability of that happening is actually relatively low. So people have done analyses to show that in fact what's

interesting about gradient descent is that the higher the number of neurons you have the more synapses you have, the less likely it is that you're going to get trapped in local minima and thus the better it is to do gradient descent. So what we've kind of discovered in AI is that in fact these gradient descent algorithms work better, the larger the system the more we scale it up.

| Matt: | 36:15 | These things are so high dimensional it seems like you never really settle on anything if it's a dynamic system. |

| Blake: | 36:22 | Right. So in fact what you can show is that basically says sometimes what can happen to these algorithms is that they'll get trapped in what's called a saddle point. And that's where you've got a local minima in a few directions but non minima in other directions. And if you happen to get trapped at exactly the middle point of this saddle then your algorithm can get stuck. But people have worked out a variety of tricks to get past and and with those tricks in place, basically the only time that your algorithm ends up converging is when it gets pretty close to what we think is something like a global minimum of a function essentially. |

| Matt: | 36:58 | So this gradient this finding they're trying to find the best place to be in that n dimensional space is what back propagation enables because we can see the complete state space. So now enter ideas about apical credit assignment and how that could work. |

| Blake: | 37:18 | Right. Right. So to be clear back propagation is one possible way of doing gradient descent and what my lab has been proposing is that there - Well we know there are other ways of doing gradient descent and I am personally convinced by the idea that our own brains do something like gradient descent but there are a variety of reasons that the specific details of propagation are just biologically infeasible and one of those things is that in order for back propagation to work you have to do a full path forward through your hierarchy and then do another path towards your hierarchy and these need to be two separate things that you do and there's no evidence that our brains engage in this sort of separate forward and backward past. |

| Matt: | 38:14 | Right |

| Blake: | 38:14 | What's interesting though is that when we look at the physiology of the neurons in our brains a lot of the feedback that arrives at the neurons, the principal neurons of our forebrain, which are called pyramidal neurons, a lot of the |

feedback that arrives at a pyramidal neuron is in its apical dendrite which is a special dendritic compartment that these cells have that basically actually goes up towards the surface of the brain. So what my lab's been interested in is the idea that these apical dendrites might actually provide a way of integrating some of the feedback information that you need to do gradient descent with without disrupting the ongoing processing that's happening in other parts of the cell. And that in this way you could estimate the gradient of your cost function without disrupting without having to do separate forward backward passes or disrupting the processing that's occurring.

Matt: 39:17 So instead of duplicating or doing the past twice you're adding an additional sort of computational unit to each neuron.

Blake: 39:25 That's right. Exactly. So if each neuron has its own little computational unit where it can calculate its gradient information then you don't have to worry about the separate forward and backward passes.

Matt: 39:39 So I just want to relate this to HTM because that's my audience you know. We talk a lot about distal basal dendrites and it sort of having its own computation for a cell that can predict or change its behavior. It's similar to that I think...

Blake: 39:54 So you know this is something that I really like about the model that you guys are building at Numenta is I think thinking about things in this way where you say okay what might this dendritic compartment be contributing to learning and how might that be a distinct computation is something that has rarely come into artificial intelligence but which I suspect is critical to understanding what's going on in the brain because just when you look at the diversity of shapes of neurons in the brain it's pretty clear that there's, the brain is using these different dendritic compartments to do different computation somehow. And so that's got to be part of the answer.

Matt: 40:35 Absolutely. So with these new ideas how do, how can we change current deep learning frameworks? Because this is sort of going to the core of what a neuron is the definition of a neuron how can people earning change to incorporate these new ideas. Do you see a path forward?

Blake: 40:54 Yeah so I think that probably the most important thing for the people is in terms of incorporating some of these ideas is about hardware implementations potentially because you know the fact is that gradient descent works so well that one of the

things, it drives some people who are purists nuts. Gradient descent works so well that what we've seen over the last few years in AI is just an explosion of people saying OK well I'm just going to define this cost function and this particular architecture and then I do gradient descent to it and voila I have now got a new state of the art on some task. And to some extent there's no reason to expect that that has to stop in the near anytime in the near future. It will probably peter out at some point. But as it stands we're still seeing continued improvements from just applying gradient descent to new problems and new cost functions and new architectures.

Matt:	41:57	Five years ago, I wanted an app that I could take a picture of a leaf and tell me what kind of plant it was. That exists now because of gradient descent I imagine.
Blake:	42:06	That's right. Precisely precisely. And so you know I think that where where that might end up failing a bit more though is that if you're actually trying to you know actually build a circuit that does in the hardware deep learning for you not just you know simulating it on a GPU then maybe you'd want to think about potentially having circuits where you've got different compartments where your gradient signals are being calculated or predictive signals like you guys have are being calculated. And this might end up being a more, much more efficient architecture for running deep learning systems.
Matt:	42:54	Do you think there's software changes that can be made to current deep learning frameworks that are in production right now that can incorporate these things or is this going to be like the next phase of AI development that incorporates it?
Blake:	43:09	I think would be more like the next phase. As I said I think this kind of stuff - I think there are a variety of things that neuroscience can still teach deep learning as it stands and we've seen some of that with respect to the incorporation of things like memory and attention and other things. But really I think in terms of some of these ideas about dendrites and how are they going to help it's only going to be when we come to the sort of hardware systems that it might be useful, because you know at the end of the day I suspect that's why the brain did it as well because it was having to implement it in hardware
Matt:	43:47	That makes perfect sense.
Blake:	43:48	And you couldn't just toss around you know numbers to wherever it wanted to at any time.

Matt:	43:51	Yeah that's a good insight there. Well, this has been a great conversation really. I have one more question for you. Are you a theoretical neuroscientist?
Blake:	44:02	Yes I think I am.
Matt:	44:03	That's great. It's hard to find those sometimes.
Blake:	44:07	Yes.
Matt:	44:08	People who admit they are theoretical neuroscientists, so it's nice to see some people claim that because I don't think there's anything wrong with it.
Blake:	44:15	No indeed. And I think that's, that's part of this shift we, this is maybe a good place to come full circle on is part of the shift we're seeing in neuroscience towards this broader perspective that incorporates things like machine learning and other parts of mathematics into neuroscience rather than just taking this very biological approach is that we need to shift to something a little bit more like physics in terms of having people who are theoreticians that are really just thinking deeply about the data that's coming in and trying to integrate it in order to generate mathematical models that can really guide our experiments and guide the hypotheses that we're generating.
Matt:	44:56	Especially with all the data that is right around the corner, these theories are going to be dated or invalidated pretty quickly.
Blake:	45:04	Yep that's right. That's right.
Matt:	45:06	It's an exciting time to be in the field honestly and it's been a pleasure to talk to you, Blake. Thanks for being on the podcast.
Blake:	45:12	Yep thank you. It was a real pleasure, Matt.
Matt:	45:15	Thanks for listening to the Numenta On Intelligence podcast. For more information on our company or our mission visit numenta.com.