AI, Speech Data, and Human Intelligence

Human child might learn language after about five years of, you know, speech. Our AI systems require maybe like a thousand or ten thousand times as much speech data.

our AI Systems are creeping up on us, right? And at some point we're going to have superhuman AI system and those are going be smarter than us. And then who knows what's going happen, everything's gonna fall apart.

I think that's just a misleading way of thinking about intelligence. When you are using the AI rather than thinking something through for yourself, You're training the A. I.

system. You are not training your own brain. Well, hey everybody, we're going to get right back to the podcast, but I want to share something with you that is very personal.

You know, for decades, I've been a practicing neurologist. I have been asking a single question, and what makes a good brain go bad? More importantly, obviously, what can we do about it?

And that question has led me to write my new book, which is called Brain Defenders. It's now available for pre-order. And in this book, I'm going to take a hard look at where we are today and offer up some challenges as to where are we today in terms of what we're being told.

You know, despite all the attention on targeting proteins, misfolded proteins like beta-amyloid for Alzheimer's, we still don't really have any meaningful treatments for diseases like Alzheimer.

and Parkinson's disease per se, and we certainly can treat symptoms. But as it relates to treating the underlying issue, we're falling short. Many of these approaches manage the symptoms at best, but without really addressing what's driving the underlining process.

And this science is evolving. What's really exciting is the emerging research on what I've called the microglia. I didn't call it, that's what they're called.

These are the brain's immune cells. When they balance, they are protecting, repairing, But when they're dysregulated, and we can talk about why that happens, we do talk that in the book, they can drive inflammation and lead to brain degeneration, often long before symptoms appear.

And here's the empowering part. These microglial cells are deeply influenced by your day-to-day lifestyle choices, your nutrition, your environment. And in the new book, Brain Defenders, I share both the science and provide a practical roadmap that'll help you calm inflammation, supportive microglia, and really ultimately take control of your brain's destiny.

If you pre-order your copy today at braindefenders.com, you can receive some special bonuses like a sneak peek at chapter one. the full glossary for the book, and a curated offers from my favorite brain health products, many of which I've discussed here on the podcast.

This is a new way of thinking about brainhealth. And I truly believe that this book hopefully is gonna change everything.

Introducing Brain Defenders and Microglia

I can't wait for you to read it. and I think you're gonna be as excited as I am about the information you are about to receive. That said, let's get right back to our podcast Well, hey everyone, I'm Dr.

David Perlmutter and welcome again to The Empowering Neurologist. Thanks for joining us. We have a very interesting and deep conversation today. It's going to take us into an area that's certainly familiar to all of us, but certainly also profoundly mysterious, and that is the nature of the human thought process itself.

My guest is Tom Griffiths, a cognitive scientist. His work explores one of ambitious questions in all of science and that is can we understand how our minds work using mathematics?

The book is called The Laws of Thought and in this book he walks us through centuries of inquiry into this exact question. from philosophers like Descartes and Leibniz to modern approaches using artificial intelligence, all in an effort to answer a deceptively, I think, simple question, and that is, what are the rules that govern how we actually think?

And what makes this conversation, so compelling, is that it challenges something we often take for granted, that our thoughts are intuitive and even ineffable.

Griffiths has a suggestion that instead of thinking may follow our various underlying principles, much like the laws of physics, if we can understand these principles we better understand not only ourselves but also the rapidly advancing technology that are beginning to mirror the various aspects of human cognition.

That is what we are challenged by. You know, the goal is, can AI ultimately mimic human cognition and even go beyond that? And we're going to discuss that today.

We're gonna unpack a topic that can feel abstract at the beginning and, even, intimidating. I know that in looking at his book, reading his books, it's challenging.

But, you know we look at things like logic probability. AI, of course, has mentioned, but we'll do it in a way that makes it relevant to everyday life.

because ultimately this isn't just about machines or mathematics, it's about how we make our decisions, how learn and how perceive all that's going on around us, our reality.

So this is going to be a deep and I think perhaps challenging podcast, but I the payoff is gonna be really valuable, so let's get started. Well, Tom Griffiths, welcome to The Empowered Neurologist, glad to have you here today.

Thanks, David, great to to here. I'm going to say it's heavy, not in its weight. It's a deep read. And I think you get to one of our fundamental questions, and that is, how do we even think?

What makes us what we are in terms of being able to be better decision makers and what were able recruit to make our decisions? But let me take a step back and ask you why your interest?

what brings you to explore how we think, the laws of thought? I mean, to me, it's really one of these great mysteries. And that's what drew me to psychology in the first place.

I'd done a lot of math and science in high school. When I got to college, I wanted to explore something different and really find an area where we had questions that we didn't have good answers to.

So I think when I learned science and school, we'd sort of just have equations written on the board and we told, oh, that how this works. That's how it works." And when started taking psychology classes, It was kind of a little more like, these are the questions we have, these the answers we've have so far, there's lots of territory for us to continue to explore.

And so, I think that's part of what's exciting about being a cognitive scientist and diving into these questions is we're still very much trying to figure out the answer to these fundamental things about what it is to be.

Well, we are getting right to the core of it because when we look at those equations on the board in your earlier

education, Those are finite. They are defined, they are not analog, there's an answer.

There's no guesswork, There is no nuance to it, right? That's what mathematics is all about. And I think that one of the central questions is, is there the ability to ultimately, if we had the tools, determine exactly what is going on mathematically with respect to the whole issue of thought?

Tom Griffiths and the Laws of Thought

That's been the aspiration of people trying to answer this problem. And that's really the story of the book is people are trying take different pieces of thought and turn them into mathematics.

I think that is the right way to think about it. It's kind of like an expanding circle where we've got some piece that we got a good handle on and then we turn to the next piece and the piece after that.

We don't really know where that process is going to end up. But the very first pieces were logical reasoning, right? So, you know, going from things where you're certain from, things that are true to other things, it's reasonable to infer a true.

And then realizing that wasn't enough, figuring out how do we handle uncertainty? How do capture that? Had we capture things like learning really, the project of a scientist is taking mathematical structures that mathematicians have figured out and then trying to figure out, how does connect to the world?

and the project of a cognitive scientist is trying to figure out what pieces of minds can we express using those mathematical systems. So cognitive science has been around quite some time and had some pretty landmark breakthroughs, at least in terms of theory, and now you are building upon established cognitive science, but you're taking it a step further.

How is what you are describing in the book different from now that you've built on existing cognitive sciences? Where do you diverge? A lot of the story in this book is the stories of how we got to this moment in terms of these different threads that people have used to try and understand how thought works.

So each of those being a different kind of mathematics. In the books I talk about systems of rules and symbols, so things like logic as one sort of first attempt.

And then discovering that didn't quite work, thinking about representing ideas as points in space, and then needing some kind of story about how you do computation in spaces and that being what's called artificial neural networks.

Then the third thread being one that comes from probability of statistics, which really gives us a tool for answering sort-of why questions, like why is it that this system learns?

Why is that it doesn't learn better? Why does it do the things that it does? And so those three threads have started to come together in modern AI systems, where you have systems that are trained on symbolic data, things like language, that using giant neural networks to learn how to approximate that structure, and that train with a probabilistic goal, something like predicting the next word that appears in a sentence.

And that combination of three things turns out to be very powerful. But it also doesn't capture everything about how minds work. And so that's, I think, the big project for cognitive science moving forward is trying to figure out, yeah, again, how much of what human minds do is captured by the mathematics that we have.

So I was going to do this later in the interview, but then how do we compare where we currently are in understanding of human thought process and consciousness to, we'll get back to consciousness a little bit later, to artificial intelligence.

How close is AI to mimicking what you believe is going on in your brain? I think our AI systems right now have a couple of important differences from human minds.

And those are really more a consequence of just the paths that humans and AI system find to solving the problems that they're posed by their environment.

So our current AI Systems have, I'd say, two main issues. One of those is that, they require much more data than humans to learn the things that the learn.

language off for about five years of speech. Our AI systems require maybe like a thousand or 10,000 times as much speech data to get to the point where they get you in terms of the language input that they're trained on.

As it relates to time for that training to take place, that is certainly a lot shorter with the AI platform. Yeah, but so they can be trained in less sort of wall clock time, they are having the equivalent experience of, you know, like a 1000 years, much more data than a child is getting when they learning the structure of language.

So that's one difference. And then the other difference is generalization. So our AI systems generalize in ways that are often surprising to us, or fail to generalise in way that is surprising, where they can solve one problem And you can give them another problem that seems right next door to us, right?

It seems like it's exactly the same kind of thing, maybe with some tweak to it, and they fail miserably at that alternative version of the problem. And so, you know, our intuitions about intelligence, if you have a friend who's really good at solving math problems, You expect there to be other kinds of things they'd be good.

At that's not necessarily how our AI systems work. You can make an AI system that is really going to solving Math problems. But then it is going not be as good at solving other kinds of problems.

It's not sort of generalizably smart in the way that the humans are. And so- Your model of human brain, explain intuition? Intuition in terms of like the- Being intuitive, yeah.

Yeah. I mean, I think a lot of what I do is trying to map these what we call sort folks psychological terms, right? Things that are in way we talk about minds.

to computational and mathematical ideas, right? So intuition, one way of thinking about that is something which is sort of like a kind of implicit learning, it's something where it is not necessarily verbalizable.

It's a way that we're solving a problem without sort of thinking through it consciously. Those kinds of solutions. That's exactly the kind of thing that neural networks normally do.

They're, they're good at learning patterns and finding associations between things instead of making connections. What's unusual is actually the way we use neural network now where we are training them on language and then getting them to generate language in order to answer the questions that they ask them.

And so that's pushing them maybe out of that more intuitive mode and into something which is more like an explicit reasoning mode. But you just mentioned pattern recognition.

Aren't these large models certainly incredibly more advanced as it relates to pattern and recognition in comparison to our brain or not? What they're able to do is process just much more data than a human can.

And so the thing that always impresses me about these models is they know as much as an expert, or at least they can answer questions in the way that an experts would for just very different domains.

So if you're talking to a large language model, you can be asking it detailed medical questions, as I do, detailed mathematical questions. You can, be ask it for like, chemistry, how does cooking things work, what combinations of flavors might be good, it can give you answers to all of those kinds of questions.

It can even do things like make analogies between them. Right? So this is one of my favorite tricks is you find someone who just as a human being happens to be extremely expert in two things, right?

Someone who's a doctor, but also a bluegrass musician and knows a lot about medicine and blue grass. And then you can ask it to, you know, form an analogy or to explain the connections between two very specific things that that person knows, and the model can do that.

And it can that for any human expertise, right? So that's a way in which the models excel. But I think that knowledge is also often surprisingly brittle.

I wouldn't have thought you would come up with bluegrass as your example, but it's been surprising. What about the notion of then creativity? How does your model lend itself to explaining creativity.

So that's actually a big challenge for the current AI models. Although some of the recent work that we've done suggests that they're not as bad at this as has previously been suggested.

Can you define creativity? I use the term, but can you find it for us? Yeah, so by creativity, I'm saying in this case, sort of coming up with new ideas, right?

That are things that are, you know, useful to us as well. Right? So it's not just enough, because you could just randomly generate lots of ideas that wouldn't be that creative in terms of they're not, they are either sort making a list of all of the ideas.

That's, not very interesting. It's coming with ideas are good and new, is I think what we want out of our out of our creative partners. So yeah, large language models are not great at this because what they do is learn the probability distribution over language that they're trained on, right?

So they are learning what the sort of kinds of things are that people say. And that's really good if you want to be able to make a system which is able emulate what a person would do in a particular situation.

It's able play act being a doctor or play acts being chemist or whatever it is that you're asking it to do. Um, it's not so good. If you, if want produce things that those people have never done before.

Right. And by definition, part of it, the wisdom. Yeah. Just agree. Yeah, that's right. A creative idea is often something which, you know, no one has thought before and those things are harder for the models to produce because they're going to be biased towards the kinds of things that people normally do.

And so that imposes a constraint on how useful the model are sort of coming up with sort out of the box ideas, right?

AI vs. Human Cognition

Out of box is not a thing that they generically do That said, some of the work in my lab recently has been looking at, you know, comparing creativity and models and people, and it has some surprising findings, like, um, on a task where we have the models come up with new products.

Uh, so both humans and, people are sort of asked to come with, new product designs. And this is work that was led by LLU in MyLab. They actually come up with product designs that are considered by people to be about as good as the product design people come with.

One interesting difference though is that Human creativity is enhanced by doing things like asking people to form analogies. So if you say, come up with a design for a car based on properties of an octopus, that helps humans come with new, more creative ideas, and that doesn't seem to help the large language models.

One possible explanation is they're already drawing on a big repertoire of knowledge that they are using, so priming them with that knowledge doesn' help.

There are still meaningful differences in the mechanisms of creativity between humans and language model. Hey

everyone, we're going to get right back to the podcast, but I have an important message for you.

If you're caring for somebody with Alzheimer's, any other form of dementia or even what we call mild cognitive impairment, or maybe you've received one of these diagnoses yourself, Or if you are focused on preventing cognitive decline, I want to speak directly to you for just a moment about some serious research that's going on Looking what we call 40 Hertz light and sound simulation.

We've been actually talking about that on the program. Here's the challenge. Not all 40 hertz light devices are the same. Most use what's called stroboscopic light, and that's.

The type of light that flashes then you can see the flashing. And that can cause nausea, it can cost headaches. And if you can't tolerate that, you won't use it.

So there is a company called Optosudix. They've solved this problem with a patented technology that still gives you the 40 Hertz light flashing, but through light that looks and feels quite normal.

It's the device I actually have on my desk when I'm working. And that's why this company sees a 94% adherence rate and significant improvements across various metrics, including mood, energy, focus, sleep, and memory.

And the light is called the EV light, EVY. You can use it as I do when you're working on your computer, when your reading, watching TV, eating breakfast, whatever.

It becomes really part of your day, not just another burden. Its kind of passive in the background. So I recommend giving it a try for 90 days. And if you and your family don't see value in it, then you can return the device for a full refund.

They cover the shipping both ways. You can learn more about Optosudix at optosüdix.com forward slash pearlmutter. youcan use the code to get a $200 savings.

The code is pearl mutter, my last name 26. This is a device that really is risk free in terms of getting involved with 40 Hertz stimulation. Important information, let's get right back to our podcast.

So when do we get there with these large language models? What's your sense in terms of scale and the speed at which things are being developed? I think there are two different goals that you could have.

One goal is my goal, the kind of understanding how human minds work goal. I actually think the kinds of things that we have in these models are not too far from at least being the sort of basic ingredients that you need to explain where it is that human intelligence is coming from.

The main thing that's missing there is what we call in machine learning and cognitive science, inductive bias, right? Which is the reason why we're able to learn from less data is that our human brains are sort of biased towards certain kinds of solutions.

And so we don't need to see as much information in order to reach those conclusions. So our brains sort bias towards, yeah, like certain kind of languages, that's right.

We're biased toward reaching certain, being able learn certain things more easily. And that's a consequence of our evolutionary history and the constraints under which human minds have evolved.

So I think there are probably ways to capture those things. That's one of the things I work on in my lab is trying to think about how do we put more human-like inductive biases into our AI systems.

And that, that will bring us closer to understanding key parts of human intelligence. If you want to build. I mean, I understand human thought, but is, is the goal to make these platforms more human-like or what?

I. No, no, if your goal is to. An AI system, right. then maybe you're on a different track, because I think our expectations for an AI system go beyond our expectation for a human, right?

Just like when people are building a self-driving car, it's not enough to make a car that is as safe as a driver. It needs to be as safer as the human driver in all the context where the driver would have done the right thing, the car will do the thing.

We want it to not make dumb mistakes. and probably has to be even safer than a human driver for us to say, okay, this is something that we're willing to sort of give up our human autonomy and let the cars take over the driving for.

And same for an AI system. I think we have higher standards for how reliable that system has be in order for to think about using it in places where previously a we'd be doing something.

And so if you want to make a system that has that goal, I think there's a lot of things that we need to do in terms of enhancing the kinds of approaches to AI that already have.

So let me take this to the idea of how the laws that you describe in the book, how does it play upon our understanding of mental health disorders? A lot what I focus on in book is, we talk about this in cognitive science in different levels of analysis.

We can think at the most abstract level, we can ask a question like, what is it that human minds are doing? What are the kinds of problems that humans are solving?

And what do ideal solutions to those problems look like? Right. And so that abstract level is really trying to figure out what are the sort of general principles of intelligence.

These would be things that are shared between humans, between AI. If aliens we encounter are sort equivalently intelligent to us, you know, anywhere in the universe, we would expect intelligent organisms to follow the same principles.

That's what I'm calling the laws of thought. At the next level, We have how it is that, You know minds approximate those solutions. Right? Like it's not really possible for us to, you know, do logic perfectly or do arithmetic perfectly, or any of these kinds of math perfectly.

We're sort of approximating them with the finite resources that we have. Um, and so that's a question about, yeah, what are the kinds strategies that human minds use for solving those problems?

And then the level below that focuses on how is it that those things are then instantiated in brains, right? In neurons or for our computers in silicon chips.

So, whatever it is, that the sort physical substrate of intelligence. I was gonna say, so when you ask about psychiatric disorders, I think you can think about those at each of those different levels of analysis, right?

You can about, in some cases there might be something where somehow people are solving a different problem from what they should be solving. It could also be that something's gone wrong with the strategies that they're using for solving the problem.

So they are just like paying too much attention to one kind of information and that's causing them to become more anxious or whatever it is. Or it could be there's something has happened in the brain that's making it hard for the brain to execute the algorithms and strategies that it normally executes and that blocking, you know, being able to successfully solve those problems.

And so I think when people think about computational psychiatry, they're often thinking at all of those different levels of analysis.

almost always an approximation is really, I hate to say best guess, but in any decision, any thought process, if we get close, generally there's a proximity to reality that we always seem to be approaching and pretty much works for us.

I wonder if there is a breakdown in that, in the proximity of the right decision that might relate to mental illness. Do you understand where we're going with that?

Because I don't know if I'm able to elucidate that. Yeah. So some of the work that we do is… That doesn't allow us to get close enough to whatever that answer is to explain our experience.

Yes. Some of work I do was on an idea that called resource rationality. And there the idea is that We're not able to reach those perfect solutions because, you know, we have finite brains, right?

And we're trying to do the best job we can using the computational resources that are available to us. And so that is a way of understanding a sort of paradox that you see in psychology and computer science where on the one hand, psychologists are going to tell you we are bad decision makers because we follow these simple heuristics that lead to systematic biases.

Worked by people like Danny Kahneman sort revealed what that structure looks like. On the other hand, computer scientists still hold up humans as the thing that we're trying to beat when we are building our intelligent systems.

That's our best example so far we have of a system that's smart. One way of understanding that paradox is to say, well, we try to solve really hard problems.

We're doing the best job we can of solving those problems with the resources that have. It turns out that that best is often using some kind of heuristic, and it results in some biases, but that is the kind best you could do.

So you can imagine if you change the resources that are available to you, the kind of strategy that you would follow would change as well. And so that's one way of thinking about, I think, some of these kinds of cases that inhabit psychiatry, where it's that something is not working in your brain in the way that it should.

You're trying to do the best that can. despite that constraint, and then that's something that produces then what looks like unusual or disordered behavior.

But the input based upon the platform has changed. And I think that seems to be one of the major goals is to reestablish input from a different, more functional platform, whether it's through modern psychiatry.

That makes us really think about medication as an inroad to a bit myopic if we recognize that the is playing such a fundamental role in nuancing the entire thought process.

Changing serotonin levels seems to be a little bit narrow-minded to me respectful. I think you could think about this as there's different kinds of interventions that might make sense at those different levels.

Sure, if it's something that's fundamentally gone wrong in that physical substrat, then maybe medication or something like that is a reasonable intervention.

But if it's more about the strategies that people are following instead of following maladaptive strategies, then there are all sorts of other kinds of techniques that you can imagine that are more sort of cognitive behavioral.

And those maladapter strategies again are a manifestation of the platform, the input. I'm going to go to a place that, you didn't necessarily cover, but I think I would suspect you've thought about.

And that is there's such a big push these days to at least explore the idea of psychedelic-assisted interventions to

basically rewrite that operating platform.

Any thoughts on that? And then we'll get back to the book, but I'd love to hear your input from that. Yeah, we've been sort of gradually walking outside my area of expertise, and we're taking a step outside.

You're a smart guy. But no, I mean, it's intriguing to think about how you can explain those kinds of effects in the terms of the kinds systems that I'm talking about here.

And I have colleagues who are interested in doing that. So my former colleague, Alison Gopnik at Berkeley, has a, you know, sort of somewhat worked out computational sort story about, uh, impacts of psychedelics in terms of changing attentional resources, right?

So when I was talking about resource constraints, one of those resource, constraints for us is attention that, um, Uh, that means we have to make decisions about what we're attending to and we sort of narrowing down the spectrum of our experience.

One of the things that psychedelics do is, you know, sort-of broaden out that focus of attention. And that might be something which means, yeah, You're getting information that you weren't getting because of all the default strategies you can follow.

Let's get back to the book then. I think that anyone who is going to pick up your book and read it, but I some people, let's just say, two words come to mind and that would be free will, right?

I mean, that's been the age old question. Do we have freewill or is this predictive? So can you unpack that for us a bit through the lens of the laws?

Yeah. So the things that I talk about in the books, so logic, neural networks, probability theory, None of them I think is inconsistent with the notion of free will because the kinds of problems that human minds have to solve are things that are extremely complicated in a way that means that there's a lot of choice points in whatever you're doing, right?

So you take the example of logic. On a surface, logic seems like something that's really deterministic, like it sort of tells you How to go from things that are true to things, that a true it has rules.

You know the way that it works when you're doing logic is you have some facts that you know are. And then you use these rules to figure out what other facts might be true.

The problem is that there are so many rules that could apply in any one situation. that you have to make good decisions about which rule you're going to apply.

This is what makes math hard, right? The sort of math that mathematicians do is proving something is a matter of starting with some things you know and then following a very sort secure path, where you are making a series of choices about the deductions that make until you get to the end, or you can think about something like a game of chess.

Game of Chess has rules, tells you where to put your pieces when you start the game, it tells how those pieces move. And from those rules, you can now generate all possible games of chess.

Your goal when you're playing chess is to find a path through that, vast array of possible choices that you could make that gets you to the point where you win the game.

But that's extremely difficult, right? And so just the complexity of that problem means that there's plenty of room for a human being to have optionality and flexibility in the choices they're making when they are playing that game How then, how might we reconsider our approaches to education based upon what you've described in your book?

How we try to inculcate knowledge into the developing mind, our techniques for that, How might that change? All of the approaches that I talk about in the book are things that have made contact with education.

I talked about this rules and symbols approach. One of insights that was at the heart of that, was Noam Chomsky's characterization of structure of languages.

That's something that led to kids diagramming sentences in school, trying to figure out the kinds of structures that, that underlie language. Um, uh, neural networks tell us about how learning works.

And from that you can design interventions that are about, you know, like teaching people, um, for example, a classic sort of psychological result is that it's better to study something, You know a little bit every day rather than wait before the tests and then spend several hours studying it all at once.

Right. And those are principles that you can get from thinking about how neural networks work. And then probability and statistics tells us about, um, how the assumptions that learners make about what they're learning is something that can influence the rate at which they learn.

So if you are fully expecting, yeah, if your expecting to get good examples, right? Cause you think your teacher is a good teacher. That is. Something that changes.

the rate at which you learn. If you just think I'm getting random examples, you'll learn more slowly than if you think you're getting good examples. That's just because your expectations about the relationship between the data you are being given and the conclusions that you should draw is different.

Can we remove that block and utilize what you've just said day in and day out in our learning process? I think that we can. Yeah, I mean, there are colleagues who are in education schools that are interested in understanding these kinds of mechanisms and thinking about how you translate them into the world.

I think there's certainly challenges involved in doing that. So a lot of these things are really trying to characterize just what the ways are that human minds by default work, right?

And then you can think about engineering systems to better engage with those human minds. And mostly we've done that in the lab. We've got a couple of things where we looked a little further out in terms of doing things like diagnosing how someone learning algebra might have the wrong concepts.

Some work they did with a former student, Anna Rafferty, where were able to use probabilistic methods to solve that problem. But I think there's lots of room to take what we've learned from this science of the mind and translate it into more effective educational interventions.

You know, a lot of my work has focused on what call environmental evolutionary mismatch that our various bodily systems have evolved over time to function in harmony with a well-defined set of environmental cues.

As it relates to the process of thought, I'd be interested in hearing what you have to say about The idea that we're confronting the whole notion of thought with technology in a way that is unlike anything the human brain slash thought has ever experienced.

So that's a bit of a mismatch. Like the mismatched between ultra-processed foods and our metabolism, for example. What are your thoughts on how our processing and processes have been affected by being confronted by technology in a way that even in the past 10 years has never been predicted?

Let me first sort of talk about, I think this is actually a good lens for thinking about some of the differences between humans and AI, which is that A lot of the things that we've talked about as ways in which human minds differ from AI systems come down to kind of like biological constraints on how humans work, right?

So humans need to solve a particular set of problems. We only live for a few decades. You know, we only have the

compute that carry around inside our heads, a couple pounds of neurons.

And we can only communicate with one another by doing things like making honking noises or moving our fingers, right? So we're very constrained in terms of the things that we do.

But those constraints are key to the way that our intelligence works. So, we learn from small amounts of data because we have to. We are efficient in the ways that use our computational resources because were not getting more, and we are able to come up with things language, society, companies, mechanisms that we create to allow us to pool our resources together to achieve things that go beyond what it is that were able to do individually.

That set of things, that defines what humans are, is different from the way that our AI systems work. Our AI system can learn from many human lifetimes of data, can get more and more compute as they need it, you know, share data between them transparently.

You can take one system that's being trained on one thing, train it on something else, you can sort of copy them around. They're very different kinds of organisms from us.

And so that accounts for these differences I was talking about in terms of inductive bias and how quickly systems can learn and what kinds Because we've been shaped by being these biological organisms and our AI systems are not subject to those same constraints.

And as a consequence, we're sort of aliens. We just call them organisms. That's a step. I would scare quote it as organisms, but yeah. So I think that's actually important for us developing better intuitions about AI, because I we project our human intuitations about intelligence, the intuitons that have been shaped by interacting with other humans onto these systems, by really thinking about them as alien and different from us and finding different kinds of solutions.

Even though on the surface those solutions look quite similar, it's probably a good mode to be in when you're interacting And it sort of maybe sets us up in a way where, as you're saying, we might expect that using these technologies is going to have different impacts on human minds than interacting with another human being mind.

Hey, We're going get right back to the podcast, but I do have an important message. You know, for decades we've been told that our genes are basically our destiny as it relates to health and that everything's locked in place, predicting even our future brain health.

But let's be clear, your DNA isn't a verdict, it's a roadmap. And I'm talking to you today about 3×4 genetics. I actually described them in the new book, Brain Defenders.

And this is a committee that helps you understand your genetics, what you've inherited, and I think more importantly, how your daily choices can influence how those genes are actually expressed.

In other words, how you can control your own genetic expression. And that matters for your health and it really matters to your brain health, including things like memory and mood and cognitive resilience.

You may carry genetic markers associated with cognitive decline, like as we've talked about APOE4, but possession doesn't equal expression, your lifestyle, you're nutrition and metabolic health.

All the things that we talk about here on the podcast can powerfully influence which genes are turned on and which stay quiet. The 3×4 genetics test looks at how your genes impact brain function at a cellular level and, which then through the 3X4 health journey, in which they provide you, guides you through a clear and a science-backed program focused on inflammation, methylation, oxidation, oxidative stress, things that we routinely discuss here on the podcast that are the key drivers of basically what makes a good brain go bad.

Free Will, Mismatch, and Future AI

and we have spoken with the team at 3×4 Genetics and they were able to get a special offer for our podcast community. So head on over to 3X4Genetics.com 4 slash drpearlmutter and get that special off and start really your personalized brain health journey today.

Let's get your genes working for you and let's go back to the podcast. What about uncertainty? How does uncertainty influence cognition? Is that a positive event experience?

And if it's ongoing, or how does that affect the ability that we have to think? I would say a lot of what characterises our strengths as intelligent systems, for humans, is our ability to deal with uncertainty.

Logic, they said, takes you from things that are true to things are that true. That turns out to be a pretty small amount of the kinds of things humans do, right?

So when cognitive scientists started with logic as their model of what thought might be like, They made a bunch of progress on things like problem solving, you know, reasoning, playing games, doing arithmetic, other kinds math.

There's a set of thing that they could describe well using that structure. But then there was another set of things that they couldn't describe at all.

Things like how vision works, right? How it is that you perceive the world based on the sensory signals that hit your eyes. Or same thing for audition, for hearing.

How is it that people discover new causal relationships between things? how it it's that we learn language, these are all problems where there's actually some uncertainty at the heart of them.

So you can take the example of vision. You just don't get enough information from the visual world to be able to determine what the three-dimensional structure is that's around you, right?

Your brain is constantly solving this problem of taking the two- dimensional information that is hitting your retina and turning it into something like a sort of three dimensional picture of the world.

And so the fun of this involved in that, is really quite remarkable. When that has been distorted by prism glasses or flipped upside down, how quickly the brain can adapt to that.

Yep. And it's a big chunk of your brain, right? Like sort of like the back, you know, broadly construed, the black half of brain is doing, trying to solve this very complicated problem.

So they're big. Yeah. The whole structure of that problem is one that has uncertainty in it because you just don't have all the information that you need.

You don' t notice that because it' s not effortful to us. Your brain does it for you. The only time when you realize that there might be something funny going on is when see a visual illusion and you're left with, say, multiple interpretations of what an image is, right?

You might able to switch back and forth between different interpretations or you are sort of surprised to discover that something is different from what it looked like.

But other than that, it's something which you by and large don't notice. And so that same framing applies to things like how we figure out causal relationships in the world.

Learning causal relationship, no one actually gets to see a real causal relation, unless it is like billiard balls hitting each other. The things that are the deep causal relations in world around us, are things that we infer from observations that make that are imperfect.

And so we have to deal with uncertainty when we do that. Or thinking about someone else's mental state, right?

Someone behaves in a certain way and now you come up with an interpretation for it.

That is, you know, again, making an uncertain inference. So I think this is pretty intrinsic to what it is that human minds do. Then it's sort of the edge cases where we're trying to do some sort perspective, thinking, some future planning where you end up feeling uncertainty.

But mostly our brains just do a really good job of dealing with it. Best guess, giving it the best guess. But your model could predict that, right? Your model can predict what that guess would be.

Yeah. And that's in the third part of the book, it's about probability and statistics. It's really about how it is that minds and brains solve that problem.

So there's an approach that's called Bayesian inference. It's named after an 18th century British minister, the Reverend Thomas Bayes, who kind of came up with the math for this.

But it's basically using ideas from probability theory to tell you which guess is the best guess, right? Which of many hypotheses is that one that you should take most seriously based on the data that.

And still applied to scientific assessment of probability, Yep. Yeah. But it also gives us a good foundation for thinking about human cognition, because we can think about things like, what are the solutions that people are biased towards being measured in terms of those probabilities?

So you can say something like before people see anything, before they hear any language, they have a bias towards certain languages being the ones that they might learn, and we could express that in some probability distribution of languages.

Let me go to intelligence, if I may, and that is there seems to be that many misconceptions about intelligence seem to abound. What do you think are the biggest misconceptions that you would like to rewrite for people in terms of how they understand what intelligence is?

Yeah. So I think a lot of the challenges that people have in thinking about AI at the moment are really a consequence of using our human idea of intelligence and generalizing it to an AI system where, you know, if you have a friend who can solve an international math Olympiad problem, that friend is going to be able to do all sorts of other things, right?

And our AI systems that can solve international mathematical problems might have difficulty solving other more basic problems that are the kinds of things that would be surprising to us if your human friend couldn't do it.

And so that's playing into a sort of notion that maybe there's a one-dimensional space of intelligence. Intelligence is a thing. We have maybe some animals down here.

Humans are here, and then our AI systems are creeping up on us. And at some point, we're going to have superhuman AI system, so those are going be smarter than us, then who knows what's going happen.

Everything's gonna fall apart. I think that just a misleading way of thinking about intelligence, that it's better to think about it as You know, more like a sort of Darwinian picture, right?

That, that we have in nature different kinds of creatures that have adapted to solve different types of problems. Those creatures can solve problems better than we can in many cases, you can like have spiders that can hunt really effectively and you know ants that do collective coordination things better then humans and there's lots of nice examples of animals being sort of very smart within their niches.

And human minds in the same way have evolved within a niche that's constrained by these constraints on lifespan and computation and ability to communicate that means that we're really good at solving certain kinds of problems.

We're building our AI systems using data from humans and trying to get them to solve similar kinds of problems but without those same constraints. And so they have evolved towards maybe a different set of solutions that look kind of like us, but also differ in systematic ways that mean that you get these errors in generalization and so on.

And so I think thinking about it as not one scale, but, you know, yeah, like a branching tree where there's lots of directions that you can be going in as an intelligence system is probably a healthier way of us engaging with AI.

And it's something that I didn't remove some of the anxiety about, being overtaken because it was really that we're sort of just growing out in those different directions.

Let's, for our viewers, spend a little time on the idea of optimization of the process of thinking of thought. And where we could start maybe with sleep, where would you like to start that and just give our viewer some take-home ideas?

Okay. So sleep is something that's maybe an intervention at that, you know, biological level, at the level where your brain is working and so on. Most of the things that I work on are interventions that are at a level above that and thinking about the kinds of strategies that you can use.

And so, um, I have another book called algorithms to live by with my coauthor, Brian Christian, which is about exactly that. It's, it's telling you, if you're solving a certain kind of problem, what's the optimal algorithm for solving that problem?

And then how can you implement that? So, from. deciding when you've seen enough apartments and you're ready to start writing checks to being able to find parking spots or figuring out whether you should be organizing the books on your bookshelf.

We actually have practical solutions to all of those problems. In terms of the kinds of ideas that are in this book, the way that I would translate this into sort of better ways of thinking is by thinking about some of these kinds of questions about how it is that we can be complementary to the kinds If people are making bad decisions because of resource constraints, because they don't have enough computation to be able to look far enough into the future to determine what the consequences of those actions might be, that's actually an interesting opportunity for an AI system to help us, where you can have the AI pre-compute what some of the those consequences look like, so you could make a more informed decision.

just based on having the information accessible to you. A lot of the things we've been doing in my lab or along these lines are thinking about how do we put more computation into human decision environments in a way that helps the human to make better decisions.

Let me put you on the spot a little bit here. Take us five years in the future. How far has the evolution of the interaction between AI and the common man in terms of his or her common person, his of her life, where are we five years from now in term of what that relationship looks like?

Right now, people are using chat for various things, writing letters, or writing books, scripting a podcast intro and outro, whatever it may be. But what does it look like five years from now with its evolution?

Yeah, so I think we should expect that AI systems are going to continue to get better. You know, along the dimensions that they've been getting better already, which means doing a better job of solving problems that are already solving and expanding the set of problems they're able to solve well.

And the reason for this is basically that the way that these AI systems are created is by training them on data. And a lot of the AI companies at this point are making big investments in getting more specialized kinds of data in the settings where people want to be using AI.

So if you're using a medical AI, they're getting medical data. If you want use it in a legal setting, it's more legal data, and so there's a big focus on filling in some of gaps in behavior of these systems by getting targeted data sets that allow you to create AI systems that are able to solve those problems.

And so that process doesn't require any new technology. It's really just a matter of building those data sets. And we should expect that those AI systems are going to continue to get better.

I think without a change in the fundamental technology that underlies it, we're going to continue to be in this same mode though of having AI systems that produce weird errors in unexpected contexts.

And the frequency of those errors might go down, but it's really a consequence of the fact that these are sort of alien systems, they're solving a problem that's unlike the problem we are solving in a particular way where they are using this kind of neural network technology.

Unless there's a changes in what the underlying technology looks like, I think that's going to continue to be the problem that we have. Well, what happens to our problem solving and decision making abilities, F, once we really start to outsource this much more aggressively than we're doing now?

Yeah. So, I mean, think you can look at other technologies as an example where, you know, like calculators are a good instance of this, where before there were calculaters, if you wanted to solve a math problem, and you had to have a human do it for you.

Or a Spies rule. Yeah, that's right. We're very good at doing those kinds of computations. And, and even today, um, in places where they use an abacus rather than a, uh, a calculator, kids actually learn internal abacas algorithms where are they're able to, you know, do, yeah, really complicated.

Computations, right? Just using their brains. That's something where I think we were willing to let our mathematical abilities slide. Right. So. the expectation that you would have access to something like a calculator in any context where you'd need to do that kind of mathematics seems reasonable enough that people said, okay, you don't really need learn how to three digit times three-digit multiplication problems in your head, right?

That's something which you're probably going to be able to use a machine for. And so I think this is a societal question about What are the things that we think one needs to be able to do to, be a functioning member of society without augmentation, right?

And then that's something that really need to target when we really think about education and so on. That just like you have your classes where you learn math before you use a calculator and then you get to use the calculator later on, that something I think we need very conscious of as a society in terms of making decisions about what are, the places where we allow the AI system to, to step into, you know, the environments where people are using these things.

And then also thinking about that in the rest of our lives, right? That when you are. Using the Ai rather than thinking something through for yourself, your training the ai system.

You're not training your own. Yeah, and there's a risk to that, should these things not be fully available at some point in the future. They're highly energy dependent, right, as is the human brain.

You know, we're losing, I think we are setting ourselves up for a loss of some basic skill sets that were inherently good for what they could do in terms of problem solving, but also, the sheer action of solving those problems was good, for the brain as well.

Yep. The complementary side of this though is that there's another set of skills that we are developing in order to work with these systems. And that's a set a skill that, you know, we're sort of automating cognition.

and then that means that. We have to learn metacognition, right? So met acognitions is. thinking about your own thinking and the strategies you might follow, or even thinking other people's thinking, and who it is you may ask to do a job, who you ask for information.

So as we're building internal models of what the capabilities of these systems are, making decisions about what

tasks to give them, We're building out some of those skills, which are more like the kind of like planning, metacognitive reasoning sorts of skills.

And so I think, yeah, there's some complementarity. It's almost like everyone has become a manager rather than a worker in our cognitive economy, right?

Where you now have your AI be your worker that's going to go off and do those things. You have to figure out how to manage that worker effectively in order to get the best outcomes from it.

And the best managers are usually people who've had that job already and know what it looks like to do a good job of that and that are set up to, you know, help guide an employee through it and produce good products.

And so that means you still have that goal of really learning how to a do good of doing something in order to be someone who's able to actually use the AI in the way best way possible.

Tom Griffiths, The Laws of Thought, thank you for joining us today. This is, you know, it's something that a lot of us are thinking about, especially as it relates to current developments and what the future looks like.

So thank for sharing time with us. Thank you very much. Wow, this has been a fascinating, a deep dive, you know, an exploration into the architecture of thought.

Something, we're all experiencing every moment, yet rarely stop to take a look at, to examine. And I think what is so powerful about this conversation is the reminder that our thinking isn't random, at least as we learn today.

It is shaped by patterns, by probabilities, that our brains interpret and organize information both in real time in terms of our experiences and certainly tap into information that we've stored away from the past.

And as we discussed, understanding these patterns isn't just an academic exercise, it has real implications for how we make decisions, how navigate challenges and uncertainty, and even how relate to emerging technologies that are increasingly capable of mimicking aspects of human cognition.

And I will assure you, as we talked about, those technologies are advancing day by day, week by week. You know, at the same time, one of the most important takeaways is that while machines can replicate various aspects of thinking that we are seeing today, they still fall short in meaningful ways, particularly when it comes to things like intuition, creativity, and the ability to learn from very limited information.

So far, we hold the high cards as it relates to these indices. So this isn't just a conversation about AI, it's a conversion about our intelligence, human intelligence.

What defines it, what shapes it and how we might better harness this. So what a podcast. Thank you for joining me on the Empowering Neurologist today.

If you found this episode valuable, I encourage you to share it. And as always, especially after the podcast today, stay curious. I'm Dr. David Perlmutter for the empowering neurologist.

We'll be back soon. Bye for now.