

/mlst

The Brain Abstracted - The Simplification of Reality in Science

September 26, 2026



Machine Learning Street Talk

Podcast Transcript

Revision: 9491798c

Topic headings include retention indicators: ● above average, ● average, ● below average (vs. similar YouTube videos)

Interested in licensing ReScript for your podcast? Get in touch with Tim Scarfe.

Contents

The Wood Louse & The Spherical Cow ●	2
The Necessity of Abstraction ●	3
Simplicius vs. Ignorantio: The Boxing Match ●	4
The Kaleidoscope Hypothesis ●	5
Is the Mind Software? ●	7
Critique of Causal Patterns ●	9
Temperature is Not a Thing ●	10
The Ship of Theseus & Ontology ●	12
Metaphors Hardening into Reality ●	15
The Illusion of AGI Inevitability ●	16
Prediction vs. Understanding ●	17
Climbing the Mountain vs. The Helicopter ●	20
Haptic Realism & The Limits of Knowledge ●	21
References	25

The Wood Louse & The Spherical Cow ●

Speaker 0

00:00:00

Let me tell you a little story. 19 sixties in the summer, a little kid named Karl was playing around in the back of his garden and he noticed all of these wood lice crawling around, know, the little insects that can curl up into a ball. And what he noticed was that depending on whether they were in the sun or in the shade, they would move faster or slower. They behaved differently. And that's it. Karl grew up to be Professor Karl Friston 1 of the most cited neuroscientists alive. He's been on this channel before more times than I can count and that childhood observation about wood lice, it never left him. He spent decades developing what he calls the free energy principle, which tries to explain all of behavior with 1 equation. Perception, action, learning, why you scratch your nose, all of it, Friston claims, comes down to minimizing a single mathematical quantity. There's an old physics joke, assume that we can model a spherical cow in a vacuum. The joke is about how scientists grotesquely simplify messy reality to tame it. The free energy principle might be the ultimate spherical cow. It promises to explain self organization, this bewilderingly complicated phenomenon with something so emaciated we might as well call it tautological. Even Friston himself agrees with this by the way. This is what he said to us last time we spoke with him.¹

Speaker 1

00:01:32

The free energy principle is not meant to be complicated or difficult to understand. It's actually almost tautologically simple. So the whole free energy principle is just basically a principle of least action pertaining to density dynamics, the dynamics of the evolution of not densities but conditional densities. That's just it. This is before thermodynamics, this is before quantum mechanics, it's just about conditional probability distributions.

¹Karl Friston (UCL) — Person Cited as the originator of the Free Energy Principle.

The Necessity of Abstraction •

Speaker 0

00:02:04

So what do we do with this? Has Friston actually found some deep truth about how minds work? Or is he doing what many scientists do which is mistaking the simplification for the actual thing? Well it turns out there's a philosopher who has spent an incredible amount of time thinking about this exact problem. Professor Mazviita Chirimuuta teaches at Edinburgh University Her book, *The Brain Abstracted*, is basically about what happens when neuroscientists simplify brains to study them. What gets captured? What gets lost?²

Speaker 2

00:02:37

1 of the answers that might seem obvious to people is that we pursue science because we are curious. We just want to know how the world works. We want to reveal, discover the underlying principles of the universe which apply in all cases. Switching off the idea that you are just interested in nature for its own sake out of curiosity and saying okay how can we engineer these systems to actually do things that we want. Getting them to behave in artificial ways. If those simplifications sort of allow you to achieve your technological goals. There is no in principle problem with oversimplification if you are going to say I am not just interested in nature for its own sake, I just want applied science.³

Speaker 0

00:03:23

I should say by the way that *The Brain Abstracted* probably influenced my thinking more in 2025 than anything else. She's an inspirational lady. I look up to her very much and certainly thinking back on many of the episodes we've done in 2025, I can see her influence in the questions I ask and how I think about things. So here's her starting point. Scientists have to simplify. We're limited creatures trying to wrap our heads around systems way more complex than we can actually comprehend. Our working memory holds maybe 7 items. Our attention is more scattered than a group of toddlers with iPads. We die after 80 years if we're lucky. So we build models. Right? We leave stuff out on purpose. We tell ourselves stories about how the world works. But the question is, why does any of this even work at all?

²The Brain Abstracted — Book Prof. Chirimuuta's book discussed as the primary influence on the episode's theme.

³The Brain Abstracted Patreon interview — Interview.

Speaker 3

00:04:11

Science is a humanistic endeavor. Right? The purpose of science in the universe is to make the universe intelligible to us, not to control it, not to predict it, and not to exploit it. Now, you can do all those wonderful things if you like, but in the end, as far as I'm concerned, science is no different from poetry. It's that we're trying to make sense of the world, trying to give it meaning in relation to our own existence.⁴

Simplicius vs. Ignorantio: The Boxing Match ●**Speaker 3**

00:04:11

Science is a humanistic endeavor. Right? The purpose of science in the universe is to make the universe intelligible to us, not to control it, not to predict it, and not to exploit it. Now, you can do all those wonderful things if you like, but in the end, as far as I'm concerned, science is no different from poetry. It's that we're trying to make sense of the world, trying to give it meaning in relation to our own existence.⁵

Speaker 0

00:04:42

If you'll allow the indulgence, I want to tell a little story. It's a boxing match. In the red corner, Simplicius. He thinks science works because the universe is actually simple underneath. Find an elegant equation and you've hit the real thing. Simplicity tells you that you're on the right track. And in the blue corner, ignorantio, he thinks we simplify because we're too dumb to do otherwise. Our models work well enough for our purposes, but they're approximations, just useful fictions if you like, the map not the territory. Now, both of them agree that scientists need to simplify. But where they disagree is what that means about reality. Simplicius had history on his side or at least a certain type of history. Galileo, Newton, Einstein, they all believed pretty explicitly that nature was fundamentally orderly and that finding simple laws meant you'd found something true. Einstein famously said, God doesn't play dice. And no, he didn't actually think god had anything to do with it, but he was expressing faith that the universe is at the very bottom legible. Now, Chiromuta has gone all in on ignorantio's position. She thinks successful science tells us we've become good at building useful simplifications and that doesn't prove that nature is simple. The philosopher Nicholas of Cusa had a phrase for this attitude, doctor ignorantia. Basically, learned ignorance. You study hard, you learn a lot and what you learn includes

⁴David Krakauer's presentation on intelligence. — Interview

⁵David Krakauer's presentation on intelligence. — Interview

what you don't know. Now, when we interviewed Chirimuuta, she'd been following Francois Chollet's videos and for those of you who don't know Francois is a friend of the channel. He's our mascot. He's 1 of my heroes And he's got this idea called the kaleidoscope hypothesis, which is basically that the universe is made out of code and underneath all of the apparent gnarly mess that we see, there is intrinsic underlying structure.⁶⁷

The Kaleidoscope Hypothesis ●

Speaker 0

00:04:42

If you'll allow the indulgence, I want to tell a little story. It's a boxing match. In the red corner, Simplicius. He thinks science works because the universe is actually simple underneath. Find an elegant equation and you've hit the real thing. Simplicity tells you that you're on the right track. And in the blue corner, ignorantio, he thinks we simplify because we're too dumb to do otherwise. Our models work well enough for our purposes, but they're approximations, just useful fictions if you like, the map not the territory. Now, both of them agree that scientists need to simplify. But where they disagree is what that means about reality. Simplicius had history on his side or at least a certain type of history. Galileo, Newton, Einstein, they all believed pretty explicitly that nature was fundamentally orderly and that finding simple laws meant you'd found something true. Einstein famously said, God doesn't play dice. And no, he didn't actually think god had anything to do with it, but he was expressing faith that the universe is at the very bottom legible. Now, Chiromuta has gone all in on ignorantio's position. She thinks successful science tells us we've become good at building useful simplifications and that doesn't prove that nature is simple. The philosopher Nicholas of Cusa had a phrase for this attitude, doctor ignorantia. Basically, learned ignorance. You study hard, you learn a lot and what you learn includes what you don't know. Now, when we interviewed Chirimuuta, she'd been following Francois Chollet's videos and for those of you who don't know Francois is a friend of the channel. He's our mascot. He's 1 of my heroes And he's got this idea called the kaleidoscope hypothesis, which is basically that the universe is made out of code and underneath all of the apparent gnarly mess that we see, there is intrinsic underlying structure.⁸⁹

⁶On Learned Ignorance — Book Refencing Nicholas of Cusa's concept of "docta ignorantia".

⁷Francois Chollet — Person Cited for his "Kaleidoscope Hypothesis" regarding intelligence and abstraction.

⁸On Learned Ignorance — Book Refencing Nicholas of Cusa's concept of "docta ignorantia".

⁹Francois Chollet — Person Cited for his "Kaleidoscope Hypothesis" regarding intelligence and abstraction.

Speaker 4

00:06:39

Everyone knows what the kaleidoscope is. Right? It's like this cardboard tube with a few bits of colored glass in it. This is just like few bits of original information gets mirrored and repeated and transformed and they create this tremendous richness of complex patterns. You know, it's it's beautiful. The kaleidoscope hypothesis is this idea that the world in general and any domain in particular follows the same structure that it appears on the surface to be extremely rich and complex and infinitely novel with every passing moment. But in reality, it is made from the repetition and composition of just a few atoms of meaning. A big part of intelligence is the process of mining your experience of the world to identify bits that are repeated and to extract them, extract these unique atoms of meaning. When we extract them, we call them abstractions.¹⁰

Speaker 0

00:07:52

Now she's not saying that Chollet is wrong She's saying that he's making a philosophical bet. Might be right, might be wrong. It's the same bet that Plato made.

Speaker 2

00:08:02

Seeing that as a philosopher, I thought that's Plato because Francois precisely says we have the world of appearance. It's complicated. It looks intractable. It's messy. But underlying that, real reality is neat, mathematical decomposable.

Speaker 0

00:08:20

Now, feel like I should defend Chollet a little bit here, you know, because obviously we love Chollet He's not making any weird metaphysical claims. At least I don't think he is. If scientific theories actually explained reality the way it is, you would expect fewer u turns. Now, the biggest simplification in the 20 first century, the final boss of simplifications is this idea that the mind is a computer or that the mind is running a software program. So we have inputs, we have processing, we have an output. This metaphor has become so established in the collective zeitgeist that no 1 even questions it anymore. It barely even registers in our brains as a metaphor. So is it or isn't it? Isn't it a little bit weird that computation is this abstract formalism like, you know, automata that makes these state transitions. Something completely non physical. And we're describing the mind as if it is that abstract thing. That sounds a little bit weird. There are many movies made about this who talk about uploading their minds into the matrix. Neuralink talks about interfacing with your brain's software. Joscha Bach thinks that consciousness is a software program running on your brain.

¹⁰Machine Learning Street Talk interview with Francois Chollet. — Interview

Is the Mind Software? ●

Speaker 0

00:08:20

Now, feel like I should defend Chollet a little bit here, you know, because obviously we love Chollet. He's not making any weird metaphysical claims. At least I don't think he is. If scientific theories actually explained reality the way it is, you would expect fewer u turns. Now, the biggest simplification in the 20 first century, the final boss of simplifications is this idea that the mind is a computer or that the mind is running a software program. So we have inputs, we have processing, we have an output. This metaphor has become so established in the collective zeitgeist that no 1 even questions it anymore. It barely even registers in our brains as a metaphor. So is it or isn't it? Isn't it a little bit weird that computation is this abstract formalism like, you know, automata that makes these state transitions. Something completely non physical. And we're describing the mind as if it is that abstract thing. That sounds a little bit weird. There are many movies made about this who talk about uploading their minds into the matrix. Neuralink talks about interfacing with your brain's software. Joscha Bach thinks that consciousness is a software program running on your brain.¹¹

Speaker 5

00:09:30

That this is the universal, that you have these invariances in nature, that you can have patterns, that have causal power, that have the ability to reproduce themselves, that have the ability to shape reality, that are invariances that you cannot simply explain more simply by looking at what atoms are doing in space, but you have to look at these abstract patterns to make sense of them. Every other explanation is going to be more complicated in the same way as money is going to be impossibly complicated if you try to reduce it to atoms. So you have to look at these causal invariances and spirits are actually such causal invariances, they are actually disembodied, right? They are not bodies, they are not stuff in space, they are not mechanisms in the same way but they are causal mechanisms, abstract mechanisms and so we put the spirit back into nature using the concept of software. A lot of people think that's metaphorical, but I don't think it's metaphorical at all. It's the literal truth. Software is spirit.¹²

¹¹Neuralink — Company/Org Mentioned as an example of the "brain as software" metaphor in action.

¹²Joscha Bach Patreon interview.

Speaker 0

00:10:27

We're all just talking about this stuff without even batting an eyelid. Like, where's the skepticism, man? It just sounds so plausible to us. So we assume that it just kind of has to be the case.

Speaker 5

00:10:39

There's something super interesting about computers. What a computer ultimately is, is it's a causal insulator. The computer is a layer on which you can produce an arbitrary reality. For instance, the world of Minecraft. You can walk around in the world of Minecraft and it's running very well on a Mac and it's running very well on a PC and if you are inside of the world, you don't know what you're running on. And it's not going to have any information about the nature of the CPU that it's on, the color of the casing of the computer, the voltage that the computer is running on, the place that the computer is standing in in the parent universe, right, our universe. So the computer is insulating this world of Minecraft from our world. It makes it possible that an arbitrary world is happening inside of this box. And our brain is also such a causal insulator. It's possible for us to have thoughts that are independent of what happens around us. That we can envision a future that is not much tainted by the present. We can remember a past that is independent from the present in which we are. And it's necessary for us, our brain has evolved as such a causal insulator as well to allow us to give rise to universes that are different from this 1, For instance, future worlds, so we can plan for being in them.

Speaker 0

00:12:00

Bach says that money is an example of a causal pattern. It's not the ink on a banknote. It's not the electrons in your bank server. It persists across and ensconces in various physical instantiations. So paper, coins, gold, digital ledgers. And yet they say money causally affects the world. It gets you fed. It starts wars. It builds cities. He says that software is the same. A program is an abstract pattern that can run on many types of chips, maybe even neurons. And that pattern has causal power because it controls whatever substrate it's running on. The same algorithm produces the same effects regardless of what physical stuff implements it. So the invariance, that sameness across substrates is the causal mechanism, the pattern itself, at least according to Joscha He even accepts that physics is causally closed. He says that the abstract description and the physical description are 2 ways of looking at the same causal structure. Neither is reducible to the other, both are real. But I'm pretty sure Chiramoto would ask, who identifies that invariance? When we say the same algorithm runs on different chips, completely different things are actually physically happening. Right?

Different voltages, different electrons doing different things. The sameness is something that we impose. It exists in our description, not in nature. And as for the money example, money only works because of human interpretive practices. Right? If you take away the humans and their agreements, it's just paper. Right? Money is just paper. And the causal power is actually in the social substrate that participates in it. Now I think Joscha has taken a useful way of talking about complex systems and promoted it to metaphysics. And that's simplicios all over again, right? Mistaking the elegance of our descriptions for the structure of reality itself. I mean, maybe information really is more fundamental than matter, but that's another philosophical wager and we've made these bets many many times before. Just look at the history of all of this. So Descartes thought that the nervous system worked like the hydraulic automata in French royal gardens. Fluids pumping through tubes, pushing levers. That was the high-tech metaphor of his day. Later, when scientists figured out that nerves carry electrical signals, the brain became a telegraph network. Then it was a telephone switchboard, signals traveling down wires, operators routing calls. And now in our era, the brain is a computer.

Critique of Causal Patterns ●

Speaker 0

00:12:00

Bach says that money is an example of a causal pattern. It's not the ink on a banknote. It's not the electrons in your bank server. It persists across and ensconces in various physical instantiations. So paper, coins, gold, digital ledgers. And yet they say money causally affects the world. It gets you fed. It starts wars. It builds cities. He says that software is the same. A program is an abstract pattern that can run on many types of chips, maybe even neurons. And that pattern has causal power because it controls whatever substrate it's running on. The same algorithm produces the same effects regardless of what physical stuff implements it. So the invariance, that sameness across substrates is the causal mechanism, the pattern itself, at least according to Joscha. He even accepts that physics is causally closed. He says that the abstract description and the physical description are 2 ways of looking at the same causal structure. Neither is reducible to the other, both are real. But I'm pretty sure Chiramoto would ask, who identifies that invariance? When we say the same algorithm runs on different chips, completely different things are actually physically happening. Right? Different voltages, different electrons doing different things. The sameness is something that we impose. It exists in our description, not in nature. And as for the money example, money only works because of human interpretive practices. Right? If you take away the humans and their agreements, it's just paper.

Right? Money is just paper. And the causal power is actually in the social substrate that participates in it. Now I think Joscha has taken a useful way of talking about complex systems and promoted it to metaphysics. And that's simplicios all over again, right? Mistaking the elegance of our descriptions for the structure of reality itself. I mean, maybe information really is more fundamental than matter, but that's another philosophical wager and we've made these bets many many times before. Just look at the history of all of this. So Descartes thought that the nervous system worked like the hydraulic automata in French royal gardens. Fluids pumping through tubes, pushing levers. That was the high-tech metaphor of his day. Later, when scientists figured out that nerves carry electrical signals, the brain became a telegraph network. Then it was a telephone switchboard, signals traveling down wires, operators routing calls. And now in our era, the brain is a computer.

Temperature is Not a Thing ●

Speaker 0

00:12:00

Bach says that money is an example of a causal pattern. It's not the ink on a banknote. It's not the electrons in your bank server. It persists across and ensconces in various physical instantiations. So paper, coins, gold, digital ledgers. And yet they say money causally affects the world. It gets you fed. It starts wars. It builds cities. He says that software is the same. A program is an abstract pattern that can run on many types of chips, maybe even neurons. And that pattern has causal power because it controls whatever substrate it's running on. The same algorithm produces the same effects regardless of what physical stuff implements it. So the invariance, that sameness across substrates is the causal mechanism, the pattern itself, at least according to Joscha. He even accepts that physics is causally closed. He says that the abstract description and the physical description are 2 ways of looking at the same causal structure. Neither is reducible to the other, both are real. But I'm pretty sure Chiramoto would ask, who identifies that invariance? When we say the same algorithm runs on different chips, completely different things are actually physically happening. Right? Different voltages, different electrons doing different things. The sameness is something that we impose. It exists in our description, not in nature. And as for the money example, money only works because of human interpretive practices. Right? If you take away the humans and their agreements, it's just paper. Right? Money is just paper. And the causal power is actually in the social substrate that participates in it. Now I think Joscha has taken a useful way of talking about complex systems and promoted it to metaphysics. And that's simplicios all over again, right? Mistaking the elegance of our descriptions for the struc-

ture of reality itself. I mean, maybe information really is more fundamental than matter, but that's another philosophical wager and we've made these bets many many times before. Just look at the history of all of this. So Descartes thought that the nervous system worked like the hydraulic automata in French royal gardens. Fluids pumping through tubes, pushing levers. That was the high-tech metaphor of his day. Later, when scientists figured out that nerves carry electrical signals, the brain became a telegraph network. Then it was a telephone switchboard, signals traveling down wires, operators routing calls. And now in our era, the brain is a computer.

Speaker 6

00:14:40

To be precise about what we mean by physical. And everything has to be physical because even GitHub, you know, has to store its data in some sort of hard drive or magnetic field or whatever technology, but it's not storing it in in nothingness. You know? So so knowledge information always has this form of physical embodiment. Now, I think we tend to think about it as non physical because it is a thing that is not a thing, which is the same as temperature. You wake up, you look at your phone, and you see the temperature, and you decide how you're gonna dress, and nobody has any doubt that temperature is something that can be measured. But it took about like 2000 years for us, you know, as a species to figure out, you know, what temperature was and the fact that it could be measured. And there were 2 fundamental difficulties that I would say made it difficult for us to understand, you know, temperature. The first 1 is that first people thought that hot and cold were 2 separate things. Okay? So that temperature was like a mixture of the 2. It's like when you make green out of blue and yellow. Okay? And it took a while for people to understand that cold was the absence of heat and not that cold and heat were 2 different quantities that were tempered together, that were mixed. So temperature actually mix means mixture, not, you know, like what we now mean by temperature. The other thing that was very difficult to understand is that people thought that temperature was a thing, was some sort of fluid that grabbed onto things. So let's say if you had a steel rod that is hot, is that steel rod kind of like has this sort of invisible fluid that is heat and they had good reasons to believe that it was an invisible fluid because it could flow, let's say, you could connect that rod to something that was cold and that cold thing was going to warm up because that fluid was going to be flowing in that direction and so forth. So, they thought that it had a physicality as a thing. A brilliant Englishman, Jowell, basically figures out that that is not the case, that, you know, temperature is not a thing. And the way that they do it is through this observation which I don't know if you know how cannons used to be built, you know. So, if you just grab a piece of sheet metal and you make it into a cylinder and you try to

make a cannon out of that, the moment exactly that you that you shoot the cannon, that's gonna open up like a flower in a cartoon. You know, like like, you know, like a Looney Tunes type of situation. So what they would do is they would make these solid, you know, cylinders of metal and they would bore a hole in it, you know, to create the cannons. And boring those holes released an enormous amount of heat. So Joule thought, well, how come all of that heat is there It's like an infinite amount of heat if I continue to bore a hole in a piece of metal an infinite amount of time, I'm not gonna it cannot be a thing then. And that, you know, leads him to realize that temperature is actually something that has to live in things, but it's not a itself. It's related to the kinetic energy of the particles in the thing. But it's not a thing itself. It doesn't have its own particle. There isn't kind of like a temperature particle. Temperature is kind of like a property that matter has and that holds onto things. Knowledge is similar, you know, in that it holds onto you and to me, you know, and to the collective to exist. But it doesn't have kind of like a physicality in itself. But it always exists in some sort of physical medium or substrate. So, that sense, it's always going to be physical. No matter how virtual it gets, it has maybe a different type of physicality. But even electromagnetic waves that are transmitting, you know, data from your WiFi router to your laptop are technically a physical embodiment.¹³

Speaker 0

00:18:10

Now, spoke with professor Luciano Floridi a few years ago and it was actually 1 of my favorite ever episodes of MLS I I think very highly of him, which is why we're gonna show some clips of him in in this show because it's very apropos. But this is what he had to say about it.

The Ship of Theseus & Ontology ●

Speaker 0

00:18:10

Now, spoke with professor Luciano Floridi a few years ago and it was actually 1 of my favorite ever episodes of MLS I I think very highly of him, which is why we're gonna show some clips of him in in this show because it's very apropos. But this is what he had to say about it.

¹³Cesar Hidalgo, MLST interview. — Person

Speaker 7

00:18:24

Ontology on the other hand is how we structure the world in the sense that we think that that's the way it is. With the kind of eyes we have and the kind of light around the world that those are the colors we we perceive. But certainly a world full of colors is the world which I take it to be the world. That's my ontology. Re anthropologizing means changing some of that particular nature. Allow me a distinction so I hope it's not too confusing. Reality in itself, call it system. Description of reality as we perceive it, enjoy it, conceptualize it, live through. Model of the system. Ontology to me is the ontology of the model, it's not the metaphysics of the system. I hope I haven't, you know, made a complete mess here. Okay? So metaphysics, nomenon, system, whatever the source of the data that we get, fantastic. The data don't speak about the source. The music of the radio is not about the radio. But there is a radio. Of course, the music is what we perceive. Music has its own ontology structure, etcetera. The model. The model is at that point what we enjoy. Why did the digital revolution has changed the the nature of the world around us? Not metaphysically, but ontologically, so the re ontologize. Because some of the things that we have inherited from modernity, a sense of the world that is now being restructured, and a certain understanding of the world, so re epistemologizing as well of their world. We go back to this temptation of talking about reality as if it were something that we need to grasp, catch, portray, hook, spears. When in fact, the the way I prefer to understand it is as malleable, understandable in a variety of ways, something that provides constraints. It doesn't mean that you can interpret in any possible way, but leaves room for different kind of interpretations. So if the flow of data that come from whatever is out there, and again I'd rather be sort of agnostic about it, can be modeled in a variety of ways. 1 way is to, especially 20 first century given the technology we have, etcetera, to interpret that as, you know, an enormous computational kind of environment. It's perfectly fine as long as we don't think that there is our right metaphysics. It's the correct ontology for the 20 first century. Now, this is not relativism because on the other hand, different models of the same system are comparable depending on why you're developing that particular model. And let me give you a completely trivial example. Suppose you ask me whether that building is the same building. That question has no real answer because it depends on why you're asking that question. If your question is asked because you want to have directions, I'm gonna say, oh, yeah. That's the same building, sir. The same building? Yeah, absolutely. No. Go there, turn left, no traffic lights. If your question is like same function, I say, no, it's a completely different building. It was a school, now it's a hospital. Next question. So is it or is it not the same? That, that question is the mistake. An absolute question that provides no interface or computer scientists call level abstraction chosen for 1 partic-

ular purpose so that I can compare whether an answer is better than another. Let me crack a joke for the philosophers who might be listening to this. Tazeel Shep. Is it the same or it's not the same? Who is asking? Why? Because if it is the tax man, the tax man, you're doomed man. I mean, there is no way you can play any, oh, I change every plank that you're gonna pay their tax. It's the same ship. I don't care. But if it is a collector, that ship is worth 0. You change all the planks? You must be joking. Is worthless. So is it or is it not the same? Depends on why you're asking that particular question. Tell me why and I can give you the answer. Now why, in other words, no frame within which we have chosen the interface that provides the model of the system, no potential answer. So the question is like, is the universe a computational gigantic, yes or no, meaningless? Is it worth modeling the universe as a gigantium for the purpose of making sense of our digital life? Oh, yes. Definitely. Because we are informational organisms. So metaphysics. No. I meant in the 20 first century the best way of understanding human beings today is as informational organisms. Last century we thought that biologically not made much more sense. A lot of water and sprinkle a little bit of extra and so on. Mechanism, big car time, etcetera. Not absolute answers, not relativistic answers, but relational answers. The relation between the question, the purpose and the actual answer. But it takes 3, not 2.¹⁴

Speaker 0

00:23:33

So the computational model isn't literally true, but it's useful. The mistake is forgetting that it's a model. So the early cybernetics guys, so Wiener and McCulloch and Pitts they knew that they were working with analogies. McCulloch and Pitts wrote their famous paper showing that neurons could theoretically work like logic gates. Now, weren't claiming neurons actually were logic gates, but they were using it as a kind of functional description. But somewhere along the way, the metaphor hardened. A lot of neuroscientists today don't say that the brain is like a computer, they say it is 1 and the metaphor became the thing itself. Now Chirimuuta, borrowing from Whitehead by the way she said that this is the fallacy of misplaced concreteness. This is another 1 of those leaky abstractions I was talking about. By the way, there's a great book called *The Brain Abstracted* by Mazviita Chirimuuta I interviewed her recently. And she said that 1 of the most pervasive myths in neuroscience is that we use these leaky abstractions and idealizations to talk about cognition. And usually, it's using the most recent technology at the time. So, you know, a few hundred years ago, we were describing the brain in terms of pulleys

¹⁴Luciano Floridi, MLST interview.

Metaphors Hardening into Reality •

Speaker 0

00:23:33

So the computational model isn't literally true, but it's useful. The mistake is forgetting that it's a model. So the early cybernetics guys, so Wiener and McCulloch and Pitts they knew that they were working with analogies. McCulloch and Pitts wrote their famous paper showing that neurons could theoretically work like logic gates. Now, weren't claiming neurons actually were logic gates, but they were using it as a kind of functional description. But somewhere along the way, the metaphor hardened. A lot of neuroscientists today don't say that the brain is like a computer, they say it is 1 and the metaphor became the thing itself. Now Chirimuuta, borrowing from Whitehead by the way she said that this is the fallacy of misplaced concreteness. This is another 1 of those leaky abstractions I was talking about. By the way, there's a great book called *The Brain Abstracted* by Mazviita Chirimuuta I interviewed her recently. And she said that 1 of the most pervasive myths in neuroscience is that we use these leaky abstractions and idealizations to talk about cognition. And usually, it's using the most recent technology at the time. So, you know, a few hundred years ago, we were describing the brain in terms of pulleys¹⁵¹⁶

Speaker 8

00:24:39

and Pulleys and levers, yes.

Speaker 0

00:24:41

That's right. And and you know, and then it was, you know, as a prediction machine, as a computer and all this kind of Yeah. At the end of the this is an example of a these are grounded things that we understand. They're really good models because we can both talk about computers. We both know what computers are. But the brain doesn't work like that in any sense. Jeff Beck put it even more bluntly when we spoke.

¹⁵A Logical Calculus of Ideas Immanent in Nervous Activity — Paper The seminal 1943 paper by McCulloch and Pitts comparing neurons to logic gates.

¹⁶Science and the Modern World — Book Reference to Whitehead's "fallacy of misplaced concreteness".

Speaker 9

00:24:59

It will always be the case that our explanation for how the brain works will be by analogy to the most sophisticated technology that we have. Is that how's that for a non answer? Right? So so, you know, you know, a couple thousand years ago, right, how'd the brain work? Was like levers and pulleys, man. I mean, duh. Don't be ridiculous. Why? That was the you know, at some point in the Middle Ages, it became humorous. Right? Because fluid dynamics was like the, you know, was the kind of techno you know, the technology that was like the most advanced or technology took advantage of of water power was like the most advanced technology that we had. Now, the most advanced technology is computers. So duh, that's exactly how the brain works.¹⁷

The Illusion of AGI Inevitability ●

Speaker 0

00:25:41

Now here's something that kind of bugs me. Right? You go into any AI conference or you drink from the well of San Francisco by spending too much time on Twitter and you develop this mindset that AGI is inevitable. You start feeling the AGI and you'd be forgiven for thinking this because I've been using Claude code and my god, I feel that there's been more interesting stuff happening in the world of software development in the last 6 months than there has been in the previous 20 years. This this technology is genuinely amazing but it is automation technology. It's it's not really intelligence, which means it's really as good as your ability to specify and supervise and delegate to the system, but it is absolutely amazing. But why do we have this view?

Speaker 2

00:26:24

It's not an argument that AI is impossible so much as why does it seem so possible, so inevitable to people and saying that what I'm arguing is that if you look at the history of the development of the life sciences of psychology, there are certain shifts towards a much more mechanistic understanding of both what life is and what the mind is. Which are very congenial to thinking that whatever is going on in animals like us in terms of the processes which lead to cognition, they are just mechanisms anyway. So why couldn't you put them into an actual machine and have that actual machine do what we do? With all of that mechanistic history in the background, AI could seem very inevitable but if that mechanistic hypothesis is actually wrong, then these claims for the inevitability of a biolog-

¹⁷Jeff Beck, MLST interview.

ical like AI would not actually be well founded, but we could be subject to a kind of cultural historical illusion that this is just going to happen.

Speaker 0

00:27:34

Cultural historical illusion. I've been thinking about that phrase. Maybe our confidence says more about what we've inherited intellectually than about how minds actually work. Now another thing that Mazviita has inspired me to think about a lot is the difference between prediction and understanding. Indeed, when I interviewed the Nobel Prize winner John Jumper at Google DeepMind a couple of months ago, this was the question I asked. And he had quite an interesting way of distinguishing those 2 things. It's almost like it's at any point learning how to refine and optimize the structure.

Prediction vs. Understanding •

Speaker 0

00:27:34

Cultural historical illusion. I've been thinking about that phrase. Maybe our confidence says more about what we've inherited intellectually than about how minds actually work. Now another thing that Mazviita has inspired me to think about a lot is the difference between prediction and understanding. Indeed, when I interviewed the Nobel Prize winner John Jumper at Google DeepMind a couple of months ago, this was the question I asked. And he had quite an interesting way of distinguishing those 2 things. It's almost like it's at any point learning how to refine and optimize the structure.¹⁸

Speaker 10

00:28:06

Okay. So I think we should distinguish 3 things. Predict, control, understand

Speaker 8

00:28:10

Yeah.

Speaker 10

00:28:10

First. So predict means that you say, I'm gonna do a thing. What am I gonna what will be this value of my machine? What will appear on my computer screen in the future? That is predict. Control is I want to measure this thing in the future, and I want it to come out 17. Right? That's control. Understand is a lot like pre-

¹⁸Highly Accurate Protein Structure Prediction with AlphaFold — Paper Referencing John Jumper's work at DeepMind (Nobel Prize winning).

dict, except there's a human in the loop. Understand means that I have such a small collection of facts that you will predict, and you will do it with facts that I can communicate to another human in kind of this compact fix fits on an index card. That's almost understand. And so I think these machines let us predict. They let us control. We have to derive our own understanding at this moment. Right? We can experiment now on the artifact. We can look at the 200,000,000 predicted structures, not just the 200,000 experimental structures in order to help us understand, but it doesn't do the act of understanding for us, it does the act of predict and maybe control.

Speaker 0

00:29:14

The problem is these 2 goals actually pull against each other.

Speaker 2

00:29:18

I think we're at this moment in science now because we have these tools like LLMs for language and convnets and visual neuroscience are being used as predictive models of neuronal responses which don't have that mathematical legibility that originally, so when I was trained in the field that people aspired to have. And so you have this possible conflict, you can either pursue that goal of understanding or you can pursue the goal of prediction but it seems like you can't have both at the same time.

Speaker 0

00:29:55

Now on the 1 hand people go into neuroscience because they want to understand the mind. They want that feeling where something clicks and you suddenly get how it works. That's what drew Chiromuta to the field in the first place. That's what keeps people up late at night reading papers. But on the other hand, there's just prediction. Building tools that work. If your model forecast data accurately, maybe you don't care whether it's true in some deeper sense. So LLMs are getting unreasonably good. They are winning math Olympiads. They are I mean, as of last week actually, GPT 5.2 apparently discovered a new theory. Well, it's it's solved 1 of these problems that Terence Tao had on his website This is insane. But does it actually understand anything? And does it matter if it does or doesn't, as long as it works? Chomsky had an amazing commentary on this a few years ago when we spoke and I think it's still as relevant today as it was then.¹⁹

¹⁹Terence Tao's Blog — Person Mentioned in context of AI solving math problems he posted.

Speaker 11

00:30:48

Suppose that I submitted an article to a physics journal saying, I've got a fantastic new theory and accommodates all the laws of nature, the ones that are known, the ones that have yet to have been discovered. And it's such an elegant theory that I can say it in 2 words. Anything goes. Okay. That includes all the laws of nature, the ones we know, the ones we do not know yet, everything. What's the problem? The problem is they're not going to accept the paper because when you have a theory, there are 2 kinds of questions you have to ask. Why are things this way? Why are things not that way? If you don't get the second question, you've done nothing. GPT 3 has done nothing.²⁰

Speaker 0

00:31:32

Classic Chomsky. So maybe theories are overrated. Maybe prediction is enough. But Chirimuuta worries about that trade off Right? When you give up on understanding, you don't know when your tools will break. You're stuck with black boxes. They work until they don't and you won't see it coming when they don't. I spoke with philosopher Anna Ciaunica about this recently and she had a beautiful way of describing it.

Speaker 12

00:31:56

Suppose you want to climb a mountain and you arrive on the top of the mountain, what's the argument to say that actually it's only when you're on the top of the mountain that that what's the climbing on the mountain is? I mean, you cannot really arrive on the top of the mountain if you don't do the first step. Every single step matters. First step is as important as the last 1. Actually, are more conscious when we take the first steps in climbing the mountains than when we are on the top of the mountains and we have all this like full blown capacities and sometimes we shut ourselves in the legs.²¹

²⁰Noam Chomsky, MLST interview. — Interview

²¹Anna Ciaunica — Interview

Climbing the Mountain vs. The Helicopter ●

Speaker 12

00:31:56

Suppose you want to climb a mountain and you arrive on the top of the mountain, what's the argument to say that actually it's only when you're on the top of the mountain that that what's the climbing on the mountain is? I mean, you cannot really arrive on the top of the mountain if you don't do the first step. Every single step matters. First step is as important as the last 1. Actually, are more conscious when we take the first steps in climbing the mountains than when we are on the top of the mountains and we have all this like full blown capacities and sometimes we shut ourselves in the legs.²²

Speaker 0

00:32:28

And of course, I brought this up when I debated Mike Israel And the biggest misconception in all of AI, what all of the folks in San Francisco believe in is this philosophical idea called functionalism. That we're walking up the mountain and when we get to the top of the mountain, we have all of these abstract capabilities like being able to reason, play chess. But that disregards that the path that you took walking up the mountain is very important and not only the path, the physical instantiation, the stuff that the mountain is made out of. So Mike's view is that if something produces intelligent outputs, why does the substrate matter? Silicon, neurons, it doesn't make any difference. It's all information processing. Needless to say, he pushed back hard.

Speaker 8

00:33:10

You can climb mountains, you can touch stuff, but you never truly embodied experience anything if you push on that philosophical button hard enough because you can always abstract out to like, these are just neural network pings from groups of neurons. And so, you don't truly deeply know anything in some kind of weird philosophical way because it's just neural network calculus all the way down. You know, you climb the mountain, that's cool. Helicopter can climb a mountain much better than you. Does not have the ability to reason and abstractly and plan and predict things at all.²³

²²Anna Ciaunica — Interview

²³Mike Israel debate on functionalism.

Speaker 0

00:33:42

So it's possible that what you can do or how you can function isn't the whole story. Or maybe if that's wrong, we should just start using helicopters. So individual minds are limited. But what about collective minds? What about humanity as a whole? We've built this incredible thing over centuries. Right? Libraries, universities, Wikipedia, an expanding store of knowledge that no single person could ever hold. Doesn't that escape our individual limitations? So there's this dream of universal knowledge, accessible anywhere, perspective free.

Speaker 6

00:34:15

There is a tacit, an implicit idea there that knowledge is something that something can have while my view is that knowledge is a much more collective phenomenon. Okay? So and it's not something also that you can put in something like a book. In my opinion, the book doesn't have knowledge. The book is an archival record of some ideas that I was able, you know, to put together in a nice structure, but you cannot have a conversation with the book. Knowledge only can go to work when it's embodied. You cannot throw, like, you know, a bunch of engineering manuals and cement into a gorge and expect to get a bridge because the books don't have knowledge, teams have knowledge, organizations have knowledge.

Haptic Realism & The Limits of Knowledge ●

Speaker 6

00:34:15

There is a tacit, an implicit idea there that knowledge is something that something can have while my view is that knowledge is a much more collective phenomenon. Okay? So and it's not something also that you can put in something like a book. In my opinion, the book doesn't have knowledge. The book is an archival record of some ideas that I was able, you know, to put together in a nice structure, but you cannot have a conversation with the book. Knowledge only can go to work when it's embodied. You cannot throw, like, you know, a bunch of engineering manuals and cement into a gorge and expect to get a bridge because the books don't have knowledge, teams have knowledge, organizations have knowledge.

Speaker 0

00:34:53

Yes. Knowledge is social. Communities accomplish what individuals can't. But collective knowledge is still knowledge from somewhere. This matters, right? It's shaped by particular questions, particular tools and particular blind spots.

Speaker 2

00:35:08

I think 1 of the interesting things about this phenomenon, not only of LLMs but the internet as this idea that it's the repository of all human knowledge is that it goes along with this idea almost that knowledge doesn't have to be perspectival, it doesn't have to be like of a place of a community, it kind of can float free of the situation in which this knowledge was acquired. That's kind of the aspiration of these ideas sort of a universal repository of knowledge. But what this perspectivalist position actually sort of points us to is actually knowledge is inherently of a place, of a community. We acquire knowledge not by being like completely open minded to everything that's possible to know, but actually by sort of narrowing our view, discounting possibilities actually is what allows you to pursue a line of inquiry and actually pin down some information about say the natural world which is humanly achievable. So the contrast I'm trying to make here is between a view which says that knowledge is perspectival. It's inherently from a human point of view, which means that it's inherently finite. We cannot aspire to this sort of universal free floating knowledge because as finite human beings, we can only achieve knowledge of the world through recognizing our limitations. And this notion of like, you can have non perspectival knowledge like everything in the internet based on like all of the different possible perspectives all blended together. This somehow gives us a god's eye view. LLMs aspire to be this like every person voice, but it's precisely because they don't have a particular socialization into a finite community that they're not reliable, that we can't pin them down to actually what would be a kind of honest trustworthy perspective.

Speaker 0

00:37:19

So Chirimuuta has this idea that she calls haptic realism Most of the philosophy of science treats knowledge like vision. You stand back and you observe reality from a distance. She thinks it's more like touch.

Speaker 2

00:37:32

We just look around, we absorb how things are, our knowledge is sort of entirely objective, it's almost like a God's eye view on reality. But if you think that scientific knowledge in particular is more kind of touch like, you can't ignore the fact that we sort of run into things. We have to pick things up, engage with them, ultimately change them in order for us to acquire knowledge of them. So you cannot discount the fact that we're kind of meddling with things in the process of bringing about our knowledge.

Speaker 0

00:38:05

Neuroscientists are more than passive observers of brains. They poke them, they prod them, they stimulate them, they model them. And in doing that, they change what they find. The patterns that emerge are real, but they're also partially created by the process of investigating itself. It takes all the messiness of biological cognition and it reduces it to 1 imperative, minimize free energy. Everything else supposedly follows from that. Now Simplicius loves this. I mean, finally, the simple truth, the 1 principle to explain it all. But ignorantio says, wait a minute. The math is elegant, the framework is unified, but does that mean it's captured what brains actually are? Or did we just build another beautiful simplification and started forgetting that it was a simplification? So Chirimuuta said to me that we should ask different questions right? Not is this true, but what does this help us do? What does this light up? What does it leave in the darkness? And the other thing of course is that we are finite biological creatures. Right? We There are limits to our cognition. And Chomsky spoke about this fascinating concept of a cognitive horizon when we when we chatted with him.

Speaker 11

00:39:18

If we are organic creatures, we're going to be like other organic creatures, and that there are bounds to our cognitive capacities. So for example, a rat can be trained to run pretty complicated mazes, but it can't be trained to learn a prime number maze. Turn right at every prime number, it just doesn't have the concept. And no matter how much training you do, you're not going to get anywhere. Well, I suspect there's reasons to suppose we're like rats. We have capacities, we have a nature, we have a structure, they yield all sorts of extensive range of things that we can do, but they probably impose limits. And I think we could even make some guess about what these limits are.

Speaker 0

00:39:59

So our best theory is they bump up against the walls of the limits of our cognition, of our cognitive horizon. And maybe that's fine, but maybe even knowledge of where the walls are is useful in of itself. Science makes things simple and it's not a flaw. Right? Without simplification, we'd have nothing. You can't study everything at once. But simplification has risks. Right? You forget your model is a model, you mistake elegance for truth and you think you found solid ground when really you're just building another floor. So look at Opus 4.5, right? Foundation models today, they are artifacts of staggering complexity. We've trained them on everything humans have ever written. We treat their outputs like they came from somewhere authoritative, somewhere outside of us, somewhere that knows. But the knowing was ours all along, right? Just compressed, refracted, reflected back to us from the silicon. Whether that reflection captures the actual thing, that is a question that we're barely starting to ask. You can use powerful frameworks like the free energy principle, but just remember they're frameworks, right? They're tools for building, they're not the final word. So the brain is not a hydraulic pump, It's not a computer. It's not a telephone network. It's probably not a free energy minimizer either. I mean, least not in some like literal way. What the brain actually is, we will only ever caps glimpses of. Right? That is through our limited instruments and theories. And that's okay because that's what it means to be finite. So Chirimuuta had this amazing example from Greek mythology called Proteus Right? And if you could pin him down, he'd have to answer your question correctly. But if you let go and you let him get away, then he would shape shift and shape shift. Nature is like that. Right? You can pin it down, you can ask questions, but it's always perspectival. As soon as you let go, there's always a myriad of other perspectives that can be interpreted from reality. Karl Friston's Wood Lice, they were doing something very similar right? So slow down in the sun, move faster in the shade. But Friston isn't a woodlouse and neither are you.

References

- [1] **Karl Friston (UCL)**
<https://profiles.ucl.ac.uk/1236-karl-friston>
Person Cited as the originator of the Free Energy Principle.
- [2] **The Brain Abstracted**
<https://mitpress.mit.edu/9780262548045/the-brain-abstracted/>
Book Prof. Chirimuuta's book discussed as the primary influence on the episode's theme.
- [3] **The Brain Abstracted Patreon interview**
<https://www.patreon.com/posts/brain-abstracted-124479979>
Interview.
- [4] **David Krakauer's presentation on intelligence.**
<https://www.youtube.com/watch?v=dY46YsGWMic>
Interview
- [5] **On Learned Ignorance**
<https://www.amazon.com/Nicholas-Cusa-learned-ignorance-translation/dp/0938060236>
Book Refencing Nicholas of Cusa's concept of "docta ignorantia".
- [6] **Francois Chollet**
<https://fchollet.com/>
Person Cited for his "Kaleidoscope Hypothesis" regarding intelligence and abstraction.
- [7] **Machine Learning Street Talk interview with Francois Chollet.**
<https://www.youtube.com/watch?v=JTU8Ha4Jyfc>
Interview
- [8] **Neuralink**
<https://neuralink.com/>
Company/Org Mentioned as an example of the "brain as software" metaphor in action.
- [9] **Joscha Bach Patreon interview.**
<https://www.patreon.com/posts/joscha-bach-deep-141884561>
- [10] **Cesar Hidalgo, MLST interview.**
<https://www.youtube.com/watch?v=vzpF0JRteeI>
Person
- [11] **Luciano Floridi, MLST interview.**
<https://www.youtube.com/watch?v=YLNGvvgq3eg>

[12] A Logical Calculus of Ideas Immanent in Nervous Activity

<https://link.springer.com/article/10.1007/BF02478259>

Paper The seminal 1943 paper by McCulloch and Pitts comparing neurons to logic gates.

[13] Science and the Modern World

<https://amazon.com/dp/0684836394>

Book Reference to Whitehead's "fallacy of misplaced concreteness".

[14] Jeff Beck, MLST interview.

<https://www.youtube.com/watch?v=9suqiofCiwM>

[15] Highly Accurate Protein Structure Prediction with AlphaFold

<https://www.nature.com/articles/s41586-021-03819-2>

Paper Referencing John Jumper's work at DeepMind (Nobel Prize winning).

[16] Terence Tao's Blog

<https://terrytao.wordpress.com/>

Person Mentioned in context of AI solving math problems he posted.

[17] Noam Chomsky, MLST interview.

<https://www.youtube.com/watch?v=axuGfh4UR9Q>

Interview

[18] Anna Ciaunica

<https://www.patreon.com/posts/dr-anna-ciaunica-144509970>

Interview

[19] Mike Israetel debate on functionalism.

https://www.youtube.com/watch?v=4yYcN_mFi18