Rajesh Rao reflects on predictive brains, neural interfaces and the future of human intelligence

Twenty-five years ago, Rajesh Rao proposed a seminal theory of how brains could implement predictive coding for perception. His modern version zeroes in on actions.

18 DECEMBER 2024 | by PAUL MIDDLEBROOKS

This transcript has been lightly edited for clarity; it may contain errors due to the transcription process.

[music]

Rajesh Rao
The goal of the brain, or at least one goal of the brain is to learn a generative model of the world, an internal model of the world. It's constantly making these predictions or doing essentially hypothesis testing. It's essentially generating hypotheses and it's trying to then match those hypotheses with what's coming in through the census. When there are deviations or mismatches, those are called prediction errors, and those errors are what are used to then update the hypothesis.

If you want to have a very rich way of modeling the world, then what you want to do is have the higher level modulate the dynamics of the lower level. You need to change the function being computed at the lower level depending on the task. We say, "Why do we use these primitive modes of interacting with the brain so EEG and TMS? Let's build a direct brain-to-brain communication system with them."

[music]

Paul Middlebrooks
This is “Brain Inspired,” powered by *The Transmitter*. Hello, I am Paul. Today I am in conversation with Rajesh Rao. Rajesh is a distinguished professor of computer science and engineering at the University of Washington where he also co-directs the Center for Neurotechnology. Back in 1999, Raj and Dana Ballard published what became quite a famous paper, which proposed how predictive coding might be implemented in brains.

What is predictive coding? You might be wondering. It is roughly the idea that your brain is constantly predicting or at least back then it was roughly the idea that your brain is constantly predicting incoming sensory signals. It generates that prediction as a top-down signal that meets the bottom-up sensory signals. When those two signals meet the top-down prediction with the bottom-up sensory input, the brain computes a difference between the prediction and the actual sensory input.

That difference, that error is sent back up to the "top" where the brain then updates its own internal model to make better future predictions. That was 25 years ago, and it was focused on how the brain handles sensory information. Raj just recently published and proposed an update to the predictive coding framework, one that incorporates perception and action, suggesting how it might be implemented in the cortex specifically which cortical layers do what. Something that he calls active predictive coding.

We discussed that new proposal. We also talked about his engineering work on brain-computer interface technologies like brain net which basically connects two brains together and like neural co-processors, which use an artificial neural network as a prosthetic essentially, that can do things like enhance memories, optimize learning, and help restore brain function after strokes, for example. Finally, we discuss Raj's interest and work on deciphering an ancient Indian text, the mysterious Indus script.

All right, a link to all of the work that we discuss in the show notes at braininspired.co/podcast/201, 2-0-1. I'm in the process of setting up our next live discussion for Patreon supporters. If you're a Patreon, check your Patreon messages soon or check in on the “Brain Inspired” discord. Thank you for listening. Here's Rajesh.

[transition]

Paul Middlebrooks
Raj, by the way, this was the first long motorcycle ride I've taken since I got a motorcycle, and I'm in a hotel, so-

Rajesh Rao
Fantastic.

Paul Middlebrooks
-I may not be on the top of my game, not that there is a top of my game, but anyway.

Rajesh Rao
You look great, and I think it should be interesting to see what interesting questions come out of you having done that long motorcycle ride.

Paul Middlebrooks
I actually wanted to start with-- Well, maybe you won't tell me this. What year were you born?

Rajesh Rao
Oh, it's on the internet. It's on my Wikipedia page. It's 1970.

Paul Middlebrooks
'70. When it was the '60s and you were already good at academia perhaps before you were born, I understand that you excelled in early years in your schooling. Here's what I want to get to, I would like you to tell me the origins of the predictive coding story. When it was in the '60s, before you were born and you were thinking about predictive coding, how did you start thinking about that?

Rajesh Rao
[laughs] That's right. It was in my previous life. I've been reincarnated now to bring up predictive coding. Basically, the story is essentially that I grew up in India in Hyderabad and did my high school there. I got a chance to come to the US as an 11th grader because I got a rank in the Indian central science exam that they have. As part of that rank, they were able to send me with another group of people to do some research. Basically, I got hooked on to research right in 11th grade.

I went to University of Maryland, did some research. I would call that, put that in codes, research and superconductivity. It was time of high temperature superconductivity at the University of Maryland. I got a taste of research then and then took the SAT exam. Never applied to any universities. I was just waiting for universities to contact me, thinking if my SAT score is high enough, somebody will contact me. Obviously, that never happened [laughter] because that's the way it works in the IIT exams in India. You write an exam. If the rank is high enough, you get to go to an IIT.

Paul Middlebrooks
They will contact you.

Rajesh Rao
Yes. They will let you know. At least, there will be a ranking of the announcement and things like that of who got in, who didn't so I was under that impression. Then luckily for me, there were a couple of universities that did write to me saying, "Hey, look, we have scholarships for international students." One of them was in Texas, a tiny university called Angelo State University. Another one was in Alaska, University of Alaska Fairbanks. Given my preference for weather, you can imagine where I went. I had some family in Texas as well. I went to Texas for my undergrad.

Paul Middlebrooks
I grew up in Texas, so I was curious about that aspect of your development and how you found Texas in general. That was in the, what, early '80s?

Rajesh Rao
No, not early '80s. I'm not that old. I'm old but I'm not that-- It was late '80s.

Paul Middlebrooks
Late '80s.

Rajesh Rao
It was basically late '80s and it was the time before the internet. It was interesting. It was an interesting culture shock for somebody coming from a bustling place in India to a small town in Texas with cowboys for roommates. You can imagine all the interesting things that happened, and I could talk about that over beer sometime.

Paul Middlebrooks
Yes, we'll do that over a beer. Maybe fast forward, I suppose, because I do want to buy you a beer now and talk about Texas.

Rajesh Rao
Predictive coding, right?

Paul Middlebrooks
Yes. I don't understand the origins of it. I understand Helmholtz. We don't have to go that far back. Hermann von Helmholtz with inference and predicting the best outcome, predicting what you will perceive. I know that origin story, but in terms of neurons, what is that origin story?

Rajesh Rao
I think for me and my advisor at the time, Dana Ballard, I went to University of Rochester for my PhD. I was going to do a PhD in theoretical computer science, but then I happened to meet Dana Ballard in the copy room, and he said, "Hey, I have this summer RA, so can you work with me for a summer?" I said, "Okay, I'll give that a try." Sure enough, that got me hooked to, at that time, computer vision. We were looking at this problem of how do you reconstruct objects that are behind occlusions?

The idea was let's take the representation from the visual cortex. We have these Gabor filters, these oriented filters. Now, if you can recognize objects using this representation based on the responses of these filters, can you now also reconstruct what's behind these occlusions that the objects may be behind? Then it turned out that those filters, you cannot just combine them linearly with the responses of the filters and then reconstruct an image like you would do for PCA filters, principal component analysis, or the eigenvectors because they're not--

Paul Middlebrooks
Which is linear. I'll just interject.

Rajesh Rao
Linear, yes. Even if you do a linear combination of these Gabor filters, they are non-orthogonal to each other. That's when we got into-- The idea was can we then do gradient descent on a reconstruction error cost function? You have this idea that we can reconstruct images that are based on non-orthogonal filters, the responses. Then it leads to this idea of optimization of the responses of neurons that are trying to reconstruct an image that's been presented based on these Gabor filters.

Then if you flip that and say what if we cannot just estimate the responses but also learn those filters? Then you get this really interesting idea of, at a fast timescale you do inference, which is similar to what Helmholtz was talking about which is inference of what object you're perceiving right now. You do that using this dynamical system, which is based on gradient descent. The interesting thing there, which was not very appreciated at that time, and even I didn't appreciate it as much was the equation that you get from gradient descent. You have this cost function of minimizing the prediction error or the reconstruction error.

When you do the gradient descent equations, it basically falls out of the equation that you need an architecture that has feed-forward and feedback connections. The feed-forward connections in the architecture would be carrying the prediction errors, and the feedback connections would be carrying the predictions or the reconstruction of the image.

Paul Middlebrooks
For fear of losing the audience, I'm just going to back up and we'll go Lego bricks.

Rajesh Rao
Sure, yes.

Paul Middlebrooks
What is the idea of predictive coding? Do you want me to say it and then you can correct me?

Rajesh Rao
Sure, yes, go ahead.

Paul Middlebrooks
Predictive coding is basically that you have some idea forward in the brain, somewhere forward in the brain, some idea of what you're expecting, and you're sending that backward through the brain. Those incoming sensory signals that are coming up toward forward of the world of shapes, if we're talking about vision, then they get met with this prediction, and then there's a difference between what you're expecting and what those signals are propagating forward. It's the difference between the prediction and the signals that gets propagated forward. How does that sound?

Rajesh Rao
That's great. I guess we were good at conveying the message in those papers, so I'm glad to hear that that's-- Basically, the idea is the traditional model, going back to Hubel and Wiesel was always that you had this feed-forward pass. You flash an image, you get a feed-forward pass all the way from V1, V2, V4 in the visual system, up to IT, and then you get recognition, and then you get cognition of that. If you make a decision, maybe you press a button. That's action.

Paul Middlebrooks
Back in those days, was it always voila, there's the cognition of it? If it just builds up to an abstract enough edges to a table leg, to a table, "Oh, there's a table," and it magically happened?

Rajesh Rao
That's the way it was taught, even in the AI field. The whole field of AI was partitioned into all these different there's people doing vision, there's people doing motor control robotics, there's people doing higher level AI based on logic. It was very similar to how even in neuroscience and cognitive science, we have these people focusing on particular modalities and so on. Then, for me, as somebody coming into that area of neuroscience and looking at the anatomy, it seemed really surprising that people were completely ignoring the feedback connections.

If you look at every cortical area, it's sending not just the feed forward connections, but also receiving feedback connections from a higher order area. Then if you talk to people at the time, who are really famous visual neuroscientists and ask them, "What about those feedback connections?" They're like, "No, that's maybe doing attention, it's doing something."

Paul Middlebrooks
Really? Would they really say that?

Rajesh Rao
Yes, sure. People thought that those were not really critical for the traditional perception of objects and those were maybe brought into play when you're doing something like very specific kinds of attention, maybe during sleep and things like that. A lot of people were not necessarily even acknowledging that those feedback connections could be playing a very dominant role. Predictive coding flips it on the head. It says the goal of the brain, or at least one goal of the brain is to learn a generative model of the world, an internal model of the world.

It's constantly making these predictions or doing essentially hypotheses testing. It's essentially generating hypotheses and it's trying to then match those hypotheses with what's coming in through the senses. When there are deviations or mismatches, those are called prediction errors. Those errors are what are used to then update the hypothesis. You're sending them back in the feed-forward pathways. The feed-forward pathways are actually not carrying the raw signals, but they're carrying the deviations or errors, mismatches. That was the big insight, I think. That was different from what the conventional thinking was at that time.

Paul Middlebrooks
All right, let's stick with that insight for a second. Was that insights an aha moment for you?

Rajesh Rao
I would say the idea was in the air. I was at that time reading papers by people like David Mumford, for example, who was also talking about things like that between the thalamus and the cortex and cortex to other cortical areas. There were people like James Albus who was talking about it in the context of AI, and he'd already done work, obviously in the cerebellum. That was his main contribution. Neuroscience was the model, the more obvious model of the cerebellum. He also had really interesting ideas about hierarchies in AI and in robotic controllers and things like that. Then finally, if you really go back in time, there was somebody named Donald MacKay. He's actually the father of David MacKay who wrote that book on AI information theory. He actually proposed-- He had a paper called “The epistemological problem for automata” back in the 1950s.

Even he proposed this idea that what if you can send error signals from one module to another, and then you could then build up this sort of abstract representations at multiple levels of the hierarchy. What we did was essentially take many of those ideas that have been around and then say, "Let's actually implement that mathematically. Let's see what comes out of it." Then we showed that some of the things that come out of it, if you interpret neural responses as prediction errors, is that you can explain some of these puzzling effects like end stopping, contextual modulation, orientation, contrast effects, things like that that it'd be harder to explain if you just have a feed-forward model. This actually gives you a normative explanation in terms of natural images and in terms of natural behaviors.

Paul Middlebrooks
Even if you look back at the McCulloch-Pitts articles, they knew feedback was important. Maybe you can educate me on this a little bit better, but it was almost like a nod to the fact when they were drawing their little neuron diagrams, McCulloch-Pitts neurons, one to the other, and then there'd be a feedback loop. I think they wrote about how it would be important one day, perhaps.

Rajesh Rao
Yes. I think they call them loops. They definitely add this idea that loops are pretty important. I think that at that time, there were a lot of people interested in characterizing essentially what we call dynamical systems now, but they were trying to characterize the properties of these loopy networks. I think the correspondence between that kind of work, which is very fundamental, rigorous, basic elementary work, to now trying to map that onto anatomy is I think the leap.

That's what Mumford had suggested, and that's what we also tried to suggest was that let's try to map some of those ideas now onto the cortical anatomy. The idea of there's six layers in the cortical area, there's feed-forward connections coming into the middle layers, layer four, and then there's feedback coming from the superficial deep layers. If you look at that, you have the Felleman-Van Essen hierarchy. Can we actually--

Paul Middlebrooks
That monstrosity of a diagram.

Rajesh Rao
The way that we looked at that Felleman-Van Essen hierarchy was what if we interpret that as a generative model that evolution came up with for essentially modeling the world? Once you interpret it in that sense, then it says if you have this idea that you can sample from that generative model and generate examples of what the animal faces and its interactions with the environment, then inference becomes this idea of trying to update those estimates you have, those hypotheses you have about the world.

I think the key idea there was inference as a fast update of the neural responses at the population level in all the different cortical areas. At a slower timescale, using the same mismatch or errors to then update the weights, the learning part, the synaptic plasticity part. That was happening at a much slower timescale.

Paul Middlebrooks
Your famous paper is the 1999 - that must be your most cited paper. Am I correct?

Rajesh Rao
It is, yes.

Paul Middlebrooks
You had that, and in fact, I've seen you in a talk point to it with optimism for graduate students saying, "See, stick to your ideas, and eventually perhaps they will come to fruition."

Rajesh Rao
Perhaps is a key word there.

Paul Middlebrooks
Well, perhaps. What do you think is the, oh gosh, I would say 3%? What do you think?

Rajesh Rao
I think it's hard to say. I would say one of the reasons why that paper, it was published at a time when I think there were still not the techniques you would need to really test some of these predictions. For example, these prediction error neurons. Can we look at specifically layer two and three, which is where you would expect these prediction error neurons to be because those are the feed-forward connections to the next cortical areas?

Similarly, can you look at the deeper, like layer five neurons and look at what kind of responses? Because in the predictive coding model, the deeper layers are the ones that are storing the estimates of state or in a motor responses and so on. You'd have to have the techniques, which were not available at that time. I think if you fast forward, 10 years from then, and then start looking at in the late 2000s you start to see more citations of that paper. Then now I think there's so many people trying to look at different aspects of that theory with evidence for some people against. I think it's become very, very rich area of research. I think the theory was maybe a little too early for its time at that time. I think the fact that *Nature Neuroscience*, which is where it was published, was willing to publish it as opposed to saying it's just pure speculation, which they could have done. I think is credit to them for having taken that particular paper.

Paul Middlebrooks
Occam's razor, it's theoretically great, right?

Rajesh Rao
Yes. It had an elegance to it, which I think we ourselves didn't quite appreciate as much. I think Friston actually wrote a *News & Views* on it 10 or 20 years after it was published talking about how that one paper really influenced him as a researcher-

Paul Middlebrooks
Oh.

Rajesh Rao
-to really pick up the mantle at that point, because we had actually-- I had moved on to more general models on the Bayesian Brain Hypothesis and belief propagation in Bayesian networks and things like that. Then, of course, Friston went on to do the free energy principle based on predictive coding ideas, and then more recently active inference. There's a whole bunch of work that is [crosstalk].

Paul Middlebrooks
I'm sorry to perseverate on this, but at the time, did you know that it was going to be so influential? Not know, but did you have some confidence, some sense that this is-- I know you had a sense that it was a good idea, but I have good ideas every day, and they're never good ideas.

Rajesh Rao
I think at that time, I would say, I was not too confident that it would actually be something that would inspire people to do experiments because there was a *News & Views* that was written on that paper by Koch and Poggio and it's an interesting *News & Views*. You can see them, they code a Sherlock Holmes story and then talk about how the fact that the dogs were not barking at night. The *Silver Blaze* was the story of Sherlock Holmes, is an example of predictive coding because the person was not surprising. Then the deduction is that the person who stole, I guess, the horse called *Silver Blaze*.

Paul Middlebrooks
That's a very positive review.

Rajesh Rao
It was a positive review. Before that, I was told by the editor that the previous person there has refused to write *News & Views*. It's like, "Oh, there's already people that are not necessarily going to like this." Maybe it goes against their way of thinking about how the cortex works. Like I said, there was not as many citations. There was not a huge number of people writing to me and saying, "Hey, that's a great thing." I was like, "Okay, we wanted to get it out. It's out there, let's see what happens." Then it took 10, 15 years right before it really caught on.

Paul Middlebrooks
Caught on. Were there early detractors also?

Rajesh Rao
Yes, I think there's always been detractors. There's been people even now. Basically, a lot of people don't buy into this notion of the feedback really influencing the responses and then the prediction--

Paul Middlebrooks
How can that be? Oh, specifically, with the prediction errors, that's a different. Feedback is obvious in the brain.

Rajesh Rao
Even feedback carrying predictions from a higher area. Even-

Paul Middlebrooks
That's not obvious.

Rajesh Rao
-notion of that is still debated for particular tasks. I think there's really clear evidence now that for sensory motor tasks, where the animal makes a motor movement, and then there's an inference copy that allows you to make a prediction. I think there's a lot of evidence now coming from David Schneider's lab from Georg Keller's lab and so on, showing that when there's locomotion or there's a four-limb press of a lever and there's an auditory response, you get a prediction in the auditor cortex, there's a suppression that looks like a prediction error signal there.

Definitely, for the motor, I think you see that. If it's purely sensory, which I would argue that in some sense, a purely sensory experiment typically has a fixation, and then you're flashing something, it may not necessarily be the naturalistic behavior that would elicit predictions and prediction errors. There have been some reports about these error-like responses in those situations as well.

Paul Middlebrooks
Go ahead.

Rajesh Rao
That's where I think there was initially some resistance in terms of the fact that there wasn't at that time any evidence for feedback conveying predictions and feed-forward conveying prediction errors except for these contextual effects like these extra classical receptive field effects, and even that some people are explaining, but just through lateral inhibition which could be one way to explain that.

Paul Middlebrooks
What is the lesson here? It's not that predictive coding is universally accepted still, but now the technology is better, and we can have better and better measurements of hypotheses generated by a predictive coding framework. One lesson is keep going grad students. It might get better, but for a-- The jury is still not out, I guess, right?

Rajesh Rao
Yes, I would say there's evidence for some parts of the theory. I think as a theorist, I feel like our job, all the theoreticians out there, our job is to propose theories that are explicit enough in terms of connecting to the anatomy-

Paul Middlebrooks
Specific enough.

Rajesh Rao
-that they can be shot down. They can be falsified. We basically want to have the ability to falsify theories. In the process of that, if it generates some new data that then really sparks a new theory, and leads the field forward, I think that's the job. We should not be afraid to propose these theories as long as they're sensible, they match the existing data which is what we were trying to do at that time was saying, "We have these puzzling responses called end stopping or contextual effects. We have the anatomical evidence saying there's feedback connections. Let's propose this theory and then let's see what happens if you start doing experiments."

I would say the same thing now. Look at what's out there, what people have collected in terms of the data and look at the computational aspects also. What is the brain really trying to do? Let's look at evolution.

Paul Middlebrooks
That's what I was about to ask you. Do you see predictive coding or the family of predicting coding approaches as some sort of general theory about what the brain is doing? There's the free energy principle now. There's Jeff Hawkins. Everything is prediction, these things go back years, obviously, and everyone points to Helmholtz and there was someone before him.

Rajesh Rao
There's ChatGPT and transformers, which are also based on prediction, right?

Paul Middlebrooks
That's right. We love them.

Rajesh Rao
[laughs] I would say it's just part of the story. Prediction is important because animals have to predict to survive. Ultimately, it goes up. It's this idea of at certain point in evolution, the brain start building models of the world to be able to predict, to anticipate. I think that's an integral part of if you're learning a model of the world that comes with this benefit of you can use that to generate hypotheses, you can use that to generate predictions.

That, I think, is an integral component, this idea of world models or predictive models of intelligence, either artificial or natural. It gives you a lot of benefits and it wouldn't be surprising if evolution came up with that way of handling the uncertainties of the world and to compensate for delays as well from your sensors and your muscle sensors all the way to the brain. There's always going to be delays. If you're too late to react, you're probably going to get killed.

Also, if you have the ability to predict and have a model, then you can do planning. You could essentially go forward in time and then you can do planning. Then that leads to abilities that allow you to have much more sophisticated behaviors than if you're just a reactive organism. I think there are a lot of reasons, at least, from a computational point of view, but probably also from an evolutionary, biological point of view to have something like an internal model and prediction that comes with it.

Paul Middlebrooks
Rao and Ballard was in 1999 and then the sensory motor theory of the cortex, I think I have the name wrong, but I'll get it right and I'll flash it up there, which you've recently presented, which you've been working on for, I think, a long time now.

Rajesh Rao
Not so long. I would say maybe like three, four years.

Paul Middlebrooks
That's not so long at all in research age. That's what I wondered. I'm not sure if you flipped in that regard. It sounds like you've somewhat flipped, but if you have like how that-- Did someone drag you out of perception or did you climb out yourself or?

Rajesh Rao
I think it was multiple things. I think it was just this realization that if you think about what is the purpose of having a brain, the context of evolution so going back to some of the early creatures that were trying to, for example, move towards places that have more nutrients because you don't have enough nutrients.

Paul Middlebrooks
Hang on, let me just stop you there. I'm so sorry to interrupt. They're moving to a place with more nutrients because they can already perceive whether there are nutrients there, right?

Rajesh Rao
You need perception as a way to minimize the error. Ultimately, you're trying to minimize the error.

Paul Middlebrooks
So perception came first?

Rajesh Rao
You need both. You can do random movements, but you still have some notion of perception because when you taste the nutrients or when you actually sample the nutrients, then that is like perception, if you want to call it perception. I think it's like the sensory and motor part are integrally tied. The motor part allows you to change your location or sample a different part of your environment. Then if you want to know was that successful or not, then you do need sensation of some sort to know, "What I did was good for me or bad for me." That's what we're calling perception in a very coarsest level.

Paul Middlebrooks
All right. I guess we have sensation without brains or sensation, perception. We could use them interchangeably, although that's dangerous. Let's say light sensors, right?

Rajesh Rao
Maybe artificial sensors.

Paul Middlebrooks
You could do artificial, but I meant biological pre-movement. What I'm wondering is evolutionarily, why did brains develop? That is in some sense a moot point because they are integrated, like you said. I think you have to perceive before you can move.

Rajesh Rao
No, I don't think it's one or the other. Essentially, you could say perceive because you're perceiving the fact that maybe you're low in nutrients inside your body, you're hungry. You could think of that as a internal perception. That may drive you to then move but then when you go out and you actually consume some nutrients, then your internal state may change because you're not perceiving it as I've consumed food, and so I'm more satisfied.

Paul Middlebrooks
But if you move without any perception, then you can't decide.

Rajesh Rao
Exactly. You do need perception, but at the same time, if it's just a creature that doesn't move at all, but it's just perceiving, that also doesn't quite make sense.

Paul Middlebrooks
Does a Venus flytrap move to you?

Rajesh Rao
In the sense that it does have an action of grabbing the insect and digesting. It doesn't move in the sense of creatures moving but, in some sense, maybe it does if you consider the movement, but again, it's not a moving organism like the one that we associate with brains.

Paul Middlebrooks
Right. It's an interesting-- Go ahead.

Rajesh Rao
These are great questions to think about in terms of--

Paul Middlebrooks
Oh, no. I'm sorry. I'm getting this off. I'm telling you, man, I had a long motorcycle ride and my brain is fried, and so I'm sorry if I'm a little squishy myself.

Rajesh Rao
Sensory motor coordination, that's what it is. [laughter]

Paul Middlebrooks
Back to this, what is it, 25-year divide between these two papers. I'm just going to focus on these two. I think you've published in total probably four papers or something like that. Slightly over?

Rajesh Rao
Yes. [laughs]

Paul Middlebrooks
In total?

Rajesh Rao
In total, yes.

Paul Middlebrooks
Sorry. All right. It felt like a big reveal when you added the motor aspect to the predictive coding. We just talked at length about how you have come around to thinking that action and motor behavior is important. Can you maybe just give a summary of how you incorporated it and what people can expect if they read this paper?

Rajesh Rao
I think the first observation that really prompted me to go on that whole journey was the fact that there were all these new results coming in showing that even primary sensory areas or sensory areas like V1, S1, or A1, were being influenced by motor activity or actions. There was papers

from the Carandini-Harris lab. There was papers from David Schneider's lab showing that you can actually get these responses that are motor-related, but they're happening in the traditional sensory areas like V1 or A1.

Not only that, but then the other interesting observation was anatomically again. If you look at anatomy, the layer five cells, there's 5A, 5E, but the layer five cells, even in the primary sensory areas like V1, A1, S1, somatosensory cortex, primary somatosensory, the layer five in all of these areas are sending axons to subcortical motor areas. The V1 sends it to superior colliculus, A1 sends it to inferior colliculus.

Paul Middlebrooks
It's all completely connected. It's a ridiculous system for an engineer.

Rajesh Rao
It's not completely connected in the sense that you don't have the superficial cells sending axons to superior colliculus.

Paul Middlebrooks
We don't know that yet.

Rajesh Rao
There's always a possibility. I think that was the other interesting anatomical constraint that we were looking at and then the funny thing there is it's not just that you're going from layer 4 to layer 2/3 and then to layer 5 and then that's it. Layer 5 is sending information back to layer 2/3. There's a loop there between the superficial deep layers. What is that loop doing. If the superficial layers are receiving sensory information from layer 4 and then the deeper layers are sending "Outputs to motor centers," then there is this loop going on between the sensory and motor within each cortical area. That was the first insight was, "What if we take some inspiration from reinforcement learning?" We all know about the Markov decision process idea and the fact that there is--

Paul Middlebrooks
We don't all know about it. If you can-- Even for non-scientists, the most basic version.

Rajesh Rao
The basic idea with the reinforcement learning is that there's a world and there's something called the state that the animal or agent is in and then when you take an action that state changes. Basically, that's called a Markov assumption because you're saying the next state only depends on my current state and current action but that is typically called the forward model or the dynamics of the world. You're saying the world is in this state or I'm in this state and then I take this action the world changes to this next state. If you repeat that it becomes a dynamics, the physics of the world.

That is the forward model or what we call the generative model of the world but then reinforcement learning the goal is to learn a policy which is given a state of the world, what is the best action I can take? That is called a policy. It's a function. For any state the animal is in, it's going to say, "I want to take this action because that action is the one that let me achieve my current task." If you now combine the model which is at the top and then the policy at the bottom, you have one that is making predictions of what the next state is going to be based on your sensory inputs and the other that is saying, "Given that I'm in this state, here's what the action is that's most suitable."

Then you can have this cycle between prediction, action, prediction, action, prediction, action. That is the general idea and that's what we suggested as a module like a sensory motor module for any cortical area and it's going to operate at its own spatial temporal scale depending on the kind of inputs it is getting. Every cortical area has its own spatial temporal scale and modality depending on the kinds of inputs it's getting and the kinds of outputs it's controlling or potentially controlling.

Then the last thing we threw in there was what about the different hierarchy and the different cortical areas interacting? Why is there a feedback connection or a reciprocal connection? It turns out that from a computational point of view, if you want to have a very rich way of modeling the world then what you want to do is have the higher level modulate the dynamics of the lower level. You need to change the function being computed at the lower level depending on the task.

If my task is to drive to the grocery store, then you basically want to load up the program for getting into your car, figuring out where you're going and then driving so you have all those "Programs" that are policies. These are these policies that are already being learned that are being plugged in to achieve your current goal but if your goal is now to go to your friend's house, it'll be similar programs, but the sequence may be different. If it's to cook something, again, it's the same idea. You load up a different program. We're suggesting that that kind of schema or loading up new programs can be done through top-down modulation. Maybe prefrontal cortex loads up in the lower areas these programs and the corresponding dynamics of what to expect.

Paul Middlebrooks
We'll just say it's in the prefrontal cortex. How does the prefrontal cortex come to be that way?

Rajesh Rao
Our hypothesis that it's starts off with all areas operating in terms of their respective spatial and temporal scales but by the time you get to the prefrontal cortex, you've already converged-- You have information that is now operating at a much longer time scale and the abstraction level has

gotten to the point where it's now operating at the level of abstract actions so there could be a population of neurons that are coding the current context or task which is saying right now, "I want to go to the grocery store."

That's a goal that got instantiated right at the highest level but it's maintained until that goal is achieved. While it's being maintained, it is modulating all the other areas that are involved in achieving that goal. That involves even the basic lower level somatosensory. Even somatosensory cortex, visual cortex, auditory cortex, they all now have that context from the higher area saying, "Here's what my current goal is. Now let's see if we can achieve that goal."

Paul Middlebrooks
Then you get your-- I don't know what you would get. My grandmother would get cookies, definitely at the grocery store. You get those cookies, and then how does it switch?

Rajesh Rao
The idea is it's a hierarchy. As you achieve the sub-goals, so basically the idea is that a complex problem, a complex task is split up into its sub-tasks, and those are in turn split up into sub-tasks all the way until you get to the point where you're operating at the level of milliseconds and controlling your muscles. That's happening in the spinal cord loops. The spinal cord is controlling things at a very fast timescale.

Then as you go up the hierarchy through the different areas, including the midbrain, and then you go all the way up to the cortex, you start to get longer and longer timescales. Then that's where you start to get these sub-tasks being achieved, and then those in turn cause a flip to the higher levels, and we go to the next goal. You're going from one goal to the next goal to the next goal.

Paul Middlebrooks
How do those goals get programmed?

Rajesh Rao
That's all about learning. You have to learn. Some people call that curriculum learning in the AI literature, but as babies after we were born, or even before we were born in the wombs, we're starting to learn these modules of what happens if I move my hand in this way, my arm in that way? Then once you're born, you're starting to grab objects, you're trying to reach for objects, you're trying to grasp different kinds of things.

You're building up this repertoire of action policies and the corresponding dynamics that go with the action policies. The idea is as you go through learning, through your toddler years and so on, you're building up a very rich set of primitives that can eventually be composed for solving more and more complex problems. There's still a lot of unknowns, but I think that's going to be.

Paul Middlebrooks
What I was asking was what-- I think we've all had the feeling like once you're done with a goal, you're like, "Well, all right, what now?" You feel lost for a moment until you realize what else you have to do. What I'm wondering is how those sequence of abstract goals, if you have an abstract goal, how does it switch? Once you accomplish one, you got to go do the next one, and then the next one. How does your prefrontal cortex know? I think you said there's a lot of unknowns.

Rajesh Rao
That's a wonderful question. I think that in the model, at least, you start off with some higher level goal, which I guess we give it, or we say, "Here's the task." Then the way it decomposes that is as you go down each higher level abstract action or policy, it generates a function at the lower level. That function basically has as the output, so remember functions are basically state to action mappings. Each of those actions is now another abstract action, which is a sub-goal.

Then that in turn generates another function at the lower level all the way down to the spinal cord level. That is what we're saying has to be learned. You have to learn each of those sequences, those functions that capture the sequence of sub-goals. We're saying it's the same module replicated at all the levels. That's one of the things that, for me, Mountcastle's whole idea of, "Hey, the cortex has something similar across," the algorithm may be similar across different cortical areas.

As a theorist, you always have that fascination with those kinds of ideas. The challenge here was, is there something like that we can come up with from a computational point of view? Then of course the brain may or may not implement it, but let's try to come up with something like that. It turns out that this active predictive coding idea seems like it is versatile enough to tackle many, many different kinds of problems, but it remains to be seen if how well does it map to something that the brain may actually be doing?

Paul Middlebrooks
Do you believe it? Do you believe Mountcastle?

Rajesh Rao
Not like exactly the cortical column is something where everything is. What I believe in this more broader idea that if you take any piece of cortex, a chunk of it, and then you give it even to trained neuroscientists, it may be hard for them to figure out which part of cortex that is. Of course, there's M1 versus V1. You can tell the difference, but I think it--

Paul Middlebrooks
90% of that is who they're connected to, not what they're doing, probably.

Rajesh Rao
Exactly, yes. I think that's a really interesting finding because if you would think that language area should be programmed very differently from vision areas and so on, if you trained traditionally, you would think visual cortex is doing edge detection. Then if you look at auditory cortex, there's no edges, so they can't be doing edge detection. You're doing something completely different. Motor cortex must be doing something very different.

It's really intriguing if you think about it in terms of what if they all have components of this sensory and motor aspects to them and it's just that in some areas, some got more emphasized than others, but they still have the basic primitives of sensory inference as well as motor control and they're operating together at multiple levels.

Paul Middlebrooks
You're a BCI, a brain-computer interface person, an engineer-minded person. This is maybe a throwaway question, but given-- What you just described is maybe it's somewhere in the prefrontal cortex. There's this higher-level, more abstract goal, and then it gets implemented at lower and lower levels until it pans out in musculoskeletal system. If you were evolution, where would you move next? Do we just get more and more abstract? What does that mean, more and more abstract? What would be a more abstract goal than honoring my great-great-great-great-grandmother's legacy by, I don't want to be stereotypical, but by doing something the way that she would do it?

Rajesh Rao
Basically you're saying what is the next evolutionary milestone?

Paul Middlebrooks
In 10,000 years, yes. How come we can't guess that?

Rajesh Rao
[laughs] Because we've been very bad at it. We can guess it, but we've been pretty bad in terms of--

Paul Middlebrooks
We're very good at guessing and being wrong, sure.

Rajesh Rao
I would say yes. I think in terms of at least human evolution, we see that we're essentially biologically, we may not have added new cortical areas or new brain structures, but we have been really making amazing strides in terms of adding tools and the cultural knowledge and tools.

Paul Middlebrooks
Perfect segue, by the way, to BCI, but I'm sorry to interrupt you.

Rajesh Rao
Exactly. Basically, I think if you think about it in terms of we basically augmented ourselves, humans augmented themselves originally using tools like rocks and made these little blades to cut meat and so on. Then we, of course, augmented ourselves in terms of wheels to move faster because our legs can only go so fast. Then, of course, we went beyond that to we were able to fly with airplanes, more recently, and then similarly, memory augmentation. We could only hold so much in our memories, but then we started writing. We had language so we started writing things down, and then we were able to pass it from generation to generation. That's another big tool in terms of cultural knowledge accumulating over time.

Paul Middlebrooks
This is the most beautiful thing about language, I think.

Rajesh Rao
Yes. It's an amazing thing which other animals unfortunately don't have. I think that leads to this idea of so if it's really tool use, and nowadays we all hang around with these devices in our hands and in our pockets. Basically we've augmented ourselves in terms of knowledge and access to information with these kinds of devices. People in the BCI world and brain-computer interfaces think maybe there is a progression there.

If you really take that idea to the extreme, then if the brain itself biologically is limited in terms of its speed and in terms of its memory capacities, then would it make sense to then augment it with artificial memories and artificial processing capacities and artificial communication capacities? That leads into the whole augmentation side of brain-machine interfaces or brain-computer interfaces, which is an area that is not something most, I think, academics want to really go too much into because there's a lot of ethical implications.

Most of us tend to stay on the medical application side. That's where the funding is, and that's where a lot of important contributions have been made and need to be made. There are companies and people in the tech world that are really interested in that aspect, the augmentation aspect.

Paul Middlebrooks
Yes, and uninterested in ethics. How do you navigate that?

Rajesh Rao
I co-direct a center for neurotechnology here. We were funded by the NSF for about 10 years as part of an engineering research center. We had a neuroethics team. We still have them. During the grant, what we did was for all the projects, we embedded a member from the neuroethics team into the engineering team itself. Every engineering team that was developing a brain-computer interface application had an ethicist who was giving them active feedback about what is the end user going to think about this? What are the long-term implications?

Paul Middlebrooks
Did your team hate the ethicist? How did that work? I don't need to ask you how you feel personally about it because it is a conundrum, I think, from a technological point of view. That's like embedding a philosopher into a research lab or something like-

Rajesh Rao
There are a lot actually, many of them.

Paul Middlebrooks
Of course, of course. That's all philosophy has left is ethics, right? No, I'm just kidding. Just kidding. Was the resistance-- I don't want to have to deal with that kind of stuff, is what I'm saying. I admire you for doing it.

Rajesh Rao
No, I think it's important because as engineers, we're really excited about building the next great thing and doing something very novel, but at the same time, the consequences may not be anticipated by engineers as much as people that are more trained to anticipate those consequences.

Paul Middlebrooks
Yes, but the consequences are always bad, right?

Rajesh Rao
Not always. I think there are some constructive things you could do. For example, there were actually some interesting observations the ethicists made when they talked to the families of patients that had, for example, deep brain stimulation. Many times a family member provide insights that may not be very obvious. They may say, "The personality of my partner changed right after they got the DBS," or the actual patient may say, "Sometimes I feel like I've lost my sense of agency. Because there's this device in there."

Then the question becomes, okay, what are the ways in which you could restore that sense? If it gets to the point where the person feels like they have lost their sense of agency, then is it okay to have a kill switch or some other way to at least provide them some relief? There are ways to have engineering solutions. If you know about some of these things that the patients are going through, I think at least that's a good way to then find ways to address issues before they become too bad to really address if they're commercialized. The technology gets commercialized maybe too late at a point in time. That's one of the hopes. You're right that many companies may not necessarily look into any of these issues because they're-

Paul Middlebrooks
Even academically and research-wise. We're talking about, you mentioned agency. We're talking about concepts-- Not that philosophy has settled on anything ever, but concepts that are very much in the air in terms of like, well, my personality has changed, but if I add a BCI or some sort of prosthetic to my brain, then is it the fault of the prosthetic or was that a deep-seated thing in me? Who's responsible? Oh, ethically, I don't want to deal with it.

Rajesh Rao
Those are great questions that-- We've been talking about them, and we've been discussing those, and I think it gets even more trickier. Just think about it in terms of, okay, now we have DBS, which is open loop stimulation, just delivering some constant current. Let's talk about AI. If you add AI into the picture that adapts-

Paul Middlebrooks
Hang on. Before you do that, this is something that you've done and are continuing to do. Just to get, and I'll introduce this in the beginning, but there's BrainNet, there's neural co-processors. I don't know. I'm not sure how you would like to describe it as an introductory idea, but you've done two people, you've done three. I don't know how many people have you gotten up to in terms of collaborating.

Rajesh Rao
I think the idea there, we were talking about brain-to-brain direct communication or, BrainNet was when we had three people collaborating directly using their brain signals to solve a problem that was, in that case, it was just a Tetris video game.

Paul Middlebrooks
This is, one person has their brain signals communicated and stimulated and the other two get to interact.

Rajesh Rao
The idea here was, the technology is not there yet in terms of actually doing brain-to-brain communication, but the idea was if we take up this challenge of, okay, let's see what we can do with the technology we have now as a way to spur people to talk about this technology so that they can put safeguards now, or at least start thinking about safeguards before it gets too far away. We said, "Why don't we take technology we have now, like EEG?"

Many people have EEG equipment in their labs. We can use the EEG to decode simple things like, if you have like a flashing light, it's called steady state, visually evoked potentials, you can decode the frequency at which you're staring. If you're staring at two different flashing lights, one of them will trigger oscillations in your visual cortex, and we can decode your intention. If you imagine making your movement, like move your hand, or imagine moving your hand, you can decode that from the motor cortex. What we said was, "Okay, we can use EEG for decoding and let's use TMS, the transcranial magnetic stimulation, to deliver information directly to the brain."

Paul Middlebrooks
Which is just for people listening, TMS is like when there's a coil outside of your brain and there's a pulse aimed at a very small part of your brain. It delivers a non-invasive, non-damaging pulse in your brain. Then depending on what function or affects, potentially affects how you think or your behavior.

Rajesh Rao
Especially if you deliver it to the back of your head, the visual cortex, then you see what is called a phosphine, which would be a little spot of light or a bar or something like that. We previously used like a Morse code idea. The flash was just conveying binary, a bit of information. We said, "Okay, why don't we use these primitive modes of interacting with the brain? EEG and TMS, and let's build a direct brain-to-brain communication system with that." This was several years ago.

We did some proof of concepts and published a few papers on that, just as a way of not necessarily saying, look, we can actually commercialize this technology because obviously we cannot. It's too bulky. If you just think about companies that are now developing implants, then you can imagine once they have implants on multiple people, then they can start doing this. If they can stimulate the brain and record from the brain, it's bidirectional BCI, bidirectional brain computer interface. Then you can start to do these kinds of experiments, and you can start to implement these kinds of systems of recording from one brain, stimulating another brain, decoding and encoding the information in another brain and so on.

That's where we get into augmentation. Now we're using neurotechnology for augmenting, in this case, the communicative and processing capacity of the human brain.

Paul Middlebrooks
Where are you now, right now with these?

Rajesh Rao
Once again, we are not in the business of really going on the commercial route. What we're doing is a lot of theory. What we're talking about-

Paul Middlebrooks
People usually ask you that and their question is really like, what future human are you creating? Is that their question?

Rajesh Rao
Yes. Some people are very much interested in that, the singularity and just like that. What's that going to be in the future? Of course, there's some people who want to merge with AI, so that's a whole other crowd who want to actually-

Paul Middlebrooks
Are you going to? Do you want to merge with AI?

Rajesh Rao
No, not really. [laughs]

Paul Middlebrooks
I don't either. I don't understand it.

Rajesh Rao
I think that what we're doing is in some sense, we're saying, okay, if you really are thinking about processing information from the brain and making sense of it and sending information back to the brain, that's what we call a co-processor for a brain. That's what we're working on is a brain co-processor, which is this idea that there's a device that is essentially a tool for the brain that can both decode information from the brain and encode information back into the brain.

Paul Middlebrooks
What that means is like, so it's listening to the brain and then it goes into an algorithm, and in terms of modern AI, it categorizes what that brain

process is signaling, almost that understands it, and then transforms it into a signal to inject into the brain? I'm sorry if these are two low level descriptions, but-

Rajesh Rao
No, I think that's one way to think about it. Of course, the way it would have to work would be that that co-processor has to have AI in it of some form. We suggested a neural co-processor, which was a neural network, so an artificial neural network. What we have there is an interaction between artificial and biological neural networks where both of them are adapting. It's a co-adaptive system. The challenge there is, how do they co-adapt to really achieve a goal?

From a medical point of view, if you, for example, are trying to replace lost function, like somebody had stroke and you want to be able to restore their movement, then this could be a device that could replace that lost function of the cortical or other subcortical circuitry. The AI has to train itself and the person now has to command this AI replacement circuit to do what the person wants it to do. Then that becomes an interesting AI problem. How do you ensure that the communication between the human brain and this artificial device achieves a common goal?

That's what we did with the neural co-processor was we said, "Hey, look, if the goal is, for example, to reach for an object when the person cannot reach the object, but then you can stimulate, let's say the spinal cord and make the person reach that particular object, then the error is now in the space of visual error. There's an error signal. Can you do back propagation, for example, right?" Which is a standard thing people do in AI. The problem is back propagation here, you'd have to do it through your body and your brain. You have to do it through the body-

Paul Middlebrooks
That's not back propagation by definition.

Rajesh Rao
Yes, but in terms of ideally, if you want to train your neural network that's delivering the stimulation, then you have to get that error signal somehow. The error signal, unfortunately, is now in the space of an external task. What we suggested was, okay, for all of these AIs, they need to have an internal model, this forward model of, in this case, your brain dynamics. How your brain reacts to stimulation. It's like a forward model of-- the action here is stimulation, and then it's going to cause your brain dynamics to move in a particular way. If you're able to predict that dynamics, then you can come up with the best stimulation to give for achieving a particular goal. It becomes, again, a control problem.

Paul Middlebrooks
These are things I just-- I'm a neuroscientist, but I ponder these things. One of the most amazing things about the brain is how adaptive it is. Sometimes I think, well, maybe that is the function as if there was a function, but the adaptability is almost the answer. What I was going to ask is, how wrong can it be for the brain to still be able to deal with it? Is that a testable question? How bad can you make the signal relative to how quickly the brain can learn it? That's a terrible, poorly phrased question, I'm sorry.

Rajesh Rao
It's actually a very important question especially in the sensory stimulation area. People that are trying to restore the sense of touch, for example, from an artificial prosthetic device, stimulating the somatosensory cortex. There's a big question there in terms of, should we make the stimulation be as natural as possible? We make the sense of touch. When you stimulate the somatosensory cortex, patients typically report, "Okay, I feel like it's pins and needles." It's not quite the nice sense of touch I get when I use my own hand.

Paul Middlebrooks
Then you say, well, just wait a week or two and it won't be, or something like that.

Rajesh Rao
People would say, "Hey, you're just not used to it. It's a novel signal right now, but if you keep using it, maybe you get used to it and that becomes a natural, sense of touch."

Paul Middlebrooks
This is like training. This is like trying to convince my children that, "Ooh, don't worry, keep practicing. It'll get easier."

Rajesh Rao
In some ways, yes.

Paul Middlebrooks
It sounds pretty painful.

Rajesh Rao
Yes. The painful is a key part there because if you are not really delivering "naturalistic stimuli," then it may take much, much longer for you to learn. I think there's a trade-off there in terms of if you're able to speak the language of the brain in some way, basically, you're delivering the pulses, electrical stimulation, optical stimulation, the way that the brain can parse it more easily, then you may be able to learn much more quickly than if it's an artificial set of pulses, like stimuli. I think that's an interesting problem.

Paul Middlebrooks
In terms of natural, then, is that just noise or is it like the natural statistics of the world?

Rajesh Rao
It's more the way that the neurons in that particular area are receiving information from other neurons. For example, if you know that in the population level, there's a particular set of frequencies that are operating at that scale, like beta frequency oscillations or something, then you can think of delivering the stimulation at some particular phase of that oscillation or you're trying to synchronize with that oscillation. Maybe that is better than you just delivering it irrespective of what's happening in that brain area.

The bottom line is you need to be recording as well as stimulating. You can't just be stimulating the brain area. You need to be recording to know what's happening in that brain area and then change the stimulation according to what's happening at that point in time to consistently deliver intelligible stimulation.

Paul Middlebrooks
Within a window, because that's antithetical to the idea that the brain can handle a wide range. A brain area doesn't get to decide the signal incoming, but what you just said is that you need to do it in a way that it's expecting it. It's a very predictive coding outlook, by the way. Otherwise, it might not work.

Rajesh Rao
Yes, it's a way to enable the brain to learn faster. Like I said, it's possible that you could not pay attention, just keep delivering something. Eventually the brain may figure out how to interpret that stimulation pattern. The question is, how long will that take? It may happen quickly, or it may take very, very long or maybe it'll take forever. In general, I think it makes sense to have some feedback about what's happening in that brain area and as you're delivering the stimulation.

Paul Middlebrooks
Can we partially get to this by doing learning studies and maybe just probing the historical literature and learning studies? Modern learning studies are becoming more better and better. I don't know if there is an answer to the proper way to learn X yet, like the most optimal. I know that there's a lot of people who make a lot of money who say that there is, but I intuitively think you got to go do it. It's got to feel really bad at first and that's how you're going to eventually learn. It also depends on the goal.

Do you think that we can-- Is this an experimentally behaviorally tractable problem that could help answer these issues? Because you're speaking about the activity in the brain, but then we're eventually talking about behavior.

Rajesh Rao
Yes, it's an interesting question in terms of, how do you study what's happening as the brain is interacting now with this external device that is sending direct input into the brain?

Paul Middlebrooks
It's just the optimal input to, if you're going to use these external devices.

Rajesh Rao
The neural co-processor idea was one way to get at it in terms of saying, okay, if you are able to adapt the neural network that's delivering the stimulation to minimize the error in the external task space, then you're relieving the brain of having to do a lot of that learning. Of course, the brain is already trying to learn, so it will be changing. It's a non-stationary system, which is why we call it a co-adaptive system. That's where it becomes tricky.

If it's only one side is adaptive, let's say the brain is adaptive, but then you have a fixed mapping for your stimulation pattern, then all the burden is on the brain to learn that. Then if you do it where both the systems are able to adapt, then the challenge is, can you make it-- they're both optimizing the same cost function or same optimization function. That's where the interesting engineering challenge is, to make it a co-adaptive, co-evolving system. That's where I think the future is probably going to go if you are going to use AI for these brain-computer interfaces.

Paul Middlebrooks
You want to do a quick segment of epiphenomenon or causal?

Rajesh Rao
What was that again?

Paul Middlebrooks
You want to do a quick segment of, is this an epiphenomenon or is it causal?

Rajesh Rao
Oh, okay, a quiz?

Paul Middlebrooks
Ready? Yes. Yes. Opinion. Ready? Oscillations.

Rajesh Rao
I would say it's causal. Yes.

Paul Middlebrooks
Consciousness.

Rajesh Rao
Epiphenomenon.

Paul Middlebrooks
Oh my God. We're done. I just heard someone say that.

Rajesh Rao
[crosstalk] I'm taking the most liberal definition there. [laughter]

Paul Middlebrooks
Okay. Those are the only two. I heard someone I respect say consciousness was epiphenomenal and then you were talking about oscillations.

Rajesh Rao
I just want to be a little bit of that cutting edge there. Yes, let's go with that.

Paul Middlebrooks
Okay. All right. Great. That's staying in, by the way. I'm not going to edit that out unless you force me to.

Rajesh Rao
Sure.

Paul Middlebrooks
Okay. I'm looking at the time right now and first of all, what have we missed about the neural--? I'm going to point everyone to all of your work, but I've been punchy and jumping around. What have we missed in terms of the neural coprocessing and/or BrainNet and the-- We haven't used the term telepathy, right? Which is the coup de gras, the pinnacle, what people-- the spooky term that these things imply. Is that a goal? I'm going to stop interjecting. What have we missed in terms of what we haven't spoken about with the neural co-processors?

Rajesh Rao
The neural co-processor idea, there was relationships to the brain-to-brain communication work that we had done early on. That work of course, got us, both people that were quite interested and quite intrigued by it. Also I think a lot of scientists were like, "Oh yes, that's just the old--" Using TMS is like frog's legs being stimulated. We got people saying, "Why did you even do that?"

Paul Middlebrooks
Oh, really?

Rajesh Rao
Yes. The main point is, it's not that we were claiming that, "Hey, this is like a grand new discovery or invention." The idea was to say, "Okay, we have been talking about things like mind-to-mind communication or telepathy for a long time in science fiction. Then the BCI really progressing at the rate that it's progressing with companies getting interested in it, this may be a good time to have that conversation. I think one of the interesting things there is, we may actually get something like very primitive forms of brain-to-brain communication if these companies have multiple patients with implants.

Because it doesn't take much to send really simple kinds of information from one person's brain to another as soon as you have the ability to stimulate. Which is something I haven't seen yet. Once that starts happening, then I think it's only a small step towards starting some really primitive kinds of what you may call like-- I'm sure the companies will claim as telepathy.

Paul Middlebrooks
They'll be called tele-- What will they be called? Because there was the telephone beforehand. It'd be a good company name for that.

Rajesh Rao
What about this one? It will be like teleneuronics or something like that. [laughs]

Paul Middlebrooks
Yes. Oh, why can't we come up with something, Raj? This is-

Raj: I'll think about it. It'll have to be trademarked first before we can talk about it. No, I'm just kidding.

Paul Middlebrooks
I know you have it. You have it in your back pocket. You're not letting on.

Rajesh Rao
I think the unfortunate consequence of some of these kinds of claims by companies, and even I think some of the experiments that we've done, is that there's a lot of people who are probably schizophrenics, who think that they're implanted with these kinds of devices already. A lot of us get emails every day from people saying, "Hey, can you help us? We think the government has implanted something in me, or there's a chip in my brain."

Paul Middlebrooks
Or the Lord did.

Rajesh Rao
Yes, people actually even come to campus. There's people who have driven miles just to come to campus, to our center, to say, "Hey, the doctor says there's nothing there. Here's an X-ray. There's nothing in the X-ray. I firmly believe it's this microscopic thing that's in my brain that's controlling my actions." Unfortunately, I think that's the price you have to pay, you start to push that frontier. There's going to be people who also think that that's happening to them, even though we can assure them that that's not really possible at this point in time. We don't have the technology to do something like that. That is something we are faced with as well. It's part of the whole ethical issue and communication issues with this field.

Paul Middlebrooks
You said that we can't telepathically communicate yet. We can't stimulate each other's brains, but we do all the time. You and I are doing it right now through language. You have an interest in-- I don't know where your interest in this script came from. Now you've at some point, you learned a ton about language and just how language works and some of the statistical properties that are indicative of what a language is. Is this a hobby of yours? It's beyond a hobby now, but how did you get involved in the Indus script and determining, trying to decipher this thing?

Rajesh Rao
Yes, I've always been interested in ancient history. I was doing one of my sabbaticals, like I think my first sabbatical, when I said, "Okay, now what if I indulge in that fascination with ancient history?" Especially ancient Indian history, because I'm from India. We have the civilization there that's more than 4,000 years old that we know very little about, partly because the script of that civilization is not yet deciphered. This was a civilization that was much bigger than the Egyptian and the Mesopotamian civilization.

Paul Middlebrooks
I'm sorry to interrupt, but one of my favorite classes that I took when I was an undergraduate was the origins of writing. Denise Schmandt-Besserat taught the course, and we learned all about cuneiform and how early writing might have been brought about by economic exchanges with little clay tablets where people put little marks to show how many goats or cattle that need to be owed or exchanged and stuff. Where does this-- and I can't remember how. What you're talking about is for, there's like 2000 BC.

Rajesh Rao
Yes, and before that, I think starting from 2000, probably 2800 BC is when they believe the civilization started evolving to the point where it was mature enough, and then it ended, or the demise of the civilization is thought to be more 1800 BC, 2000 to 1800 BC. The curious thing about them is that it was not like a huge amount of time, and there were some precursors, 3000 BC and so on, but the prime of the civilization, when they were really and their prime was probably 2800 to 1800 or so, 1500 BC.

I think the interesting thing about the civilization is the fact that they were actually trading actively with the Mesopotamian civilization, the Sumerians and Akkad and those civilizations in the Middle East, and they've found these seals. You mentioned, they had these stamp seals, so you can stamp information about typically trade, the goods that they may have been stamped on. I think the intriguing thing there is, of course, we know cuneiform and the history of writing did evolve from the Middle East. Then the Egyptians, of course, followed suit with their hieroglyphics. It's believed that the Indus civilization also evolved their own set of symbols, so about, 500 to 700 symbols that are very pictographic in many cases.

Unfortunately, there are no long inscriptions. Most of the inscriptions they have found have been on these stamp seals, which are only about five symbols long on average.

Paul Middlebrooks
Some of them squished in different?

Rajesh Rao
Yes. There's usually an animal imprint, which may have been like a symbol of a clan, like a unicorn is the most common animal symbol.

Paul Middlebrooks
Really?

Rajesh Rao
Yes, a unicorn, yes. Which people in the ancient times thought it originated in India. Unicorns.

Paul Middlebrooks
I didn't know that.

Rajesh Rao
Perhaps it was because of the Indus civilization and their seals. The seals always had at the very top a set of symbols, a sequence of symbols, which were typically on average length five symbols long. There have been no long inscriptions on tablets. Unfortunately, we cannot apply many of our AI techniques, because there's only about 6,000 of these.

Paul Middlebrooks
How do you tokenize that?

Rajesh Rao
What we did do, we did do some kind of analysis of it, which is we looked at the entropies, the block entropy of these symbol sequences in terms of which symbol follows which other symbol sequence. Then we showed that the block entropies tend to match the block entropies of linguistic scripts, as opposed to DNA, for example, or proteins, which are much more, "random," they're more flexible. The entropies are much higher. Music is slightly above language. Music is more flexible, but not as much as DNA or proteins.

Paul Middlebrooks
What's less flexible?

Rajesh Rao
Least flexible is computer programming languages like Fortran. It's below language.

Paul Middlebrooks
Oh, you're going to say Fortran.

Rajesh Rao
Yes, that was the original one that they did. We followed that. I'm sure it may start to fall, especially with now more different kinds of programming languages.

Paul Middlebrooks
Naturalistic, yes.

Rajesh Rao
Naturalistic, you're probably getting closer to the entropy of natural languages. It's an interesting way to think about it. What we did was said, "Okay, here's an undeciphered script. Let's look at all the symbol sequences and see if they fall in any of those ranges." It turned out that they fall right in the middle of the range of linguistic scripts, which means that perhaps there's at least one piece of evidence that they may have used it to encode a language, but we still don't know what language that was. There's a huge political debate in India about whether it was the Indo-Aryan language Sanskrit, as the basic civilization, or is it the South Indians have a Dravidian language family? Some people think it's that language that's encoded in there.

Paul Middlebrooks
Political factions are trying to claim it or something?

Rajesh Rao
In some sense, yes. In some sense, there's that. There's also this notion of Aryan invaders coming to India. People don't like that. Any migration to India, especially if it was these Indo-European invaders coming to India, that was the original colonial interpretation of it. The genetic evidence seems to indicate there was actually migrations happening from Central Asia into India. That's possibly how these languages like Sanskrit and many of the North Indian languages will have similarities in grammar with the Indo-European language family, compared to the South, which is more Dravidian language family, which is not related to the North Indian language families.

There's some amount of friction there. Even now you'll see people claiming decipherment of the Vedas in the seals, but it stretches more-

Paul Middlebrooks
You're an expert now, so you can judge those things, right?

Rajesh Rao
To some extent, yes. I've definitely delved into that for a long time now as I can look into it. Unfortunately, most of the time you can see that they're just stretching it and it's just not a plausible decipherment. I think you're just going to have to wait for more excavations and maybe a Indus Rosetta stone that'll help us decipher that script.

Paul Middlebrooks
I doubt we'll find it. Do you think we'll find it?

Rajesh Rao
No, it's a difficult task. It's possible it's already been found and it's just lying in some warehouse in Iraq somewhere in a museum because they were trading with these folks.

Maybe there were some inscriptions that had both cuneiform and the Indus script, but so far nobody has found anything like that. They have found Indus scripts with really odd sequences. It seems like they were using the Indus script to express the language in Iraq, like a foreign language, maybe in-

Paul Middlebrooks
Oh.

Rajesh Rao
That's really interesting.

Paul Middlebrooks
They're only with like five characters?

Rajesh Rao
Just five characters, but the symbol sequences are very different from what you would find in the Indus Valley, in the finds there. They were using it in a different way, at least. I think there are some really intriguing suggestions like that.

Paul Middlebrooks
Is this going to be a lifelong project for you?

Rajesh Rao
For me, it went from being a hobby to more of really getting drawn into the literature and the history of it. I do spend a lot of time reading about it, but I have no funding for it. It's more of a sabbatical project, where it keeps me interested in something very different from what I typically do. I look at myself as somebody who's interested in the past, which is this Indus script. Then the present is like, okay, let's figure the brain out. Then the future is like the BCI augmentations. I like to keep looking at all of those aspects. It keeps my brain at least occupied and interested.

Paul Middlebrooks
One obvious question to me then is, not many people who engineer large language models are linguists or have a sense of what language is or how it may have come about or it's early post origins, et cetera, or even its origins. Has this affected how you think about large language models? I'll preface this also by saying that whatever great AI model is working, people are going to say, "Well, maybe the brain is like that." Then map, whatever AI model is currently working to brain processes and often it works and stuff. Has this affected how you think about large language models or how do you situate how you think about large language models maybe is a better question.

Rajesh Rao
Yes, I was quite intrigued when the paper came out, the attention is all you need, paper. Then also the fact that prediction was a very key aspect of it. It's not-

Paul Middlebrooks
You must've been elated.

Rajesh Rao
Yes. It was basically predictive coding, except for the fact that they were not using the prediction errors to do inference. There was no update of the internal representation, but the learning was obviously driven by prediction errors. The fact that there was a hierarchy there as well, so that there's these soft attention layers, which is building up these dynamic representations over time. You're doing-

Paul Middlebrooks
[crosstalk] sending out the representation to every layer, like every time. It's just massive, right?

Rajesh Rao
Exactly.

Paul Middlebrooks
It's in some sense beautiful probably to you.

Rajesh Rao
It was actually very interesting to know that there's now this artificial system that's just purely based on a prediction. By predicting the next word, it was able to do so much.

Paul Middlebrooks
Not even recurrent, like there's no feedback.

Rajesh Rao
It's not a regressive model, yes, you're right. Effectively it was taking into account everything in the past, while processing the future and the future and the past were all getting processed simultaneously. I think the cool thing there was the fact that, yes, I was able to really learn these very impressive, predictive, models and then generate these sequences, which seemed to be quite in line with the thing we were thinking about in our active predictive coding model, this idea that there's a sensory processing aspect that's constantly predicting the next state based on the previous state and action, but there was no action here. The action was missing.

Paul Middlebrooks
Actually, it's just output.

Rajesh Rao
It's just like a token, basically, that the previous token may be like being put in, or all the tokens in the past are being used to process the current token. I think that that's also something that's missing is, my sense is the controller is missing. In the active predictive coding architecture, there's a policy network that is controlling the generation of the next input. We have the predictive network and the controller network, and they're closely interacting with each other.

Depending on the task, the controller network is feeding in the action to the predictive network saying, "Okay, generate this next. Because I'm going to take this action." I think we've lost that in the transformer model. I think that's where some of these RL approaches, they're applying now to the transformer models, it's getting the controller back in the picture. That's where I see a relationship. More recently people have shown that transformer networks implement a special kind of hyper networks, which is what we are using for active predictive coding.

When I said that the higher areas are modulating the lower areas in active predictive coding, the higher cortical areas may be modulating the function being computed at the lower areas. That is basically what is called a hyper network, in the AI literature, where the input of the higher level is transformed by a neural network called a hyper network, to another neural network, which is the function. It's basically a whole new network is being generated.

It appears that transformers are doing something like that in a simple way. A linear hyper network model. If that's true, then I would say there's probably a closer link between what we've suggested and what transformers are doing, except we were trying to model the cortical architecture in terms of hierarchy and laminar structure. transformers don't have that. They don't have the controller aspect either.

Paul Middlebrooks
Then overall are transformers-- I'm trying to get a sense of your judgment on them.

Rajesh Rao
I would say they're great, but they're not enough. They're great as a principle of how predictive models can really capture a lot of information about essentially the statistics of the world, the physics of the world or the dynamics of the world. They can be used to capture that, but that's not enough for intelligence. I would say, I go back to that notion that people have suggested, Paul Cisek and Buzsaki and others, that ultimately, it's all about actions and movements. If you don't have the controller there telling you what to do, then you, like transformers don't have, like chat GPT and transformer, they don't have a sense of agency. That comes-

Paul Middlebrooks
How do you get that?

Rajesh Rao
What's that?

Paul Middlebrooks
How do you get that? How do you get a sense of agency?

Rajesh Rao
I think you get it from being able to act in the world and being able to actively use action as a way to generate language. We get that because we actually act upon the world, and we use speech. That's actual motor actions, right?

Paul Middlebrooks
But squirrels have agency, they don't have speech.

Rajesh Rao
Yes. You have to have action. You don't have to have speech, but in order to have agency, you need to be able to act upon the world and be able to cause consequences. You have to be able to see the consequences of your action and then maybe achieve goals. You start to see, okay, if I take these actions, I achieve my goals. Then if I do these actions, I get something else. That's how you build a sense of agency. Then for a purely predictive system, which is only predicting, but not really using actions as a way to guide those predictions and get to a particular goal. Then it becomes hard to argue it has agency. Some people do, but I think it's a missing piece.

Paul Middlebrooks
You think that we could build agency?

Rajesh Rao
Yes, I think any system that starts to interact with the world, its own actions and generate goals, those are all things that you associate with an agent. You do require to have that policy or controller and potentially a model of the world to go along with it.

Paul Middlebrooks
I have two more questions for you, and I promise I'll let you go.

Rajesh Rao
Oh, it's been fun.

Paul Middlebrooks
Way earlier-- I'm sorry?

Rajesh Rao
Oh, it's been fun, so I could probably go on. Yes, go ahead. [crosstalk] You're tired, you just came back.

Paul Middlebrooks
I'm doing a brain initiative workshop tomorrow where I'm participating. I have to like, yes. One, way earlier on, when we were talking about active predictive coding or predictive processing, in general, you were talking about, back then, back in 1999, when people-- I said we, we're recording single neurons and that's when I grew up, I recorded single neurons. The technology wasn't quite there to be able to begin testing the hypotheses that might be issued from a theoretical framework like predictive coding.

Since then, we're starting to get connectomes, which everyone always said like, well, man, once we have a map of all the connections, we can do everything. That is starting to come to pass. We now have super high-density recording techniques, neuronal recording techniques, so that we can put high density electrodes in different parts of the brain and get lots and lots of neurons. Between those two, okay, and then now, and we have-- Sorry, there's like five things. Those two, and then we have AI models, we have high compute, we have a lot more statistical models that we can employ, that's the same as AI. In terms of modeling, which of those, if any, do you think is the most important that is happening right now?

Rajesh Rao
That's a trick question, because I think if I choose one, it's probably going to alienate a whole bunch of people working on the others. I would say we need all-

Paul Middlebrooks
No, okay, I'll ask it differently. Which of those do you, like when you wake up, you're like, "Oh, man, I could use this to do that." Current, I know they're all necessary and awesome.

Rajesh Rao
Yes, I think, it changes. For me, the amazing part, it's a fantastic time to be a theoretician in neuroscience right now, because there's so much data. This is a great time to really think about these larger scale theories. When I was a grad student, there were books and papers by Mumford, there was a book called *Large-Scale Neuronal Theories of the Brain* by Christof Koch and Joel Davis. These were all like, at that time, there were these theories, but then the data just was not there to test these theories. Now I think we're at a time where we can actually start to test some of these theories. I would say from one day to another, I feel like some days when I'm thinking about a particular theory, I may go into the literature on large-scale neural recordings, really try to find, okay, is this supported or not, and what are the things out there? Then the next day, I may actually be looking at connectomics to see, okay, is the connectivity in anatomy still, is that there to support some of these other theories?

Of course, AI is a wellspring of new ideas always. There's so much going on in AI that it's hard to keep up. Then, of course, there's some gems there that you can pick out and then see if that's relevant to understanding the brain. It just feels like we're at a time where it's this explosion of information, and really to make advances, it's almost like you have to train your brain to essentially forage and find the right kinds of information to really build up, new theories.

It's a tough task, but I think it's actually better than not having any data. When I was a grad student, we didn't really have that kind of data. We had, like you said, single-neuron, responses and so on. It was very difficult to say, okay, what's the system doing at the systems level? Now, I think we're getting to the point where we can actively collaborate. Your generation, the generation that's actually training right now, they're much more savvy with computational models. They're much more receptive, computational models.

That's really great. That means that these ideas, like predictive coding, hopefully, can be tested more quickly, and they aren't going to lie dormant for 10 years before people start taking an interest in them. I feel like this is a great time for people. I encourage people in many different areas, they can be computer scientists. They can be people doing AI. They can be people in neuroscience, psychology. I think they should all feel comfortable with coming up with theories and computational models. I think we have the training now in many areas, but this is a great time to start thinking big. Think across just one area and one region.

Paul Middlebrooks
There's maybe a paradox here. If you have tons and tons of tools, it may be harder to think big. It may be harder to think theoretically. I'm wondering how you might think to foster someone's theoretical bent, or how to think theoretically, given this deluge of tools.

Rajesh Rao
Yes, so that's a great question, again. I think it's something that you-- for me at least, it comes from thinking about it first from a computational point of view, saying, okay, here's a particular problem that let's see if we think of solving this problem, how would you do that from a normative point of view or from an AI point of view? You can then say, okay, let's look at some of the data from the brain in terms of where these kinds of problems have been explored. Then that's one way to do it, is to start from the computational and then go more deeper and deeper into the neuroscience.

The other way is to go bottom up. You could say, "Okay, I'm working on these areas, but then I want to actually go beyond just this one area from a theoretical point of view, even though I'm just recording from this area." Let's think of it in terms of this area interacting with all the other areas that are connected to it. Then the bigger picture of behavior. How is that going to work? Maybe look at people working in those other areas as well? It is a hard problem, but I think it's worth looking at it beyond just-- I remember like when I was a postdoc, I went to the lab of a very famous neuroscientist and then I was asked-

Paul Middlebrooks
Is it Sejnowski? Is it-- Sorry to-

Rajesh Rao
Oh no, I was doing a postdoc in Sejnowski's lab in Salk, but I was visiting another lab somewhere else.

Paul Middlebrooks
That's a different famous nerd.

Rajesh Rao
Yes. I was doing a postdoc in Terry's lab, but then I went to a different visiting a lab. Many different labs were visiting at that time as a theoretician. Then I was asking them, "Hey," the person who was working in V1, and then so I said, "Okay, what about V2? V2 and V1 are connected, shouldn't you be considering the interactions between the two?" The person said, "No, no, I'm going to figure out V1 first, after I understand V1, and then I'm going to understand V2."

What if you cannot do that? What if you cannot reduce, it's not a reducibility in that sense? What if V1's properties are intimately connected with V2 and other structures? Then it becomes really hard to understand just one area. I think that's the challenge we have as neuroscientists is, the brain is a very complex piece of machinery that's been evolving over millennia. If you use reductionism, then it becomes very hard to understand what's happening across the brain. At the same time, maybe we don't have the capacity to understand the whole brain, like what's happening everywhere. Then we have to find ways of picking the right abstraction level to understand and hopefully connect those abstraction levels from the behavior level all the way down to the molecules. One person may not be able to do all of that, but hopefully as a community, we can start to understand the brain at multiple levels of understanding.

Paul Middlebrooks
It's hard. It's hard. Even people who are super good at it, you will admit it. It's a little difficult.

Rajesh Rao
Definitely. Yes. Especially with all the information that's now coming out and trying to, but I don't think we can lose hope. I'm optimistic that it's going to happen. I don't know if it's in my lifetime, but I think we're on our way. It may be sooner than we think. I think we should be optimistic about that.

Paul Middlebrooks
All right. That's a great note to end it on. I meant to mention this up at the top. What I'm going to do when we're done here, I'm going to email you an invitation to join me again, because I know it takes about a year to get you on. It took a long time to get you on, but man, I really appreciate it, Raj, you coming on.

Rajesh Rao
Yes, it was fun. Thanks again for having me.

[music]

Paul Middlebrooks
"Brain Inspired" is powered by *The Transmitter*, an online publication that aims to deliver useful information, insights, and tools to build bridges across neuroscience and advanced research. Visit the transmitter.org to explore the latest neuroscience news and perspectives written by journalists and scientists. If you value "Brain Inspired," support it through Patreon to access full length episodes, join our discord community, and even influence who I invite to the podcast. Go to braininspired.co to learn more. The music you're hearing is "Little Wing" performed by Kyle Donovan. Thank you for your support. See you next time.

[music]
