THE TRANSMITTER | BRAIN INSPIRED

Eli Sennesh talks about bridging predictive coding and NeuroAI

Predictive coding is an enticing theory of brain function. Building on decades of models and experimental work, Eli Sennesh proposes a biologically plausible way our brain might implement it.

3 JANUARY 2025 | by PAUL MIDDLEBROOKS

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

Eli Sennesh
Why do things feel like stuff? Why do we engage in the behaviors we behave in? Not why in the normally scientific reductionist sense, what are the mechanisms once we hold the behavior fixed? Instead, if we don't hold the behavior fixed, what are you or any other organism going to choose, and why that choice instead of something else?

There's this problem where in neuro, we are often doing paradigms or tasks that, from a pure AI point of view, might be considered almost trivial, but from a biological plausibility point of view, that often makes them hard again.

I had actually been prepared for the concept that you might walk arrogantly into experimentation with some grand theory, and think, "This is going to totally be right," and do your first experiment and it's totally wrong. In fact, that happened.

[music]

Paul Middlebrooks
This is “Brain Inspired,” powered by *The Transmitter*. Good day to you. I am Paul. This is “Brain Inspired” podcast. As you just heard, Eli Sennesh is a Postdoc at Vanderbilt University. One of my old stomping grounds. Eli is currently in the lab of André Bastos. Andre's lab focuses on understanding brain dynamics within cortical circuits, particularly how communication between brain areas is coordinated in things like perception, cognition, and behavior. Eli is busy doing work along those lines these days, as you'll hear more about in a moment.

The original impetus for having him on this podcast is his recently published proposal for how predictive coding might be implemented in brains. In that sense, this episode builds on the last episode with Rajesh Rao, where we discussed Raj’s active predictive coding account of predictive coding. I've said predictive coding multiple times now. As a super brief refresher, predictive coding is the proposal that the brain is constantly predicting what's about to happen. Then, stuff happens and the brain uses the mismatch between its predictions and the actual stuff that's happening, to then learn how to make better predictions moving forward.

I refer you to the previous episode for more gruesome details about that process. Eli's account of predictive coding and how it might be implemented in brains, along with his co-authors, of course, they call it "divide and conquer predictive coding." You'll hear why in our discussion, the divide and conquer approach among other things uses a probabilistic approach to account for how predictive coding might be implemented in brains. We also talk quite a bit about the difference between practicing theoretical and experimental neuroscience. Eli's experience moving into the experimental side from the theoretical side. Which, well, you'll hear, it turns out everything has its own challenges, let's say.

All right. Show notes for this episode, our braininspired.co/podcast/202. As always, thank you for being here. Thank you for listening. Thank you to *The Transmitter* for your support of this podcast. Thank you to the patrons who also reach out and support. Here's Eli.

[transition]

Paul Middlebrooks
Eli, are you ready?

Eli Sennesh
Yes.

Paul Middlebrooks
We were just chatting about how you are, just a few floors up from where I did my postdoc. This is your first postdoc, right?

Eli Sennesh
Yes.

Paul Middlebrooks
In Nashville, Tennessee. I'm curious... this is in some sense a follow-up episode, because I just had Rajesh Rao on to talk about his active, predictive coding work from which updates the original predictive coding framework from 1999, that focused all on sensory. What he did here was, basically, bring in an action, part of the story into the predictive coding story.

Eli Sennesh
It's very lucky timing. We actually just read his APC paper in General Club two years ago.

Paul Middlebrooks
Oh, really? Oh, we did it a few weeks ago. It was helpful. We'll get to your related work, compare, contrast, et cetera. First, I know you have a computer science background, and I'm trying to understand your worldview. People ask me my worldview, and I can't describe it, because it assumes I have a worldview, like how you approach the world. I know you have that background in computer science. There's, I don't mean dry in a bad way, but a very computational algorithm-centric approach. I thought, "Well, maybe that's where he's coming from."

I know you did some work with Lisa Feldman Barrett. It's all about feelings and how that drives, so much of how we interact with the world. I'm just curious what your worldview in the neurosciences. Oh man, he's getting comfortable.

Eli Sennesh
Yes. Okay. There is actually-- I don't know if I have a worldview, but I have a direction and a vibe.

Paul Middlebrooks
Oh, I like that. That's a good way to phrase it.

Eli Sennesh
I feel things out for what I think could be a workable scientific approach to try and address the questions I'm interested in. Overall, I feel like the question I'm interested in is-- This is going to sound even sillier than saying consciousness. I'm sorry to say, but why do things feel like stuff? Why do we engage in the behaviors we behave in? Not why in the normally scientific, reductionist sense, what are the mechanisms once we hold the behavior fixed? Instead if we don't hold the behavior fixed, what are you or any other organism going to choose, and why that choice instead of something else?

Paul Middlebrooks
What does that mean? Hold the behavior fixed.

Eli Sennesh
Oh, okay. Now I just get to channel Lisa straight up,. So often in neuroscience experiments, and I'm thinking particularly actually, some of the animal experiments we do here at Vandy. We basically, head fix your animal, first you chair the animal, head fix, then you train them to fixate on a dot on screen. Monkeys are much smarter than mice, so they can do this in exchange for juice.

Paul Middlebrooks
You just described my entire academic career there.

Eli Sennesh
Yes. An increasing portion of mine. then you basically have them move their eyes as the only motor output of whatever you're having them do.

Paul Middlebrooks
It's the most highly constrained lab experimental setup as you can so that when you're asking the question, does, for example, frontal eye field do encode some decision process related to the behavior. You don't have to worry about all of the other conflating factors that are involved in the other behaviors.

Eli Sennesh
I'm a strong believer that experiment is theory-laden. This means that if you're doing one of these highly constrained experiments and you have a theory about what frontal eye field is doing. Great, you've controlled everything else so that you can test your theory about frontal eye field. Now back to channeling Lisa, if what you're trying to investigate is not something that you've left unconstrained in your setup, then you can't actually test theories about it.

I would say, for instance, can you use a head-fixed monkey in a chair to strongly test these physiological theories about allostasis, interoception, these other nice keywords that I wrote about with Lisa and Karen. Questionable. That is really how I ended up-- Gosh, there's a whole story actually of how I ended up working with Lisa and Karen.

Paul Middlebrooks
Yes. Let's hear it if you're willing to divulge.

Eli Sennesh
Yes. The short version is the stars aligned in a way that they never have before or since. Obviously, doesn't every academic have that story. Really it's that I had this computer science background and then late in my master's actually I started getting interested in like cognitive science, very like Brendan Lake, Josh Tenenbaum type of stuff. I had to spend a couple years studying things, to go back and try and give myself the background to engage with any of this, try to change direction. I discovered that what I was really interested in was the feeling and the why.

I started trying to figure out, well, "Who has an approach that is like this to these subjects? Where the kind of like this is very probabilistic." They were using probabilistic programs to model concept learning, and this was all working very nicely for them. They had that science paper in 2015. I was so impressed. I started reading neuroscience.

Paul Middlebrooks
You decided to continue?

Eli Sennesh
Oh, no. This gets so much worse. I'm embarrassing myself here. You could probably guess now, if someone's looking for probabilistic approaches in neuroscience, mid to late 2010s, who are they running into?

Paul Middlebrooks
Oh, well, not many people, but you've already named some of them, but I was about to ask you, why probabilistic? Maybe we could start there.

Eli Sennesh
At the time, I was an amateur and I was just vibing and trying to go from one thing that I felt like I could understand to another thing I felt like I could understand. At least partly at the beginning, the interesting part about the Tenenbaum and Lake work to me was, hey, unlike that old field of AI that I took a course in undergrad and hated it.

Paul Middlebrooks
This is symbolic style, old field, or connectionism, old field?

Eli Sennesh
Oh, I went to undergrad at UMass Amherst, so the AI class was symbolic search, heuristics, all of that, logic. The machine learning classes that I didn't take at the time were random forests, SVMs. I think there was some neural networks, but they've hired a lot more people doing neural networks since then. Oh, RL was absolutely huge at UMass Amherst to the point that they hosted the RL conference in Amherst this past year. I eventually realized, "Oh, those big guys, Sutton and Bardot, wait, Bardot, Andy Bardot, who would just walk through the hall, yes, that Andy Bardot.

Paul Middlebrooks
Isn't that a super interesting thing about academia when you meet, I don't know, maybe hero is the wrong word, but these godheads of classic things, and then, oh, they're just regular folks.

Eli Sennesh
That's why I say this is so embarrassing, is I actually went to school in a department full of such great people.

Paul Middlebrooks
When you don't know, you don't know.

Eli Sennesh
I was honestly dismissive about it because I was like, "This is all just heuristic search." It's heuristic search, you throw a lot of processing power at it, and maybe sometimes it works, but this is not actually how a brain or a mind would work. This isn't the real thing. When I started reading those Tenenbaum and Lake things, they were saying, "Well, we fit to behavior, we've done an actual experiment and checked. We're not just defining some toy task that we can then solve computationally with reasonable ease, and then go back and forth between approximations and heuristics, for the rest of our careers until an AI winter hits and wipes us out."

Gosh, maybe I did take something from the UMass Amherst, actually. Maybe I took some residual post-trauma from the AI winter. They were fitting behavior and actually fitting a wide variety or a reasonable variety of experimental tasks with human participants. I said, "Now there's something here. Now there's a real world to compare against."

Paul Middlebrooks
How did that take you to Lisa?

Eli Sennesh
Oh, sorry. Yes. I was trying to prompt you for the name Friston.

Paul Middlebrooks
Sure. That'd be Karl Friston just to be [crosstalk]

Eli Sennesh
Yes, Karl. Actually via, let's see, I was working with this embedded electronics company. I still have the hoodie over there. They had an MIT postdoc. He mentioned some of the Karl Friston stuff around the same time that Andy Clark's book came out.

Paul Middlebrooks
That's *Surfing Uncertainty*? Is that the one?

Eli Sennesh
Yes. That one had a lot of citations to people, names I already recognized. I read it. I went absolutely wild for it. He was mentioning, in the book, there's some people who are actually applying this approach to emotion.

Paul Middlebrooks
I see.

Eli Sennesh
Even better, the people who were applying this approach to emotion, Lisa and Karen, or at least Lisa and Karen locally to me. I was in Boston at the time, had a collaborator in this big interdisciplinary group that they had tried to form and maintain with varying success. It really shone for a while. I think the pandemic might have done it in a little bit. They had a collaborator, Jan-Willem van de Meent, who actually did the computational side, probabilistic programming. Of all damn things, I wrote a cold email knowing no better way to go about this. They actually answered.

Paul Middlebrooks
They were probably fairly thirsty for someone interested in it, because it's still not that widespread, right?

Eli Sennesh
Yes, as far as I know, none of this is widespread. If you take the Friston stuff too seriously, people say you're in a cult. I actually didn't join the cult until later. When I met Maxwell Ramstead, he eventually convinced me of a lot of free energy stuff, as least in part.

Paul Middlebrooks
Carl is famous for the free energy principle, and he considers it a framework, not a theory, by the way. People when pressed at least a couple years ago, he considers it a framework for thinking about the overall function of the brain instead of a theory for what it's worth. It has a lot of detractors and a lot of cheerleaders, and so you drank the Kool-Aid eventually.

Eli Sennesh
I sipped the Kool-Aid didn't go all in because by the time I was being given it to sip, I had been around Lisa and Karen enough that I had really absorbed, "No, you got to have your Evo-Devo, your neuroanatomy. You're mapping onto actual biology. It's the biology that really counts. I drunk my advisors Kool-Aid instead of the cults Kool-Aid.

Paul Middlebrooks
Okay. Good for them. Good for your advisors. Some people with your background would then, instead of embracing the biological neural plausibility would go the other direction, back to where it feels safer. I'll just say, we spoke for two minutes before I pressed record and you were talking about how NeuroAI is hard. Is it the neuro part that's hard?

Eli Sennesh
I don't want to make a public announcement. It's not like, "You know what? No one cares. No one is ever thinking about you. You're on camera, you're in front of an audience, but no one's ever thinking about you." Sure. I think I might well end up heading in the NeuroAI direction. I don't want to say as just for the money or for a career prospects thing. What I have noticed is that a lot of computational neuroscientists are renaming their work that now.

Paul Middlebrooks
Oh yes. I just got back from a brain initiative workshop called NeuroAI. I just got back from a Norway workshop called NeuroAI. That term is really being embraced because it sounds cool I think mostly.

Eli Sennesh
Yes. I've got this impression that the real difference between one thing and another is basically what was your training and what department are you looking for a job in? I think the number of departments that would do my completely ideal thing is null, and I'm sure most people end up saying that, well, before they go on the job market. I'm not going on the job market right now. Luckily, lucky me.

Paul Middlebrooks
What would that be? Can you describe what that would be?

Eli Sennesh
Ideally, really question-based or question-driven science, something close to cognitive science. In my Ph.D., I used to make up a fantasy field, computational affective science. By analogy to computational cognitive science. Now, computational cognitive science is already a fairly small subfield that often overlaps into the computer science departments because that's who will give some of them jobs. The number of cognitive science departments at universities, that do the full six-discipline, hexagonal multi-handshake thing is a handful.

Paul Middlebrooks
Or less.

Eli Sennesh
There's psychology departments who want you to do psychology experiments. There's neuroscience departments who often want you to do neuroscience, either theory or experiment, but they're defining the discipline often quite narrowly. I had a culture shock when I came to Vanderbilt and found out that what they mean by computational modeling or theory is basically biophysical or bust.

Paul Middlebrooks
That depends on who you're talking with, because you have people like Gordon Logan there also. I'm not sure if you run past him much, how active he even is still.

Eli Sennesh
I don't run into him, but at least I'm talking, let's say about my lab and a couple other labs that I interact with. There's a real emphasis on, be biophysical or don't do anything at all. Or be biophysical or give up theory and become an experimenter.

Paul Middlebrooks
Where do you sit in relation to that push? I'm trying to suss out your level of abstraction and what you think is important.

Eli Sennesh
My level of abstraction is that when I reached the end of my Ph.D., I said, "I formally did my Ph.D. in a computer science department. If I'm ever going to really investigate questions, I need to go get experimental training.

Paul Middlebrooks
Yes. You told me this a while back.

Eli Sennesh
Yes. I basically said, "All right, I'm going to go get as hardcore a postdoc as I can."

Paul Middlebrooks
That was the biggest mistake you've made. No, just kidding. Is that what you're saying the difficulty of NeuroAI is the joining of the two, that experimental and computational approach?

Eli Sennesh
Yes. It's not a mistake to go and get experimental experience, but it is a culture shock. It took me about six months to really be able to make progress on absolutely anything on the experimental side.

Paul Middlebrooks
Why is that? I know these things and people who do experimental work, we all cry together, and talk about how hard everything is.

Eli Sennesh
No, in my case, I'd rather not talk about it. It's private to the lab stuff.

Paul Middlebrooks
Suffice it to say that you run into way more problems than you would imagine you might. Would that be a summary of it?

Eli Sennesh
Yes. Way, way more. The thing is I had actually been prepared for the concept that you might walk arrogantly into experimentation with some grand theory and think, "This is going to totally be right." You do your first experiment and it's totally wrong, complete null result. In fact, that happened, but I was prepared for that. The part that I was much less prepared for is how do I even connect a theory to an experiment? The part that I wasn't, no results were a thing that I steeled myself, work it out, exercising, basically, just try to sweat until you can't be frustrated anymore that your theory is wrong. Oh, well, the theory is wrong. Even while you're submitting a theory paper about it.

Paul Middlebrooks
See, in the Popperian sense, that is the best progress, because it's an answer.

Eli Sennesh
Yes. It's an answer though. I hate to say it, but now that I'm looking at another way of analyzing the data, it might get more complicated again. Let me tell you about the actual experiment that we have in both mice and macaques. We have this thing called the glow paradigm, global local oddball. First, you give three identical stimuli per trial, AAA. This used to be done in auditory. Now we've been doing it in visual. Then, the local oddball is that fourth stimulus is B. It's something different. What the heck is a global oddball? In our manuscripts, we describe it as more complex oddballs.

A global oddball is where we set up the expectation for the animal. We try to intervene on the internal model and make it think there's a B coming, but then we give it an A. Let's say what we end up doing is testing. These are intermixed for the animal, about 80/20. 80% local, 20% global.

Paul Middlebrooks
You're really setting up the expectation.

Eli Sennesh
Yes. Actually, there's days and days of habituation followed by 50 trials of pure local oddball at the beginning of recording. That we're basically, habituating and queuing the expectations as powerfully as we can. What we're trying to do is disentangle, what happens if you have a predictable change versus an unpredictable repetition. The idea, from a neurophysiologist's point of view is that, then at the end, you're going to have a bunch of controls, those come after the main block.

After the main block, we record a series of essentially control sequences that are going to allow us to do statistical contrasts. The idea is to then eventually say, "All right, well, if you can figure out, if you can control for every other mechanism you can think of." Adaptation of the sensory neurons in V1.

Paul Middlebrooks
This is where it starts to get really messy and hard also.

Eli Sennesh
Not just messy and hard, but if you can control for everything you think of, and there's still some difference between global oddball, AAA, unexpected A. Just pure repetition or adaptation AAA, then, ah, now you've found a signature of surprise processing. For a long time, I have just been staring at this experimental setup going, "How is that surprise processing?" or, "What theory have we articulated about predictive coding in the Rowan Ballard sense that says this is surprise processing, rather than, who says the brain is tuned to look at angled gratings, moving angled gratings on a screen that flash on and off."

Paul Middlebrooks
In other words, you can't control for everything.

Eli Sennesh
It's not just that you can't control for everything. It's that, as I said, I believe, experiments is theory laden. If your theory is about the brain, predicting the continuous stream of sensory input. Then, flashing a series of angled gratings that are optimized essentially to drive V1 to a maximum degree. Under predictive coding theory, that's saying you're trying to drive-- You're trying to optimize prediction error. How do we expect to simultaneously optimize prediction error while also provoking another prediction error?

Paul Middlebrooks
That being surprise?

Eli Sennesh
Yes. That's the thing. Our setup conflates prediction error, surprise, visual change.

Paul Middlebrooks
Yes. Right. Because you're using that oddball. There's a visual difference in the oddball that you're using.

Eli Sennesh
Oh, and I should have said, actually, this is pretty much the standard paradigm as it turns out for studying predictive coding and goes back to about '09.

Paul Middlebrooks
You said that surprise and prediction errors are often conflated. What is the difference then between surprise and prediction error? Theoretically, perhaps maybe if [crosstalk]

Eli Sennesh
I would say you need to commit yourself to a theory in order for there to be a difference. The problem is if you're trying to test a particular theory, you should use the definitions from within that theory. Prediction errors within predictive coding theory, they're the residual when you subtract the prediction from the data.

Paul Middlebrooks
What the organism expects top-down signals, then it gets some observational data, bottom-up signals, and then there's a difference in the mismatch between the prediction and the actual observed data, and that's what gets passed forward.

Eli Sennesh
Exactly. How to relate that to surprise, I would reach for my information-theoretic definition, because I'm a quant person, and I would say, "Okay, well, surprisal is the negative log probability of the stimulus." Essentially, those would be two different quantities. When I eventually wrote my own computational modeling paper, prediction error was the gradient of surprise. They're related but distinct, and you have to use math to talk about how.

I guess I'm trying to just describe the culture shock of going from this environment, it wasn't oil and water, we mixed, but there was a very quantitative side that I worked on, and a very biological side. I come to this glow paradigm, this experiment, and I find that, oh, the quantitative side is just removed out from under me, I have to reconstruct it entirely myself.

Paul Middlebrooks
That's what you were getting at when you were talking about how-- What we'll call it NeuroAI is hard.

Eli Sennesh
Yes. Actually, there's this problem where in neuro, we are often doing paradigms or tasks that from a pure AI point of view might be considered almost trivial. From a biological plausibility point of view, that often makes them hard again. Then, if you're actually trying to explain neuronal data, or worse, trying to map some real theory of the brain onto neuronal data. Rather than just suggest that there could exist some mechanism explaining this behavior. There's been multiple computational models of the same behaviors. I'm thinking of the famous drift-diffusion models of decision-making.

How do you know if the brain is doing a drift-diffusion, accumulate evidence to a threshold and then decide algorithm for decision making. Or resource-constrained reinforcement learning algorithm for decision-making. There are experiments that have been fit with both these kinds of models.

Paul Middlebrooks
Yes, that's right.

Eli Sennesh
How do you know massive shock for me that there's just like, "Oh, wait, is everyone just pretending?"

Paul Middlebrooks
What do you mean pretending? Pretending that what they're doing is valid and what everyone else is doing is not or what?

Eli Sennesh
Pretending, just taking data and fitting it such that you can claim to use your theory to explain behavior, but you haven't actually tested it against substantive alternative theories rather than some null hypothesis. What the heck is our null hypothesis regarding behavior in the brain.

Paul Middlebrooks
Or alternative hypotheses? It doesn't even have to be null, just a clear alternative.

Eli Sennesh
Yes.

Paul Middlebrooks
There's something that actually that Jeff Schall, I'll just elevate him in this regard. Every year when I was a postdoc, there's a fundamental set of papers, one of which is the method of alternative hypotheses, where we tried to base. I think because of these things, because it's hard, like you mentioned drift-diffusion and I was doing drift-diffusion work, essentially, stochastic accumulator work, which is exactly what you're saying. Does the neuron ramp up to some threshold? That actuates the behavior. That's one of the things that Jeff Schall is famous for.

The idea is to look in the brain and test it and ask it through recordings. Of course, it's not super clean because we're dealing with different kinds of stimuli in this very controlled environment. The frontal eye field, as we know now, any given brain area doesn't just have a single function. There's mixed selectivity in brain areas where they're doing overlapping populations of neurons or doing overlapping functions things. If so, but anyway.

Eli Sennesh
Oh yes. Frankly, any talk of selectivity slightly makes me want to scream. I've just been re-acculturating myself to an environment where the word degeneracy, to an environment where these things are not the assumptions anymore.

Paul Middlebrooks
Wait, where degeneracy is not an assumption?

Eli Sennesh
Where degeneracy isn't the assumption, top-down influences often aren't the assumption. I'm not saying this as a negative thing in a certain way I like it, even though I don't think I can make a career out of it. Very Andy Clark quoting Quine, had this thing about desert landscapes. A neurophysiologist point of view is a very desert landscape point of view. There's the things I can measure, nothing else, nothing else exists. I'll talk about selectivity because I think I can measure it. If you tell me that that's actually caused by what I do rather than an observation of a causally independent system, then I will get in an argument with you because I think I'm measuring something real.

Paul Middlebrooks
I see. What you're describing, it's interesting that you find yourself in that world now, because in some sense, that's the old school world, which is still very much alive and thriving. Whereas there's been this recent push into a much more naturalistic types of tasks, and removing the constraints from the lab, the lab based experimental stuff. That's hard in a very different way.

Eli Sennesh
Let me make some applause or give some applause to André here. I think he doesn't do that kind of experiment yet because he's actually pushing something that's already very risky and innovative. He calls it Madeleine multi-area, high-density laminar electrophysiology. Which basically amounts to saying, "Let's have not just one Neuropixels probe in one area. Let's just cover the brain in Neuropixels probes."

Paul Middlebrooks
Neuropixels probes are like these really high-density, multi-electrode probes so that when you put them in any given area of the brain, you're getting recordings of hundreds to sometimes thousands of neurons.

Eli Sennesh
Exactly. All of our work includes the LFP, the local field potential, as well as the individual spiking signals, and then we analyze both together, which, I won't say who, but someone I really respect a lot. I went and visited their lab, actually one of my scientific heroes, I went and visited their lab at one point.

Paul Middlebrooks
You can't say who?

Eli Sennesh
Yes. I'm realizing I can't even specify this little-- The point being at one point I asked, "Do you analyze the LFPs?" They said, "No, we just look at the spiking." I think respect to André. I didn't talk about it before because it's not as native a part of my worldview. It's what I'm learning. This is actually a very ambitious thing. Even for a simple experiment, we'll have two full neuro pixels probes taking multi-unit activity, individual spikes that we sort with Kilosort, then LFP.

Paul Middlebrooks
LFP is what people talk about as measuring when people use the term oscillations. Sorry, [crosstalk].

Eli Sennesh
No, I was saying population-level signal.

Paul Middlebrooks
Oh, there's that too. It's a complimentary signal. The other thing is spikes are definitely the outputs of neurons, whereas LFP is thought to more closely track that population level input.

Eli Sennesh
Yes. Then we analyze both. We're often doing cross-correlation or coherence measures of LFP to spike. This actually tells you quite a lot and it's difficult. It's ambitious. My understanding is that it's also not easy to get grants in. I think Andre won his NSF career just this year. That was the first grant that the lab had gotten in, I think possibly three years of operation.

Paul Middlebrooks
Especially, for joint spike LFP analyses?

Eli Sennesh
For Madeline as a whole. For this research program of, let's measure in multiple areas. Let's measure the LFPs and the spikes. Let's try to capture as much as we can, so to speak, as many times as we can. Let's really try to push the limits on how dense the sampling can be in electrophysiology because of, essentially the resolution issues with imaging or EEG, that you would not want to use those. You would want to use electrophysiology.

Paul Middlebrooks
Backing up here. I was just at this BRAIN Initiative workshop and it was brought up multiple times. The idea was to think in terms of, well, what

would we need in 10 years? What's an ambitious goal for 10 years in NeuroAI? One person suggested this and then it was echoed by another person that what we need is to be able to record synaptic strengths. For example, neural networks, the strength between the units is where all the parameters are. That's those billions and billions of parameters that are in these large language models, et cetera. Those are what get changed, that strength in the connections.

If there was just a way for us to measure that in the brain, then that's an ambitious goal and it's a worthwhile goal. My immediate thought was, there's that age old question, "What would you do if you could measure all of the spiking from all of the neurons?" Would you even know what to do with it? No, the question is, "No, we don't because it goes back to the theory laden this." You have to have, you have to come from some framework or theory to then ask questions of that data. Just collecting the data is not going to get you there.

Eli Sennesh
Yes. I think that's where I'm just going to put my cards on the table and say, I think that's an open challenge for the field. I'm happy to be working on it.

Paul Middlebrooks
What is the open challenge, sorry?

Eli Sennesh
To figure out how the heck you analyze your data in a properly, theory-driven or question-driven way. I don't want to say this like it's too bad of a thing. Rather than just running statistics and then saying, I found an effect.

Paul Middlebrooks
That's interesting because that's what the AI side does in NeuroAI. It's like throwing a bunch of statistics at the data. Even Terry Sejnowski brought this up at the workshop. "What principles have we learned? What principles are there to gain from this approach?"

Eli Sennesh
Here's where I would reach back into my training with Jan-Willem as a probabilistic programmer and say, "For God's sakes, we need to be writing down generative models, fitting them to data, and then doing model comparison. We need to actually have some measure of how well does something fit the data, what theory motivates it, and then, compare them in a principled way." I think that machine learning can actually help with that. I've seen a lot of very productive, and a flurry of new work, essentially, in just analyzing NeuroData.

You also have to convince here's the hard part. Those things can get published in machine learning conferences. You have to both teach the experimenters to use them and convince them to use them. Teach it to them in such a way that they don't need you as a statistician or machine learner to actually, stand over their shoulder telling them how to encode every little hypothesis, because you want them to use it a dozen different times. They can't just keep you around forever as some consulting machine learner.

Paul Middlebrooks
Actually, so I'm going to-- It's not name-dropping because I wasn't like talking with him, but I remember Jeff Hawkins years ago giving a keynote, I think at the Annual Society for Neuroscience lecture, and I'm sure he's made this point over and over again. The traditional physics approach is you have your theorists and you have your experimentalists, and they're happy to play together. That's not the case necessarily in neuroscience. That we need to get to a point where the experimentalists are happy gathering the data to feed to the theorists who then can analyze it. That sounds awful to me, too.

Eli Sennesh
I will actually say I would much rather that experimentalists be capable and happy with analyzing their own data. The reason is that, if I say I'm going to be a theorist or a computationalist, then data analysis is something that pays the bills, perhaps. It's something that can help get a routine number of papers out the door for a machine learning person. I am actually thinking of someone, Scott Linderman over at Stanford, you'll notice that a lot of his papers are basically just machine learning based data analyses for NeuroData. That's great. That's the thing. That can build a career.

Now, personally, is that what I would want to think about as a theorist? How do we analyze data? No. That is not the thing that I have, a secret manuscript that I've been trying to finish for a year. The thing where I have a secret manuscript that I've been trying to finish for a year is, how do we explain emotion in a quantitative way or core affect, valence and arousal in a quantitative way by going all the way back to the urbiliterian and then picking C elegans as a model organism.

Paul Middlebrooks
Good luck with that.

Eli Sennesh
See, exactly. Good luck with that.

Paul Middlebrooks
You mentioned Scott Linderman. He develops a lot of tools that are being used in these naturalistic kinds of tasks. That skill-set seem seems to be what is really valuable in the academic marketplace, at least these days. Do you think I have that right?

Eli Sennesh
Yes. I'm going to use myself as an example instead of him, because, I know myself better. I don't think I could speak for the narrative arc of his career. I know that when I started my Ph.D., the starter project that I got put on was here's a new way of analyzing fMRI data in a little bit more theory-driven way and it worked.

Paul Middlebrooks
Sorry. Oh, you just needed to employ that method?

Eli Sennesh
No, it wasn't just, "Oh, there was some method and we employed it." We were building something new because our collaborator on the psychology side had some data, and he wanted to analyze it. The standard ways of analyzing it were inadequate to the theoretical question he wanted to ask. He wanted us to build something new, we built it, we published it. That gets citations, there was a follow-up. I think there's now follow-ups to the follow-up by other groups. This is going to sound horrible, but I don't mean it in a bad way. That stuff is good commodity science.

Paul Middlebrooks
It's also necessary. I can make it sound even better.

Eli Sennesh
Yes, it's like the Toyota of science. I drive a Toyota, I only bought a car this past year, but I drive a Toyota because, you know what? It's practical. That is very practical science that you can reliably never run out of new reasons to do more of it, and therefore, never run out of publications.

Paul Middlebrooks
That's right. This goes back to the idea of-- Does that contribute to progress in theory, progress in understanding principles, or is it just a very practical way to harness and say something about the data that's being generated?

Eli Sennesh
I think it has the potential to do both. By default, it mostly does the second one. That's not a criticism. That's to say, I think the field has the ingredients for a really great synthesis, laying around in different people's labs. What we need is, essentially, a small conference or workshops worth of cross-pollination. Where you can get the people with the appropriate skills all in the same room, give them the incentives to work together. I think it's actually the incentives that are the hard part.

Paul Middlebrooks
This idea of getting the people with the proper skill-sets in the same room for a couple days, it's awesome. The proper skill-set is a shifting landscape itself. Right now, we have a very specific one. People like you and Scott, whom you mentioned and stuff, where these commodities, these tools are extremely valuable, widely used, but going back to Hubel and Wiesel, they're on transparencies. They're putting just little shapes and trying to listen for the sound of neurons. Even Jeff Schall, whom I mentioned earlier, would tell us stories about, you're in lab, you'd make a little hole out of a wooden cutout, and you'd put a light up in there and is the neuron active or not. It's a very different world back then, very different skillset. I don't know how we track that. That's a meta-problem.

Eli Sennesh
Yes. that's why I say, if you're going to have a division of people's jobs or departments into theorists and experimenter, then I would want the experimenters to be able to analyze their own data because then they can do that even if it's a bit quantitative. Even if that's something of a moving frontier sometimes. Then the theorists, they can focus on asking questions like, "Well, how does the brain actually work now that we've measured it? Now that we're able to interpret the measurements."

Paul Middlebrooks
Let's get back to predictive coding though. You don't want to pin yourself into a very narrow corner, but where are you in terms of-- so the idea of predictive coding, predictive processing, is that we are constantly predicting what is coming into our senses. We have to have a model, to use the term loosely, of what we infer to be causes of things coming into our senses, inferred to be a cause in the world. We're making these predictions from our world model.

Bayesian brain hypothesis is one way to say it. Free energy principle is another framework implementation. Are you on board with this being a major function of the brain? Where does this sit in-

Eli Sennesh
Let's say major function of sensory cortex.

Paul Middlebrooks
Major function of sensory cortex?

Eli Sennesh
Yes.

Paul Middlebrooks
Why sensory?

Eli Sennesh
Going to lapse into neurophysiology vocabulary for a little bit. Sensory cortex is usually well-laminated. There's laminar sensory cortex down in these low areas. Then as you move both up the hierarchy towards cognitive areas, what we think of as cognitive areas, and also sideways over to motor, you get different patterns of lamination.

Paul Middlebrooks
The cortex is a laminar structure, meaning it has very repeated motif.

Eli Sennesh
Of six layers, like a layer-

Paul Middlebrooks
Six layers, let's say.

Eli Sennesh
Now, the one that raw sensory stimulus comes into, is layer 4. The thing is that when we talk about different lamination patterns, we're talking about, I believe they're called agranular and dysgranular, and those have either much less layer 4 or they're entirely missing it.

Paul Middlebrooks
I think that that's right. I think agranular is, has no layer 4, and dysgranular maybe has a weaker-

Eli Sennesh
Yes, a weaker layer 4. Now if you were asking yourself, "If I'm doing Bayesian computation, then my observed random variable, which is the stimulus, it has to come in somewhere. If I'm using this hypothesis about the laminar microcircuit doing predictive coding, then where's that coming in?" It's coming in, in layer four. What does the circuit doing if it doesn't have layer 4?

Paul Middlebrooks
That's where the generative network is, right?

Eli Sennesh
Maybe. Logically, it can't be doing variable-by-variable Bayesian inference. It could just store priors, but then why does it have a layer 2/3? Because that's the one that, computes errors and thereby updates the predictions now. I actually really, since we're following on Rajesh Rao's episode, I actually really like his hypothesis that, "Oh, 2/3 is the one that handles sensory data. 5/6 is actually handling chiefly motor data."

When you compute an updated sensory prediction, you might route it through there on its way somewhere else. Then, fundamentally, he would be saying, "Okay, now--" oh, and he also notes that there's thalamic projections into a cortical column that don't have to go through layer 4.

Paul Middlebrooks
The desire is to bypass layer 4, bypass layer 4 being a necessary part of predictive coding. Is that one way to [crosstalk]

Eli Sennesh
Well, or ways of reformulating the predictive coding hypothesis so that you can still have sensory data coming in even when there isn't a layer 4. Then you just have physiological and evolutionary questions about why are these areas agranular, dysgranular, laminar. What are the differences between them and the similarities? But you haven't totally abandoned your framework.

Whereas, if you're committed to layer four being where sensory observations come in, then logically, the Bayesian computation can only be done in our sensory cortex.

Paul Middlebrooks
Okay. I see.

Eli Sennesh
When I say I think I'm committed to this being an explanation of laminar sensory cortex, I'm being minimalist.

Paul Middlebrooks
Sure. You're on board with Raj's story about the incoming layer 2/3, outgoing layer 5, and how-- that's one way that it's biologically plausibly could be implemented. Your divide-and-conquer predictive coding also strives to be biologically plausible. Maybe we can start with, what is divided and what is conquered in divide and conquer predictive coding? Then maybe talk about plausibility.

Eli Sennesh
If you go look at some of the free energy papers, I think there's even one called The Graphical Brain. They tell a story about how a probabilistic graphical model has these different nodes representing different, unobserved random variables, and these get mapped onto cortical areas. Then the communication between areas is a series of messages and a belief propagation algorithm that eventually gets down to primary sensory areas where the random variable is observed.

Now, this kind of algorithm makes a very specific assumption that they call the mean-field assumption, about essentially saying, "We're going to approximate the posterior distribution with a product of independent representations." We'll have one representation for the visual, one for the audio, one that represents the integration of visual and audio, but they're actually all going to be statistically independent in the approximate posterior, scare quoting, as implemented in the brain. By the way, on the machine learning side, we know that this is quite a bad representation of a posterior distribution.

Paul Middlebrooks
Why is that?

Eli Sennesh
Essentially, it can't represent correlated posteriors.

Paul Middlebrooks
Because of the independence assumption?

Eli Sennesh
Yes. It's making a very strong independence assumption that was necessary to simplify the math in 2003. Literally, the first time variational inference was published was in a PhD thesis from 2003 or so. All my respect to people who are developing new things and make simplifying assumptions. Of course, the point of science is that we always want to try and relax our simplifying assumptions and ask, "Can we come up with a way to--" Essentially, can we assume that the real world is really complex and complexify our models over time so as to accommodate the real world.

Paul Middlebrooks
Then you're also dealing with Occam's razor, you're dealing with trying to figure out, "Well, what can we abstract away? What are the important things that we can abstract?" When you make assumptions like that mean field assumption, you are making trade-offs. It's just whether they're the right trade-offs given what you're trying to answer, right?

Eli Sennesh
Yes. What I had learned through my PhD on the machine learning side was that, if you have a complex structured graphical model as might be used in some cognitive science task, then mean field variational inference doesn't work very well. I thought, "Well, if I take a hypothesized model from neuroscience, and I apply it in AI, and it just doesn't work very well, is that what the brain does?"

No, I don't think the brain fails at things that are doable with current AI methods. Rather, I don't think the brain fails at doing things that we've observed it to be able to do in actual behavior. I think that's a case where the algorithmic model is just inadequate. I said, "Okay, let's make a better one. Instead of mean field independence assumptions, let's instead try to break down the random variables from one another so that you maintain their correlations when you update them."

Paul Middlebrooks
Is this the dividing part?

Eli Sennesh
Yes.

Paul Middlebrooks
Just say it again. What are you-- You're dividing-- Go ahead. You say it.

Eli Sennesh
You take this-- a probabilistic graphical model. It's a mathematical analogy to the brain's internal model of the world. You say this consists of a bunch of different variables that are connected to each other in various ways, like cortical columns, we can imagine. This was actually how I imagined it, was, one cortical column, one random variable. Then when they communicate with each other, those are conditional dependencies and all that.

Then I said, "Okay, let's try to divide this so that we can update each random variable in a way that takes into account the correlations with the other random variables that it's connected to."

Paul Middlebrooks
Is that the conquer part?

Eli Sennesh
That's the divide part. Then the update is the first step of conquer. Then the real conquer is that we have all of these importance weights from the world of like Monte Carlo methods for Bayesian statistics that then let us eventually write out, "Here's how good a fit to the joint model we have to the whole probabilistic graphical model." We're saying we want to do local updates that maintain some global coherence. It gets called divide and conquer because, well, frankly, NeurIPS is ultimately a computering conference.

All computing people have taken an algorithms class where they talk about divide and conquer methods.

Paul Middlebrooks
All right. I see. I didn't realize that. This is a well-known phrase in the algorithmic world?

Eli Sennesh
Yes. If you talk to algorithmicists and say divide and conquer, they'll say, "Oh, okay. You're taking some huge data structure and recursively performing the same computation on each component before going on to the connected bits."

Paul Middlebrooks
That makes a lot of sense.

Eli Sennesh
It just happened that you needed a lot of Monte Carlo tricks to make this work. When you do, it's very intuitive why you would want to do it that way if you were then going to map your probability model onto a physical circuit structure where the different random variables are spatially separated and have to signal to each other.

Paul Middlebrooks
In the divide and conquer model, what is required for it to be biologically plausible?

Eli Sennesh
The claim of biological plausibility we made is to say the computations are purely local.

Paul Middlebrooks
The local updates instead of globally.

Eli Sennesh
Right. People have talked about, could back propagation be implemented in the brain? Tommaso, our senior author has this paper on using Gaussian predictive coding to actually implement backpropagation as a substrate for backpropagation.

In this paper, we're saying, "Well, let's assume you can't do backpropagation. You don't have any kind of global computation graph or computation automatic differentiation tape in the brain, but let's assume that one cortical column can signal to another, and that if you're representing one random variable locally, then you can do really three things with it—sample from its distribution, measure the log density of its distribution, the log probability density, where density just means that you're talking about continuous random variables and not discrete ones. Three, take the gradient of the log probability density.

If you can do those three things locally, then you have the primitives necessary for our algorithm, and you can thereby obtain global coherence out of local computations. You don't need any backprop.

Paul Middlebrooks
Since we were talking about that experiment versus theory, a meta-science topic earlier, does this make clear predictions about what kinds of signals that you would expect to see?

Eli Sennesh
Now, here's where it gets biologically implausible. These were still rate-coded neurons, so they can still cross between positive and negative.

Paul Middlebrooks
Right. Brains use spikes, among other signals, like LFPs, but essentially, all of modern machine learning or AI models use rate codes. There are a lot of people working on spiking neural networks also, but I assume that if you're going to implement it in a spiking network, then you'd have to go-- it's plausible with the sampling approach, right? Because that's what spikes are all about.

Eli Sennesh
Spikes are all about sampling?

Paul Middlebrooks
Going back to the old debate on how probability is implemented in brains, there's the sampling approach versus the approach where the spike counts map onto some probability distribution types, et cetera.

Eli Sennesh
Oh, yes, with a twist. I know there's a lot of sampling approaches where you essentially say a neuron has a preferred stimulus and implements a likelihood function, and the priors are actually represented in the developmental program of the genome, not in the neurons themselves. Then those eventually make the prediction that-- they make the opposite prediction to predictive coding. They say, when the posterior probability of what the neuron prefers is higher, the neuron will fire more.

Predictive coding actually, and the free energy principle, and all of those approaches are much more information-theoretic. They say that, when the stimulus is thoroughly expected, you should see much less neuronal firing. We're in that family of theories, though we do use random sampling. My dispute with spikes being about sampling is that, of course, if you patch-clamp a neuron in vitro, then what is it? 96% of the variants, and it's spiking is explicable deterministically. There's stochasticity in the real brain, but we don't know that the single neuron is intrinsically stochastic that way.

Paul Middlebrooks
Right, that way. We do know it's stochastic, but-- Then going back to-- You started by saying that the major-- This is where it gets into non-biologically plausible mechanisms, is that it doesn't use spiking.

Eli Sennesh
Yes. Actually, I think Blake Richards Group has recently written to our rescue with their paper on-- what is it? Brain-like learning with exponentiated gradients.

Paul Middlebrooks
Brain-like because it uses spiking?

Eli Sennesh
In their case, brain-like because it obeys Dale's law. They'll have inhibitory neurons, which are negative, and excitatory neurons, which are positive, and the signs will never flip. They show--

Paul Middlebrooks
So they still use rate?

Eli Sennesh
Yes, they're still using rates there. How realistic do I think that is? I don't really know. There's areas that could use rate codes, but there's also too many experimental findings showing that precise timing matters. What it could be-- this is not an original thought to me. This is coming from a computational brain and behavior paper. I can send you the name, it's from 2020. It could be a prefix-free code.

Paul Middlebrooks
A what? I'm sorry, say that again.

Eli Sennesh
A prefix-free code.

Paul Middlebrooks
Prefix-free. What does that mean?

Eli Sennesh
That means that once you send a certain pattern of spikes, and by certain pattern, I mean the precise timing determines which codeword it is, but once you've sent a certain pattern of them, then that code word is over. Prefix-free means that no code word is a prefix of another code word. If I say ABAB, then that's either a full code word that now tells you something, or there's no full code word that starts ABAB except the one I'm already sending.

Paul Middlebrooks
What would that mean? Is that just because a rate code-- Go ahead, sorry.

Eli Sennesh
A rate code would say you listen over a certain period of time.

Paul Middlebrooks
Right. Whereas the timing or [crosstalk]

Eli Sennesh
You get M spikes, you divide by time T. Yes, whereas a timing-oriented code is you get a spike at time T, now you think, "Well, what comes next? Spike at Delta T. Delta T prime, Delta T prime, prime. You look for very specific timing, like with musical notes.

Paul Middlebrooks
In your lookup table, you figure out, "Oh, I just received this particular message."

Eli Sennesh
Yes. I just received ta ti ti ti ta. It's, gosh, has someone actually tried using 3rd grade music class on timing codes? A prefix-free code would then be a timing-based code where you say, "Once I've received a full code word, that's it. I know that I've received a full code word. I can interpret the whole message."

Paul Middlebrooks
You clean your cache and move on?

Eli Sennesh
Yes. Clean my cache and move on. Exactly. Really, gosh, on the other hand, that doesn't-- see, this is the thing that bugs me, is there's also evidence that dendrites are accumulating these precise spike timings into something more like a continuous signal that gets fed up to the cell soma. How can it be that there's precise spike timing and there's dendrites that converge from spike timing spike rate more or less?

Paul Middlebrooks
I don't know that those are necessarily problems. When you were going to say, "Is it a spike timing code or a rate code?" Because we know that some things require precise spike timing, like the interaural differences that underline how owls here locate a sound, for example. Timing is very important but maybe timing is not as important in, I don't know, frontal cortex or prefrontal cortex or something. It could be both, depending on what you're needing to accomplish.

An organ like the brain is fairly complicated, it turns out. It might be implementing lots of that degeneracy you were talking about. That could be the case in terms of how it computes. It's not maybe one or the other, but just depends on what's needed.

Eli Sennesh
Yes. That's very, very possible that, essentially-- actually, not only is that possible, that would go very well with some of our recent preprints that basically say predictive coding is a much more cognitive computation that can take place in frontal areas. Back to our glow paradigm, those global oddballs seem to get detected in frontal areas, but not in lower sensory cortex.

Paul Middlebrooks
Interesting.

Eli Sennesh
Maybe the laminar cortical column is something like a big stack of universal computational primitives that don't tell us much from just reading off the anatomy about what it is doing. Oh God, if we broadcast this, the modular-mind people are going to crawl out from under the rocks. We spend so much time banishing them.

Paul Middlebrooks
That's all right. There's room for everybody. One of the things I wanted to ask you about is-- you're mindful of what is and what isn't biologically plausible in this. You think it's important if you're going to understand—this sounds silly to say—if you're going to understand the brain that you need to implement through a model, you need to implement something that is biologically plausible, but you are willing to forego the spikes. Inevitably, any project is going to have hurdles. What hung you guys up the most in getting this thing to work and or getting it theoried out properly?

Eli Sennesh
Two big things. The first time was when I tried to write out all those waiting rules, essentially saying, "How do you accumulate the weights from doing a dozen successive updates to a random variable over a dozen passes?" I got something that looked really complicated and eventually just exceeded the numerical precision of floating point numbers in a computer.

What I eventually did was just have a meeting with Hao and talk out some options. He pointed out that one of them was essentially just cheating for getting the old importance weights and just saying, "I start with some particles." That is, "I start with some samples. I do a computation step on them. Now I have new samples. I'm going to do the same thing next time. I don't save any weights."

We ended up going with that because it turns out, once we both proved to ourselves that this was legal to do within all the rules of the game, this just turned out to be the simpler thing that was able to work.

Paul Middlebrooks
Storing the weights over time maybe is not even as biologically plausible as--

Eli Sennesh
Throwing them away.

Paul Middlebrooks
Yes. There were two things that you said.

Eli Sennesh
The other one is that, between the first preprint draft and the second one that represents our camera ready, we added this preconditioner that helps the optimization go in the right directions and respect the geometry of the latent space. This very mathematical, technical, itchy thing. The thing is, without that, stuff doesn't work, and you just don't perform very well on your test tasks.

Now we did manage to rig this up in such a way that it could be biologically plausible. It's effectively calculating a certain function of the prediction errors. If the prediction errors are locally available, then this thing is locally available. You could even nod to the free energy principle and say, "Ah, there's that precision of the prediction errors that these free energy guys are always on about." Really, it was just motivated by getting the damn thing to work.

Paul Middlebrooks
In the end, you have to have a working product.

Eli Sennesh
Yes. This is where-- I forget which famous person said that. No, two famous people have said this. It's Richard Feynman and Daniel Dennett have both said, "If you want to understand it, you've got to be able to build it."

Paul Middlebrooks
Did Dennett say that also? Feynman [crosstalk]

Eli Sennesh
He said a version of that.

Paul Middlebrooks
Did he? Okay. Feynman's is "I do not understand what I cannot-

Eli Sennesh
Build.

Paul Middlebrooks
-build.

Eli Sennesh
No. Then Daniel Dennett is completely different. He actually said, at one point, AI keeps philosophy honest. That's what I was remembering.

Paul Middlebrooks
That's interesting.

Eli Sennesh
Which is a whole other can of worms. My mistake. What I would say is, if you want to say that predictive coding is a thing that happens in the brain, based on your experimental observations, then it should hypothetically be possible to build an algorithm that does predictive coding and actually works for some of the toy tasks that we use in AI, which are still vastly more simplified than the tasks we use in neuroscience or rather the task of the brain.

An AI image-generating network does not have saccades. Unless it's one of Rao's, in which case, it does have saccades now, but that's very new for AI and completely trivial for neuroscience. I think you have to be able to build up AI to the point that it's able to do things that are trivial for neuroscience before you can really say, "Oh, a computational theory is viable now." No, it has to do the things that are most trivial for the brain.

Paul Middlebrooks
All right. Then I have two broader questions for you before we end our conversation today. One, just going off of what you just said, and been building up to this, do you need to understand the brain or brain processes or implement things in a similar manner to how the brain does things to build the best artificial intelligence? Do we need to mimic the brain, and at what level if so?

Eli Sennesh
I think that depends on how you define-- I'm sorry to be philosophical about this, but it really does depend on how--

Paul Middlebrooks
One has to.

Eli Sennesh
It depends on how you define artificial intelligence.

Paul Middlebrooks
Oh, geez.

Eli Sennesh
I don't like to commit to a definition of that at all, because what I personally want to do is understand the brain. That is the motivation for me. I want to understand the thing that actually exists, try to draw, so to speak, laws and principles from it. Then maybe I could engineer something with those in the same way that you can engineer a steam engine with Newton's laws and thermodynamics. You do have to do-- in my view, the interesting part is to do the fundamental science before the engineering.

Now, if you are engineering first, then an intelligent task is whatever the heck you have a benchmark for. There's this alternation between making a harder benchmark and beating the current benchmark. In that case, do you really need the brain? No, you need to understand your benchmark task. There's a lot of tasks where if you have a very deep understanding of the task itself, you don't necessarily have to understand how the brain would solve that task.

Paul Middlebrooks
There's all the talk of AGI in the AI world. We're going to get the AGI by next Tuesday. It's going to be the Tuesday after that. No. Then it's like five years. No. It's 20 years. Personally, going away from definitions again, I don't know what AGI is. I think that the humans are the wrong benchmark. It's like-- what's the right analogy? All we're doing is like staring at ourselves in the mirror and, yes, that's real intelligence. It's only because it's us. We think we're great, I guess.

Eli Sennesh
Oh, there I totally agree because, what is it? We got optimal chess playing at superhuman level. Maybe was that a decade before we got neural networks that could pass ImageNet classification at a human level?

Paul Middlebrooks
Yes. Half a decade, maybe. I think it was 2007.

Eli Sennesh
At the time, chess was the king task where we thought, if we understand how to play chess, we understand cognition computationally, or we've built intelligence. And then-- well, I don't even have to say "And then," there's a cliche for it, Moravec's paradox. I am very much a Moravec's paradox person where I say, "Understand embodiment first, sensory-motor stuff first, feeling first. Then maybe later in retrospect, you'll turn around and say, 'Here's all these normative principles we derived from our empirical study. We understand now how those tell us how to build what intelligence is and how to build it.'"

The term AGI almost feels like it-- I admire the sheer ambition of the people who are trying to do that and going to conferences like the AGI conference. The other angle on it is, unfortunately, that I do think, in the era of large language models, there's been a tendency to fool ourselves and define AGI down. That instead of being a name for something we don't understand and have to come to understand, only through working at it over time, it's become a name for something that we say has happened. The latest model from wherever is AGI.

Paul Middlebrooks
Is it AGI? It's got all of that.

Eli Sennesh
Right. It's like, okay, but that's because it talks. That's because it talks and we know the ELIZA effect. We know that if you talk and talk and talk, people will project personhood onto the words. To be fair to people, prior to the invention of LLMs, 100% of all linguistic stimulus we ever received came from other people. Except maybe for bad Markov chains and ELIZA, and that sort of thing.

Paul Middlebrooks
Right. ELIZA-

Eli Sennesh
The overwhelming super majority, for an optimal probabilistic reasoner, if you heard language, then the rational conclusion was that there's a person.

Paul Middlebrooks
We also know that some of us aren't that bright. For example, I've said, I think only ever Moravec, and you say Moravech. Which is it?

Eli Sennesh
I have no idea. I'm so embarrassed now.

Paul Middlebrooks
Oh, I'm sure I'm wrong. Anyway, that's the paradox that it turns out that it's easy to build--

Eli Sennesh
Is this the part from my thesis defense that I blanked from my memory where Mitch corrected me on this?

Paul Middlebrooks
I don't know. Anyway, that paradox is that the things that we think are hard to do, like chess, turn out to be easy. The things that we think that are easy to do like-

Eli Sennesh
Walking on two legs.

Paul Middlebrooks
-balancing weight might be-- or a waiter balancing a tray, walking through a restaurant or [crosstalk]

Eli Sennesh
Oh man, don't list that as easy. Talk to a waiter before you call that easy. That's hard.

Paul Middlebrooks
What I mean are the sensorimotor everyday things, the continuous sorts of behavior.

Eli Sennesh
Yes, but some things are hard even for embodied human beings. That's one of them. Go get a friend who works in the food service and ask them.

Paul Middlebrooks
I've been a server. I've been a waiter. That was the poor example. See, again, I say Moravec, I give bad examples. What do you do? Maybe none of us--

Eli Sennesh
I'm sorry, I'm not supposed to be shaming you on your own show.

Paul Middlebrooks
Yes, why are you shaming me on my own show? Sorry, I'm used to it.

Eli Sennesh
No, okay, you've been a waiter. Yes, it's easy for you because you practiced.

Paul Middlebrooks
I also have an ungodly balancing talent. No, that's not true. All right. I do have another question because you are interested in—how did you phrase that earlier—not consciousness, but--

Eli Sennesh
Feeling.

Paul Middlebrooks
Why anything feels the way it does. Another way to say it is just subjective experience in general or affect, I guess.

Eli Sennesh
Affect, the affective component of it. The classical dimensions of core affect are valence and arousal. Why do things feel pleasant versus unpleasant? Why do things feel exciting versus relaxing, what you could say, or arousing versus sedating?

Paul Middlebrooks
My question then is, and I was thinking about Anil Seth, who ties predictive coding into consciousness and that that's going to solve consciousness, essentially. What do you think about maybe that, but also it's real predictive coding's relation, possible relation to affect, the way you just described it?

Eli Sennesh
I have to say, I think the second one, the relation to affect through interoception, homeostasis, allostasis, this stuff, is a lot easier to establish than

anything about consciousness. That's why I've said, well, I'm not going to touch consciousness with a 10-foot pole. It's much too hard. Everyone's a little bit of a philosopher, but I'm not very much of a philosopher, so I'm just not going there. As to the connection between predictive coding and consciousness, here's one of the reasons I think consciousness is so hard to think about is that, oh, what is there? This classic thought experiment about consciousness? Couldn't you imagine a philosophical zombie who has the same input-output mapping and the same observable behavior, possibly even the same electrophysiological readouts as a real person, but isn't conscious?

Paul Middlebrooks
What about the affect aspect? Then they wouldn't have affect either, right?

Eli Sennesh
Right. If they don't have consciousness, it possibly makes sense-- What I would ask is, "Does a philosophical zombie have a predictive internal model? Do they have interoception?" I asked myself, "Can I imagine someone who has the same internal states and control systems at a physical level, but doesn't experience them at all?"

Paul Middlebrooks
The answer is no.

Eli Sennesh
Yes, the answer is just no because I'm like-- but there's a latent variable there. There's representations and computations going on. There's internal states maintained over time and internal dynamics. I can't imagine how there could be no one home. Like I said, I don't study consciousness because I recognize that this is, very likely, a limitation of my imagination rather than some kind of answer. It's just the way my intuitions work.

I prefer to be at least on the engineering end where I can bang an intuition against an experiment that doesn't work and bruise it until it's softer and can be remolded into another intuition.

Paul Middlebrooks
Now that you're doing experimental work, how do you think about the role of intuition? Sorry, I know this is another question and I've got to-- I actually have to go in a minute, but do you feel that your intuition has served you better from the computational world, theoretical world, or the experimental world? Because it all comes down to that. To make any progress, you have to make a guess. That's from intuition.

Eli Sennesh
Actually, I would say I don't know a good way to put those two together right now. I'm sorry, I just don't.

Paul Middlebrooks
Which two?

Eli Sennesh
The intuitions from both ends. Maybe if I was doing experiments with naturalistic behaviors, I would develop more of an intuition for how to let the experimental end drive. With the highly constrained experiments, I get an intuition for the task and the setup and the way that a particular data set or animal might behave, but not one for how do I pass from these spike trains to psychology, to the mentalizing I can do about the animal. I have no bridging intuition there whatsoever.

Paul Middlebrooks
See, now that I do "naturalistic experiments," meaning there's just a mouse running around in a box and we measure, measure, measure, now we're trying to relate neural activity to that ongoing behavior, which is continuous. They groom slightly differently. They move their paws slightly differently. Are we going to call that the same groom as the other one? How do we define that? My intuitions about experimental neuroscience, which were forged in that controlled constrained environment, I think are not serving me well. I'm trying to build new intuitions.

Eli Sennesh
If there's one thing I've learned in my life, it's really the limits of raw intuition and how you just have to bang up against experience long enough to start developing what you—let's end on a pun—call it posterior intuition rather than prior intuition.

Paul Middlebrooks
Exactly. You have to take action and update your posterior in the way that you phrased it. Eli, did we miss anything? We went haphazard, quite technical there. We remained out in the forest some. Is there anything crucial that we missed that you want to end on?

Eli Sennesh
Oh, actually, yes. There's this thing I always keep in my Twitter bio, "Abolish the value function." If I'm doing a podcast, I should tell people what that means.

Paul Middlebrooks
Yes. What does that mean? That's a great way to end.

Eli Sennesh
That means that, at one point in grad school, Jordan Theriault, who will probably listen to this—hi, Jordan—recommended this book to me entitled *More Heat than Light* by Philip Mirowski.

Paul Middlebrooks
Okay.

Eli Sennesh
Philip Mirowski is a very philosophical-leaning part economist, part historian. He wrote this whole book about the analogy between energy and the conservation of energy and economic behavior. All of this notion of there's an economic agent who maximizes utility or minimizes cost.

Paul Middlebrooks
That's the value function.

Eli Sennesh
Yes. All that stuff is the value function. What he pointed out is that, essentially, if you think like a rigorous physicist, the analogy is bunk. Economic value is not a conserved substance. People produce things that are valuable and then consume them. The amount of value is not a fixed constant number that stays the same all throughout all of this.

Paul Middlebrooks
Thinking like a rigorous physicist, would it be called an emergent property of production then?

Eli Sennesh
I'm not sure what Mirowski would say there, but his point was that in order to get all the math that was imported into economics, and then by the way, into cognitive psychology, into reinforcement learning, into optimal control, into all these things that we use in psychology and neuroscience, imported from economics—to get that from physics to economics in the first place, you have to assume a conserved substance. A conserved quantity, which represents a physical substance, on which you can then have a gradient flow, a certain kind of dynamical system.

Paul Middlebrooks
Absent that, where do we go? What is the result of abolishing the value function?

Eli Sennesh
If that's just the wrong metaphor, then I think we need to go into a much more control-theoretic frame of mind, where some signals represent references, and they can be directly compared to input signals from the bottom up by a comparator. Then when I shift from--

Paul Middlebrooks
-folks, yes. I've come around this in my theory things as well.

Eli Sennesh
Then when I shifted my point of view from all of these decision-making tasks, they're about grabbing more value, imaginary gold coins, like in Super Mario.

Paul Middlebrooks
Neuroeconomics.

Eli Sennesh
Neuroeconomics, yes, versus they're about measuring the distance between a desired outcome and a actual outcome, much more perceptual control theory sort of thing.

Paul Middlebrooks
The reference signal is somehow internally generated by which you compare, which is amenable to predictive coding.

Eli Sennesh
Yes. Then I thought, "Okay, well, now we've gone from substance to distance." These are completely different metaphors. Distance is the superior one, because as soon as you set up a mathematical model, you can measure the distance in the parameter space. By the way, that's actually the difference between reinforcement learning and active inference, in all of that free energy literature, is that the active inference people are saying, "Let's specify desired outcomes as target probability distributions, then measure the relative entropy distance from one to the other, and then just try to get closer to the desired outcome distribution." That's abolishing the value function from substance to distance.

Paul Middlebrooks
All right, Eli, I appreciate your time. Look forward to more work coming out, and good luck with the experiment. I'm sure we'll be in touch, but good luck with the work, learning more experimental research.

Eli Sennesh
The latest analysis seems to draw a very different conclusion than the ones we pre-printed, so we're going to have to reconcile those.

Paul Middlebrooks
Shocking. Yes. All right. I know you have an office mate there also needing to get back in the office, so tell them thank you for letting me take up some of your time. Thanks for coming on.

Eli Sennesh
Yes. Thank you.

[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 thetransmitter.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]
