How Anthony Zador thinks neuroscience can help improve AI

Artificial intelligence is ubiquitous and powerful, but can neuroscience still help advance it? Zador describes the "virtuous circle" of neuroscience and AI that drives progress in both fields.

11 November 2024 | by PAUL MIDDLEBROOKS

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

Anthony Zador
In fact, I'll go as far as to say, as far as I can tell, transformers are almost a counter-example to the successes of NeuroAI in that they bear, as far as I can tell, very little resemblance to anything that I expect to find in the brain. I'm particularly interested in three large questions about where you can hope to gain insights from biology, from neuroscience, and port them over to artificial systems. Alignment is, in some sense, the most fundamental. This is a crazy amount of flexibility. How does that happen? We don't know the details but the answer is, I think, that there is a developmental process. You might even call it a developmental curriculum.

Paul Middlebrooks
Hi, I'm Paul, I do neuroscience things at Carnegie Mellon University.

Tony Zador runs the Zador lab at Cold Spring Harbor Laboratory. You've heard him on "Brain Inspired" a few times in the past, most recently in a panel discussion I moderated at this past COSYNE conference - a conference Tony co-founded 20 years ago. As you'll hear, Tony's current and past interests and research endeavors are of a wide variety, but today we focus mostly on his thoughts on NeuroAI.

So, we're in a huge AI hype cycle right now, for good reason, and there's a lot of talk in the neuroscience world about whether neuroscience has anything of value to provide AI engineers - and how much value, if any, neuroscience has provided in the past.

Tony is team neuroscience. You'll hear him discuss why in this episode, especially when it comes to ways in which development and evolution might inspire better data efficiency, looking to animals in general to understand how they coordinate numerous objective functions to achieve their intelligent behaviors - something Tony calls alignment - and using spikes in AI models to increase energy efficiency.

If you like written essays, by chance Tony has written two essays, on the past and the future of NeuroAI, which are available on *The Transmitter* website and I think nicely complement our discussion on this episode. I'll link to those essays in the show notes, where I also link to a couple of the papers we discuss, and Tony's previous *BI* episodes.

Thank you to my Patreon supporters. I think it's time for a live discussion soon, so I'll be in Discord to see what kind of discussion you all would like to have soon.

Here's Anthony... Tony Zador.

Paul Middlebrooks
This is "Brain Inspired," powered by *The Transmitter*. Hey, first, just upfront, I just want to thank you for hooking me up essentially with the good folks at *The Transmitter*.

Anthony Zador
I'm so glad that worked out.

Paul Middlebrooks
Was just such a not random occurrence, but you and I were passing at COSYNE in the hall and just chatted for a minute, and then next thing you know, I'm speaking with Emily at *The Transmitter*. It just all really worked out, so I'm really grateful. Thank you.

Anthony Zador
I'm glad that worked out. I thought it would be a good match.

Paul Middlebrooks
You have NAISys coming up. I'm not going to be able to be there but I will see you maybe a month later in Bethesda at the Brain NeuroAI workshop.

Anthony Zador
Oh, nice.

Paul Middlebrooks
You're all over the place. You're organizing all sorts of NeuroAI stuff.

Anthony Zador
That's right. This is the time for-- Are we officially interviewing now?

Paul Middlebrooks
Yes, sure.

Anthony Zador
We didn't have a transition. I hope this is a--

Paul Middlebrooks
Tony, thank you for being on. Welcome. I appreciate you being on.

Anthony Zador
[chuckles] Sorry about that. I'm pretty excited about NeuroAI these days, both the work that I'm doing myself and the broader field and seeing what other people are doing, getting people with these shared interests together in a variety of settings. The NAISys meeting, the NIH meeting, which brings together not only a bunch of researchers but also, as they say, stakeholders, various branches of government funding agencies, NIH, NSF, a couple others. Then there's also one coming up at NeurIPS. That's only in the next two months.

Paul Middlebrooks
I know, it's all over the place. You may be the most excited person about NeuroAI that I know. Actually, I was just on a boat in Norway at a NeuroAI workshop, and at the very end, I was tasked with--

Anthony Zador
Wait, what? I thought they couldn't hold one without inviting me. What the--

Paul Middlebrooks
I was surprised. I was surprised, to say the least. I was assuming that you were invited.

Anthony Zador
I am hurt. No, I'm hurt.

Paul Middlebrooks
There was a no-Tony rule for sure. I polled the participants, there were about 30 people, whether they liked the term "NeuroAI". A few of them were very-- advocated for it somewhat passionately, which I was surprised by. Of course, some people don't like it. How is it sitting with you? You love the term, right?

Anthony Zador
Honestly, I wasn't a big fan at the beginning. It was actually five years ago or so that actually somebody I was working with here at Cold Spring Harbor, who's trying to help me put together a program here at Cold Spring Harbor on NeuroAI, we tossed around various ideas, and I think she can take-- Cat Donaldson, who's no longer here at Cold Spring Harbor, but I think she can actually take credit in, at least from my point of view, of really getting me to land on that as the term, because I had to write a couple proposals. I was using these three or four-word phrases, the intersection of neuroscience and AI.

She's like, that is just so clumsy. That is just so awkward. I actually did a Google Scholar search a couple of weeks ago to figure out when the earliest occurrence was. There's been a tremendous uptick. There's essentially no occurrences of the term going until the mid to late 2010s. There are a handful, but it's often used to mean something different, actually something closer to what we would call, what these days is called something like brain-machine interface or brain-computer interface. To the extent that it was being used, 20, 30 years ago. Apparently that's how people were using it.

Paul Middlebrooks
Part of my hesitation in adopting it, and I agree, it is the least clunky, clumsy way to say it, and it's clear what it means, I think. Part of my issue is even in the early days of AI-- was it John McCarthy who termed it artificial intelligence?

Anthony Zador
I think that's right. It was that meeting that he organized. I'm not sure who actually coined it. There was a 1956 Dartmouth meeting on what is now called AI.

Paul Middlebrooks
Maybe it was him who disliked it. There were some people who disliked the term. I appreciate that going back. There's been, not controversy, but there's been some debate on whether artificial intelligence itself is a good term to use. I'm on board with that criticism.

Anthony Zador
When I was a kid, when I was in graduate school, AI was the boogeyman. AI referred exclusively to symbolic AI. By contrast, the thing I was interested in was artificial neural networks, computational neuroscience, and machine learning. Very few people in my circles would have described what they were doing as AI. What they would say is that what they were doing was the opposite of AI. It actually took me a while to start even being willing to say that what I'm interested in is AI.

Paul Middlebrooks
What would they call it, connectionism? What would the term--

Anthony Zador
It's funny. The term "connectionism" was more specifically for a particular subset of people who did neural networks, the people who came more from the-- was the PDP books, the people who came mostly from psychology and who wanted a connectionist description of human cognition. Again, we could talk about all the substreams and the various approaches, but part of the research program there was to skip rule-based approaches, for example, in understanding verb conjugations, and to use a so-called connectionist approach. There was a lot of debate in the cognitive science literature about which of those is right.

Paul Middlebrooks
You've seen that historically developed in, and you've made your own transition into accepting the term more or less. Why are you particularly excited these days? I know it wasn't last week that you became excited, but because this has been a while now.

Anthony Zador
My own trajectory was that when I was starting graduate school, and I was in graduate school late '80s, early '90s, I was excited really about what at that point was one field, which was computational neuroscience/artificial neural networks. They were two sides of the same coin. There, the idea was that you would build models of how neural circuits compute, and then you would extract your understanding of what you thought-- or your beliefs, you would abstract your beliefs about how you thought neural circuits computed and apply them to build better machines.

There was no really, especially in the late '80s, early '90s, when I started getting involved with this, there was no clear distinction between those two. What I was excited about was both sides of it, because there was a long tradition in neuroscience of using quantitative models. That was certainly part of computational neuroscience, but typically those models were not constrained by the recognition that neural circuits not only have dynamics and have behavior, but they have to actually perform a function. They have to enable the organism in which the circuit is embedded to solve a problem.

I think it was really that constraint that was really crystallized by the early neural network combination or meeting with computational neuroscience. Sort of very naively. I would say that, if you ask a lot of people who worked on vision, and this is now before my time, and so I'm actually-- I was just recently trying to talk to people who were around back then to see if my inference about the history is true. My belief is that a lot of the early vision scientists, the people who were inspired by Hubel and Wiesel, didn't really see that there was a hard problem underlying vision.

In the same way that, for example, the early people who started working with computers thought that chess was going to be hard, but getting a computer to control an arm to pick up a chess piece was going to be easy. That turned out, obviously, even by the early '60s, to be false. I think that vision scientists took a while to recognize that the circuits they were studying were doing something really hard. The research program that I see based on published work following from Hubel and Wiesel, through the '60s, '70s, and '80s, was let us identify the representations in different brain areas of visual scenes, and that will be an explanation for vision.

Paul Middlebrooks
Are you talking the neocognitron days, or--

Anthony Zador
No. That is an exception, because that is somebody who's actually saying let's take what we believe is happening and put it together and see if we can build an artificial vision system. I'm talking on the physiologist side.

Paul Middlebrooks
That has stayed with us. It's not like that has really left altogether.

Anthony Zador
I think you would be hard-pressed to find, a visual physiologist who does not recognize that vision is a hard computational problem. They might say, look, what I can contribute to it is a characterization of neuronal receptive fields, and I'm going to characterize those receptive fields because that's what I can do, and I'm going to tell a story about the role that these receptive fields might play in visual processing. It would be very naive for a modern visual neuroscientist to think that that by itself was the answer.

When I was in graduate school, one might have imagined that it's pretty straightforward. You build a system with these receptive fields, and you've got a working vision system, and the reason we haven't been able to do that so far is we just haven't bothered. We haven't had enough engineers. We didn't have enough computer power. 30, 40 years ago, that might have been a viable argument, but now we know that if you try really hard and devote lots of computer power and lots of engineers, you get something now that works pretty well, but for a long time, you didn't. It remains a hard problem.

Paul Middlebrooks
In those early days, it was all Gabor filters also, and edge detections in the receptive fields. It's like, can you build up to vision from Gabor filters?

Anthony Zador
Exactly. Basically, you would break the problem into 12 or 24, or 50 sub-problems solved those individually. Like shade from shading and optimal-- I can't tell you how many papers I read at various times about different variations of optimal edge detection.

Paul Middlebrooks
Part of the reason you're excited is because of that history of-- thinking that it was an easier problem than it turns out to be.

Anthony Zador
What got me excited-- When I was in graduate school, I thought I could do both things. I thought that I could simultaneously learn about neuroscience and develop better-- take what I learned and apply it to building better systems. By the time I got to doing a postdoc, I decided that you had to pick one. At that point, the interest within the neural-- first of all, interest in neural networks was beginning to wane a little bit.

More relevant to me, I felt like there were real opportunities to learn more about neuroscience, and I felt like I should just take this approach that I had and do the right neuroscience experiments. I retooled as a pure experimentalist or as an experimentalist driven by theoretical and computational questions. For, my postdoc, I worked on synaptic physiology. I still was interested in quantitative approaches. There was an excitement at that time about using information theoretic techniques pioneered by Bill Bialek, and so I was pretty excited by those.

Paul Middlebrooks
How do you think about those-- what do you think about information these days?

Anthony Zador
It is the core of how I approach problems. I think the exercise of measuring information, which is what a lot of us, including me, were doing at that time, it is useful for thinking clearly about how collections of neurons represent information.

Paul Middlebrooks
At least their capacity to, right?

Anthony Zador
Their capacity to do so. There are lots of limitations to it, but I think it-- the actual numbers that you get out turn out to be boring and sterile, and I guess that's the complaint that people have, is that what's, for me, compelling is the framework, the way of thinking about the world, the recognition that among other things if you're going to process information, that information has to be propagated and transformed and--

Paul Middlebrooks
You have the bits, and then what the hell do you do with them, is the--

Anthony Zador
That's right. It's part of the problem. it addresses part of the problem that we have in the nervous system. I guess one thing that changed for me, and I think for a lot of the field, was that back then, a lot of us were focused more on sensory processing, and so information theory covers a lot of how you take information in from the outside world and you represent it.

How do you characterize those representations? What it left out was something that I didn't come to until really I started my own lab, which was the recognition that animals behave. This is not a shocker. Back then, many of us were studying anesthetized animals, and so the behavior that you get out of an anesthetized animal is usually somewhat less interesting. If it's--

Paul Middlebrooks
Wait. Can I just pause there for a second?

Anthony Zador
Please.

Paul Middlebrooks
There are so many instances, and I can't think of an episodic memory of one, where people point to historical experimental work like anesthetized

animals. Then they talk about how it was a necessary step to get to where we are these days. Using different animal models, using them in controlled experiments versus natural, et cetera. Do you think that that's true in this case with the anesthetized animals?

Anthony Zador
I think people going back to ancient philosophers have asked whether any particular path through history is an inevitable path. Was it inevitable? I'm not sure it was inevitable, but to this day, there are experiments that are-- let's even go more extreme. For my postdoc, I studied brain slices. We certainly can learn a lot of things in brain slices that would have been hard to learn in the intact preparation, especially back then. A crazy thing is that when a neuroscientist refers to an in vivo prep, we refer to one in which the brain is still inside the skull.

If a biochemist refers to an in vivo prep, they are talking about one in which all the proteins are still inside the cell membrane. Different preparations for different questions. I'll never ding a preparation. I'll never find fault with a line of research. What gets problematic is when the community forgets that this is a model, this is a limited model of something. When the community becomes sufficiently large, it begins to talk only to itself. Then those questions take on a life of their own that is independent of how they were originally formulated as part of an overall research program for a field of, I would characterize it as understanding how brains work to control behavior.

There are lots of different ways of characterizing it. I don't think that most people, when they started their careers, would have honed in on the representation of a static visual image in the anesthetized brain of a cat as the central question that they wanted to set out to answer. That is a super useful question to answer, and it's great to have a model system where lots of people agree on the preparation, you can really make progress.

Ultimately that, I don't think, was the fundamental question that drove all these people to work on that preparation. It was a model of a larger question, things like, how do you represent the world outside of you? Then probably people were asking, the original people who wanted to know, how do those visual representations get used by the animal? Or what is thought? Things like that.

Paul Middlebrooks
Do you think that the modern deep learning approach, NeuroAI approach, runs any risk of falling into that same error of mistaking the map for the territory?

Anthony Zador
That's a leading question. Do you have anything in mind there?

Paul Middlebrooks
What do you mean, like specific criticisms?

Anthony Zador
Yes.

Paul Middlebrooks
Because people think that, for example, a transformer is doing cognition or something. That's a very simplistic way of saying it, but it's still a model. It's not the thing. It's not the ultimate question. It's still a model, but it's a lot closer, perhaps. I don't know. what do you think?

Anthony Zador
Transformers, I don't think of as-- in fact, I'll go as far as to say as far as I can tell, transformers are almost a counterexample to the successes of NeuroAI in that they bear, as far as I can tell, very little resemblance to anything that I expect to find in the brain. Their success basically derives from the fact that they are extremely well matched to our current generation of GPU hardware. That's great. I am blown away by ChatGPT. That is awe-inspiring.

Paul Middlebrooks
It feels it's necessary to state that everyone's blown away,

Anthony Zador
I am more blown away than-- I'm especially blown away, I'll say, because right up until the week that I played with it, I was doubling down on my belief that you would never have a large language model that was any good at all without grounding. In fact, I was having arguments with people who worked at DeepMind and Google who were already playing with these. I was like, no, it won't work. They were like, no, it does work. I was like, no. They were like, no, you're wrong. Yes, I was wrong.

Paul Middlebrooks
What do you think the role of ground-- I'm sorry I'm jumping around. What do you think the role of grounding is now? Do you think it has an importance?

Anthony Zador
We can jump to that. What I think is that ChatGPT taught us something that I don't think we would have learned any other way. That was not obvious to me. Which is that in some sense, language is a closed system, in the same way that arithmetic is closed under the integers. Basically, it's

very hard to break ChatGPT by asking it questions. It gives reasonable answers and has reasonable things to say about almost-- I know there are countless examples on-- honestly, the fact that it can't do arithmetic, that it literally can't do arithmetic, I don't see as a condemnation at all.

That's a ridiculous claim. I can only do arithmetic for large numbers if I follow the algorithm, and that's how to add numbers. That's not what ChatGPT does. I think that what we've learned there is that it can give a reasonable answer to almost any question that's formulated as a string of words. That is super interesting, and to my mind, surprising. Early on, 3.5, you could still break by asking it, which I thought was significant, by asking it something like, what is the problem with making shoelaces out of uncooked spaghetti?

Paul Middlebrooks
Oh, that's a good one.

Anthony Zador
Because that required some knowledge of the physical world, and I thought that's what you wouldn't have.

Paul Middlebrooks
That's the grounding. That's your grounding.

Anthony Zador
That was an instance of grounding, and I was pretty smug. I was like, "Ha ha ha." No, GPT-4 can give you a long exposition on the problem of using uncooked spaghetti for shoelaces. I cannot find anything like that anymore that breaks it. There are a couple of the goat, the cabbage, and the man crossing a river that is said to break it. Do you know?

Paul Middlebrooks
No.

Anthony Zador
You know the classic logic puzzle. You have a goat, a cabbage, and a wolf, and you want to get them across a river, and you have to figure out the right sequence to do this.

Paul Middlebrooks
Oh, okay. They can't [crosstalk]--

Anthony Zador
What's that?

Paul Middlebrooks
Nothing. That's vaguely familiar to me.

Anthony Zador
It will give you the right answer, but then if you give it a variant of that, it will not pay attention, and it'll answer reflexively. Various people like Gary Marcus will hold that up triumphantly and say, that's an exception. I think a lot of lazy people who aren't thinking clearly would also make the same mistake. Or, "I can't operate on him. He's my son. How is that possible?" You know that one?

Paul Middlebrooks
No.

Anthony Zador
You don't know these paradoxes.

Paul Middlebrooks
I'm not part of the cottage industry of [crosstalk]

Anthony Zador
I know. It's a two-minute digression. Not worth it. Anyway, the point is that I don't find those breakages interesting. If you ask what is the limitation, it's the things that can't be done. There are hallucinations which are interesting, and that may or may not turn out to be a solvable problem. It seems like people are making a lot of progress on solving it. The point is, and this, I think, is important, is that language is only a tiny bit of what we do. For me, that's really the key point, is that we do an awful lot of things. In fact, we are the product of 500 million years of evolution.

Language, although I'm very impressed with it, and I think it's key to our success, probably emerged, depending on whether you think neanderthal had language, somewhere in the last couple hundred thousand, maybe a million years at most. We've had 499 million years of evolution. Language is just this extra on top of that. All that other stuff is the stuff that remains incredibly hard for artificial systems. We have vision systems that kind of work on static images. They're really impressive. You can take a picture. One fun thing is you can-- someone sends you a picture of themself standing in with some background in some random city, you can upload that picture and it'll say, oh, that's Sophia, Bulgaria. Wow, who would have known? Go ahead.

Paul Middlebrooks
I shouldn't have said transformers because like you said, that's an example of the opposite. I was thinking more in terms of like the convolutional neural networks with recurrence and that NeuroAI push in terms of understanding our brains.

Anthony Zador
Convolutional neural networks are the example that those of us who advocate for NeuroAI hold up every single time to the point where anyone who doesn't finds that example annoying. I'll say actually there's an even more fundamental example, which is the idea of neural networks in the first place. The idea that you're going to compute with a whole bunch of elements that are connected with variable parameters. It's not obvious that we would have gotten there had we not been inspired by squinting and making an abstract model of the brain.

If the question that usually comes up, and maybe this is what you're getting to, is what are we going to do in the future? Here's the argument that people usually have, which is, sure, brains early on inspired artificial neural networks in the same way that birds inspired planes. We do not design planes based on careful study of birds. That's the argument now. One counter-argument that Yann LeCun likes to bring up is that apparently aerospace engineers do actually study birds and get cool ideas from them. I'm agnostic to that. I'll defer to him as someone who apparently has read up on aerospace engineering.

For me, the more fundamental counter to that is simply that, we are building systems in this analogy that we want to be as birdlike as possible. We define success as building the most birdlike bird we can. It's true that a 747 can do amazing things that a bird can't do. It can fly, whatever, 10,000 miles. It can carry lots of tons of cargo. It can go 500 miles an hour. In the same way computers can do all sorts of things people can't do. They can multiply big numbers. They can serve up queries for Google, whatever. That's not what we're asking.

That's not what we would like artificial systems to do better. We would like artificial systems to do better at what birds do, which is to swoop from the sky and pick up a fish. If we wanted to build a birdlike bird that could fly through the forest fairly quietly without using too much energy and swoop out of the sky, and grab a fish out of the water, then we would probably do well to look very carefully at how birds do all these things.

In the same way, if our goal is really artificial intelligence which is the ability to do anything a person could do, if that is the most generic explanation, then we should probably figure out how people do what they do. I would argue that the path to understanding how people do what they do is to look at how animals do what they do, because people have very little, I would say, that's novel over what our ancestors did. Definitely.

Paul Middlebrooks
That's what you refer to as alignment. I asked you offline, what you're excited about. One of the things that you listed was alignment that you wrote to me that our current models for how to formulate objective functions for reinforcement learning and stuff are very limited. You think that we should look to the animals for that.

Anthony Zador
I think that's exactly right. I'm particularly interested in three large questions about where you can hope to gain insights from biology, from neuroscience, and port them over to artificial systems. Alignment is, in some sense, the most fundamental. We currently are very good at building systems that are pretty good at building systems that maximize a well-defined objective function. If we say we want multiple objective functions, the answer is usually, we'll just add a couple of terms to the original objective function. Objective one is lambda one plus objective two times lambda two, and so on.

Choosing those lambdas, that's a very impoverished way of representing objectives. In most cases, that particular expansion hasn't worked particularly well. It's brittle and it hasn't been effective. By contrast, animals are necessarily expert through evolution at balancing multiple objectives, the so-called four f's, feeding, fleeing, fighting, and romance. We have to balance all of those. When we're hungry, that might take precedence, but at some point, no matter how hungry we are, if we're about to get eaten by a predator, we should probably put our hunger on hold and flee.

Romance sometimes takes a backseat to all of those only when the other three are dealt with. This is the top-level objectives, but those are broken down into sub-objectives and sub-sub-objectives. Humans and other social animals have social objectives that are as compelling and profound as hunger. There's a biological architecture that allows evolution to introduce new objectives that interact appropriately with the existing ones. I don't think we know how that works in biology, and I don't think we understand how to do that in artificial systems.

Paul Middlebrooks
If you ask, some of the cognitive architectures were a big thing for a long time, and they still are, but one of the issues or one of the things that those people learned in trying to build those systems is that the coordination between the modules is a harder problem than the actual objective functions in the modules.

Anthony Zador
Exactly. I think that's exactly right. Yet, I think that this is a case-- I make the argument that NeuroAI is really this virtuous circle, this virtuous cycle between taking insights from neuroscience, applying them to AI, using AI as a model of neuroscience. In the same way that I don't think vision

scientists understood just how hard vision was until they took their fuzzy ideas and tried to build a system on them. I don't think even the people who work on various motivations understand how hard it is, as you say, to coordinate them.

We won't really understand how hard it is to generate behavior from an agent that has a whole bunch of objectives until we start trying to build such agents. That will define the problem better even for the experimentalists. We're still, at the experimental level, still trying to define what are the signals for reward. Obviously important groundwork, but it doesn't necessarily get us to the really hard problem or what may turn out to be the really hard problem of coordinating a bunch of these things simultaneously.

Paul Middlebrooks
This goes back to the question of what you want your artificial intelligence system to do. I don't know, replicating us is not-- maybe that's not the best use of building these things. Why would we want a system to have all the hard-won evolutionary coordination dynamics among our, whatever, 16, 17 objective functions? Why would you want them to have to battle that out, have to implement all that?

Anthony Zador
I want a robot that washes my dishes. I want that robot also not to step on me. I want it to be aware if my house catches on fire and do something appropriate about that.

Paul Middlebrooks
What about the romance part? Do you want that too?

Anthony Zador
I do not want romance. I will tell you that what drives technology traditionally has been romance. If you look at the history of various technologies, it turns out that the rise of VHS was apparently driven by, let's say, romantic movies.

Paul Middlebrooks
Oh yes. Is that true? I didn't know that. [crosstalk]

Anthony Zador
It is. Movies featuring private romance were one of the main drivers of VHS, certainly one of the main drivers of the early Internet. I believe it's still one of the main sources of traffic for the Internet.

Paul Middlebrooks
I think so.

Anthony Zador
I have no doubt that romantically inclined robots will be a huge market. Not my own personal dream, but--

Paul Middlebrooks
Wash the dishes and have a little romance.

Anthony Zador
That's right. In practice, the particular objectives that we want that robot to be guided by may turn out, will undoubtedly turn out to be very different from our objectives. What's important is not the content of those objectives, but the computational framework for trading them off appropriately. In our objective function, survival features pretty heavily for most humans. Even that, that can be relaxed. An individual ant does not really care that much about its own or I guess her own survival. An individual ant, in large part, because it's a clone of all the other ants in its colony, is very willing to sacrifice itself for the good of the colony. All other things being equal, it'd rather not die, but it's not mostly focused on not dying.

Paul Middlebrooks
There's still something at stake.

Anthony Zador
Sure. Same thing with a robot. You don't want your robot randomly walking into a lava pit. You do want your robot not to value Asimov's three laws of robotics. Because survival, I believe-- don't harm another. Do what other people tell you unless it violates the first, and try to survive unless it violates the other two. I think that's roughly what they were. Maybe we need something richer than that. Certainly we don't necessarily want it, given how we envision programming or instructing our agents these days, it's unlikely that we'll lay it out in words. Although maybe with LLMs.

Paul Middlebrooks
I had you on a really early episode. We talked about your paper that argues that most of what's useful is actually innate due to evolution over time. You're still on that, not bandwagon, you're still on that idea, but also you've incorporated development as something that would be interesting to study in terms of how-- is this related, your interest in development, is it related to this coordination of the objective functions?

Anthony Zador
Absolutely. That's right. The original bandwagon upon which I hopped was the--

Paul Middlebrooks
You built the bandwagon, right? You helped build it.

Anthony Zador
Turns out other people were on this bandwagon. I don't know.

Paul Middlebrooks
You gave it a fresh set of wheels.

Anthony Zador
Maybe I put some flags on it. I don't know. Anyway. I'm excited by the idea that much of all behavior, all animal behavior, and humans are animals, derives from deep, innate drives, and this is true at every level, that we just don't have time to learn everything from scratch. In fact, I would go as far as to say that learning can be seen as on a continuum with and really an extension of simply development. If you buy the idea that most of what we have, most of our neural circuits, most of therefore our behavior is determined by our genome, which specifies a neural circuit, then that--

It took me a while to come around to the idea that that really requires that you pay some attention to the relationship between the genome and the final brain you get. The process by which biology takes a genome and produces a brain is called development. I thought for a while that I could ignore the biology of that, but it turns out that, number one-- mostly I wanted to ignore it because I didn't know much about development and I'm still woefully ignorant.

Paul Middlebrooks
Me too.

Anthony Zador
It turns out to be pretty interesting. There are principles that you can abstract from it, and that maybe can help guide how we approach these problems. One of the core principles, for example, is the idea that the process by which you derive a brain from a single cell involves a process of the recursive application of a relatively simple set of rules, and then when necessary, those rules are modified across developmental time.

Paul Middlebrooks
People like Robin Hiesinger, who would say-- he's very focused on development. You make this point that there's not enough coding capacity in our DNA to specify the entire structure of, for example, our human brains. He would make the point, and I don't know how you think about this, that what the DNA is doing is actually encoding those recursive rules, and you have to have that. That development is necessary. You can't just go from DNA to the computer.

Anthony Zador
100%. In fact, Robin was just here for a meeting at Cold Spring Harbor last week. We had a wonderful meeting of minds. I'd never met him in person. I'd read his stuff, We're 100% aligned on that. I would say I'm now hopping on his bandwagon. The recognition.

Paul Middlebrooks
The development bandwagon?

Anthony Zador
The development bandwagon, yes. The recognition, exactly as you said, that-- just to back up. I had the idea that I was pursuing was that the genome is a compressed representation of our wiring diagram.

Paul Middlebrooks
The bottleneck.

Anthony Zador
The bottleneck. It represents a genomic bottleneck. In fact, just this week, a paper, our very first attempt to formulate that rigorously, that paper was finally published. It was out on bioRxiv for, I don't know, the last four years. It was finally published in PNAS work with Alex Kulakov, who's a fellow professor here at Cold Spring Harbor. In that vision, in that version of it, we formulated the problem of compressing a weight matrix by using another smaller neural network. We're compressing the weight matrix of a neural network by using another smaller neural network to predict the weights of the final neural network.

Paul Middlebrooks
Is that like an autoencoder?

Anthony Zador
It's not quite an autoencoder. Basically, you have a weight matrix and the weight matrix is N by N. You have N squared elements. Then you have a smaller neural network whose input is a pair of elements, two indices of the larger weight matrix, and its output is a prediction of that weight. That worked. we got great results. We were able to compress big matrices from MNIST and CIFAR and ImageNet into factor of 100, factor of 1,000, and the compressed weight matrix basically could perform almost at the same level as the uncompressed one, the one after learning just out of the box. On top of that, we showed that these compressed representations led to better transfer learning, suggesting that when you compress a weight matrix, you're throwing away the junk, and you're keeping the stuff that's important.

Paul Middlebrooks
More generalizable.

Anthony Zador
Yes, exactly. We saw compression as a regularizer. That worked. Then more recently with Blake Richards, we had another version of this inspired by some-- which is now on bioRxiv and under review. It will be another six years before it's published where we use cell types and stochastic connectivity among cell types. That also works, and it has somewhat different properties.

Paul Middlebrooks
In a sense, you've solved that bottleneck problem?

Anthony Zador
I would say that both of those were fun and great learning experiences, but the stuff that I'm working on right now with a postdoc named Stan Kerstjens is really driven by some ideas he had for how you can formulate that developmental process recursively, how you can grow a network using very simple rules. This network grows and can be guided to produce a final result that solves tasks. That, to my mind, has captured some of the key elements of development. These recursive-- What's that?

Paul Middlebrooks
Would that process also then be more efficient?

Anthony Zador
That's what we're interested in finding out. What I think is is that it represents a prior on-- the possible circuit. Any set of rules you have for compression represents a prior over the circuits that you can generate. You might not be able to articulate that prior, but if you have a small thing making a big thing, then it's only going to be able to make a subset of the big things just by information-theoretic arguments. The subset of things that it can produce and which ones it learns most quickly, that represents a constraint on the set of networks and a prior over those.

Paul Middlebrooks
You said the word I was thinking, constraint.

Anthony Zador
Exactly.

Paul Middlebrooks
I've come to think of constraints as-- It's like that coordination problem. Constraints are everywhere. In some sense, they're more important than the process is the constraint.

Anthony Zador
Exactly. People will argue that the success of artificial neural networks is really that they represent a smooth prior over data. There's a lot of work on that. It's the same idea. You can't know ahead of time what the right set of constraints is, what the right set of priors is. This is an experimental question. The proof is in the pudding. In this case, the fact that the prior at a very high level from 30,000 feet looks like the prior that guides the formation of actual neural circuits. The idea that every neural circuit ever in existence in biology arose from a single cell.

Then the set of rules that took you from one cell to many cells has to have fit in the genome. That maybe is one of the key constraints. Anyway, that's something I'm pretty excited about now. We'll see how that plays out, but I feel like that's the right way to go. There's also a very natural interpretation of that for evolution because what evolution does is it produces a circuit. You start with a circuit, and then the circuit is good at doing some things. The organism in whose brain that circuit is good at performing some behaviors and perhaps not others.

Then you select for animals that are maybe somewhat good at performing some other behavior, and then they will develop, possibly, in the next generation. If the rules of development are such that the animal's circuits get even better at that behavior, then you select for those. Then you add on to the existing circuit. You add on new abilities, new circuits, but you have to make sure that every generation, you can produce an organism that has a brain that can do all the things you want it to do.

Paul Middlebrooks
Why do you need development in there, though?

Anthony Zador
In biology, you need development because you start every generation with a single cell. You have to give a plan for how you go from one cell to a body and a brain connected to the body. Personally, I'm at this point more interested in the rules for generating the brain, but honestly, if we're ever going to understand robotics, we might want to think about the fact that bodies also are built that way.

Paul Middlebrooks
Earlier you alluded to your reluctance to take on development, and I felt-- It scares me, essentially, because it seems so hard. Then I imagine the AI field would want nothing to do with development or think that it's something that biological organisms have to go through because they're coming from one cell, but that might not contribute to AI, for example.

Anthony Zador
Sure. Let me tell you about a case that I've been thinking of. We haven't made any progress on it, so I'm giving away my research ideas, but that's okay. I think I'd be thrilled if other people pursued them as well. I've been thinking a lot about robots recently. There's a problem that we don't really have terribly good robots. They're not very good at interacting with the world. For a while, I think there's a community of people who are excited about these physics simulators. I've been playing around with them like MuJoCo.

In these physics simulators, you can specify agents that have arms and legs, and they're connected by something like muscles, and you can apply forces on them, and then you can learn policies to control them in these artificial environments. I think they're thought to be pretty realistic simulators. They try to make them as realistic as they can. It turns out to be remarkably hard to build, even in these reduced physics simulators, agents that can walk around.

It was a real teaching point to me how hard it is to make a simple agent walk around in a physics simulator. We fooled around with that a little bit, and many other people have done serious work in these. I think that my understanding is that relatively few serious roboticists spend a lot of time in simulation because the problem of taking the simulated agent and bringing it into the real world is basically unsolved. It doesn't work. It's the so-called sim-to-real problem.

Here's my thinking on that, is there is a similar, if you like, sim-to-real problem that confronts us, in that we have a genome that specifies a body and a neural circuit. The specification of the body need not be terribly closely coupled to the specification of the neural circuit. You are born with a brain that had better be able to fairly quickly learn to control whatever body you are born with.

Paul Middlebrooks
There's an exception to this. I was just talking with Karen Adolph. She studies human motor development and has spent years studying children. The rate at which they fall and run into stuff starts super high because they're exploring this space. They don't necessarily have to come out with-- Like a horse, you use the example, can walk in a minute or two of being born, but humans, we're awful.

Anthony Zador
It's a great question why humans-- Humans are always the exception. Once we understand animals, it will be interesting to understand why humans are the exception to a lot of what I'm saying. Why is it that it takes us many years to learn how to walk? I think it's pretty clear that it's not because we couldn't have learned to walk more quickly. In fact, I'll go as far as to say as a-- My kids are older now, but I remember when my kids were younger and I would have preferred them to take even longer to learn to walk because if they don't have the common sense not to do stupid things, there's-- I think you have kids. How old are they now?

Paul Middlebrooks
9 and 11, and I'm just-- There's a breakthrough where I'm trying to make my son ambidextrous, which is impossible. We've been doing a lot of throwing with left-handed and stuff. It's hard.

Anthony Zador
You remember that period where they're toddlers, where they now have the ability to locomote, but not the good sense, not to locomote to the wrong place. They have the ability to pick something up and put it in their mouths, but not the good sense not to put the wrong thing in their mouth. I think this long delay before we can even learn to stand is just a reflection perhaps of the fact that the ones who learned to stand too early made bad decisions.

Anyway, going back to animals, I would go and point out that you can mate a Chihuahua and a Great Dan. Might be a little awkward, but it can happen. Basically their DNA is compatible. The same circuit, what that tells you is that the instructions that build a Chihuahua brain are essentially indistinguishable from the instructions that build a Great Dane brain. That's two orders of magnitude. When you run MuJoCo, if you build a typical agent in a MuJoCo physics simulator, at least in our hands, if you train the agent to control its body, and then you change the body by 10%, 20%, it breaks.

We're not talking 10%, 20%, we're talking-- a Chihuahua's a couple of pounds, a Great Dane is 200 pounds. This is a crazy amount of flexibility. How does that happen? We don't know the details, but the answer is I think that there is a developmental process. You might even call it a developmental curriculum. At each step, you solve sub-problems of the overall problem that somehow enable this entire brain-body combination to learn to walk and run within a few months, even a few weeks. I think that there are other--

I'm picking Great Danes, but I think even those dogs, and I think even those probably, the evolutionary pressure wasn't as high to get things moving immediately. You could probably use a pony and-- I don't know the names of a large horse.

Paul Middlebrooks
Anheuser-Busch horses, whatever those are.

Anthony Zador
Exactly. That's probably only one order of magnitude difference in size, but I think the same argument holds. Those guys can walk within a day or two.

Paul Middlebrooks
This reminds me of-- You use the word curriculum. Curriculum learning, and you can correct me, is that idea that you just mentioned of instead of learning to-- Let's use tennis serve as the common example. You don't just go and just do the serve. You learn how to stand, you learn how to bend your knees, and then you do those separately, and that's curriculum learning, and then you can put them together.

Alexander Mathis was just talking about this and how it actually helps to teach an artificial system how to do something like that. I guess that's what you're saying.

Anthony Zador
That's right. There has been a fair amount of work in ML, machine learning, on curriculum learning. Usually, people push back and say that the failure there is that it's too hard to choose a good curriculum. I guess I would say that probably for many problem domains, that's right.

If your goal is to learn image recognition, to build a system that does image recognition, and your goal is to train it faster, I guess, it's not clear what role curriculum would play there. You might take a guess as to what are good building blocks, but by the time you've tried a whole bunch of curricula, you may as well just have used your data and trained end-to-end once.

I think that's how it got a bad name because it's been applied to solving a different problem from the one that I outlined here. Where the curriculum is potentially useful, at least one place where I see it as potentially really useful, is if you have essentially the same problem that you need to solve over and over again with slightly different constraints, slightly different formulation, a slightly different brain, a slightly different body. Now, we're in a domain where that curriculum could really pay off because I think-- Go ahead.

Paul Middlebrooks
I'm thinking in terms of Karen Adolph's studies and stuff, just because she's been on my mind. You're thinking in terms of taking inspiration from how-- Let's say, an example like as a baby crawls, the way it can even hold its head is different. It looks at different things at different times. Then as soon as it can sit up and move around better, then its outlook on the world, the way it actually takes in the world changes and it's scanning a different set of things.

Again, it's the same problem of getting visual information in. Then by the time they're walking, they're like these -- Is that what you're talking about? Using that from development as inspiration, because they're solving slightly different--

Anthony Zador
That's right. They're solving different problems and the solution to one problem provides a foundation for the solution to the next problem. This is where evolution comes in, is that evolution, in some sense, provides guidance as to the sequence of problems that were solved. I don't know if you've yet had Max Bennett on as a guest.

Paul Middlebrooks
It was a good book that he wrote.

Anthony Zador
Yes. He lays out a particular collection of five problem-- a brief history of intelligence. I recommend it to all your-

Paul Middlebrooks
He's one of the speakers at NAISys.

Anthony Zador
He's one of the speakers at NAISys, yes. Beautiful book. Just frustratingly good in that I was thinking of writing a book, and now, I don't know what I'd write. He wrote a much better book than the one I was envisioning writing, but in any case, he lays out five big problems that-- Yes, there you go. That's right. At this point, I've recommended his book so widely I feel like I deserve a fraction of the royalties that he's getting.

Paul Middlebrooks
It's a very well-written book.

Anthony Zador
Yes, it is.

Paul Middlebrooks
There's the five breakthroughs.

Anthony Zador
The five breakthroughs. I don't know if it's exactly right, but has the flavor of a really nice framework to think about the problem. The idea is that you can't get to the second breakthrough until you've had the first one. Then there's this old adage, which has a great thumb truth in it. We could even talk about why it's not quite true, but ontogeny recapitulates phylogeny.

Paul Middlebrooks
Oh my God.

Anthony Zador
In other words, development replays the evolutionary history up to a point. The natural interpretation is that you have an agent, you have an animal that can perform a bunch of things. Those things are so important it can do them at birth, and then it learns other things that turn out to be useful. Those animals that are particularly good at learning those other things quickly get selected for. The fastest way to learn something quickly is to not have to learn it at all, but just stuff it into your genome, or stuff it into your genome as much as is possible.

That's what led me to say earlier that it's extremely hard to distinguish between development and learning. Basically, the faster you are at learning something, the better your priors for learning it are, the less information you need from the world to learn it. That all happens by packing it into your genome.

Paul Middlebrooks
Are you familiar with Justin Wood? His work?

Anthony Zador
I'm not.

Paul Middlebrooks
You should check him out. He used to be a staunch nativist and now he is a staunch empiricist, and it's because of his work. He does controlled rearing of chicks. When they're hatched, they go straight into this automated box where they have complete control over what they see and what they can do and stuff. Some of his research findings have-- Everything he's finding is leading him to think that everything is learned and nothing is innate. You guys should have a conversation perhaps.

Anthony Zador
I don't think I ever want to have an arm muscle with the nature-nurture crowd. The answer is it's both.

Paul Middlebrooks
Of course.

Anthony Zador
Do you famously-- I had a friend who used to always ask me, "Do you walk to school or carry your lunch?"

Paul Middlebrooks
That's good.

Anthony Zador
That said, he's wrong. I have no idea. I just googled some of his stuff. After we're done, I'll take a look at it.

Paul Middlebrooks
Sure. I interrupted us. You were talking about packing information into the genome.

Anthony Zador
Yes. I would say that part of-- Going back to AI, one problem this curriculum issue could solve or address or provide a way forward on is it you could imagine in simulation, figuring out what the right developmental curriculum is. Then what success would look like is you pick a series of 10 things or 20 things that an agent would have to learn and sub-problems. The hope, the expectation would be that if there exists such a curriculum, learning those 10 things, the sum of the time it takes to learn those 10 things is shorter than the amount of time it takes to learn the thing that you're ultimately trying to learn, for the sake of argument, walking.

That then if those 10 things are relatively straightforward, you could maybe follow that same developmental curriculum in an actual agent with the idea that, yes, it's still going to have to learn the body, like all the differences between the body would have thought it had from development or from the simulator and the one that actually got in the real world. Maybe those differences, if you break that into pieces, are smaller than trying to learn the whole thing end to end.

Paul Middlebrooks
You wouldn't want to create a robot that develops.

Anthony Zador
I might want to create a robot that maybe is born into its body, but then learns how to control the nuances of its body because really what I want is to be able to build many robots and not have to spend a year each time I build a new robot body.

Paul Middlebrooks
You're not 100% on board with all facets of development then.

Anthony Zador
No, no, no. Look, that's always the rule from, at least-- That's always my idea when looking to neuroscience for guidance, for inspiration. I don't even know what it would mean to incorporate all the details from biology into an artificial system. The only thing that has all the details of biology is biology, right?

Paul Middlebrooks
Yes.

Anthony Zador
I spent a fair amount of my graduate work studying single channels and I think they're fascinating. I like studying channel kinetics. I think these are fun problems. I don't think they're at all relevant as far as I can tell to anything I'd ever want to put in an artificial system. How you make an action potential is a cool problem that stands on its own as far as I can tell. Maybe someone else will come around and explain why it is relevant, but I will certainly not want to build an artificial system that has sodium and potassium channels. At least it's not obvious to me why I would.

Paul Middlebrooks
That's below my line as well.

Anthony Zador
Sure, I'm open to the idea that neuromodulation is important. Obviously, it's super important for how animals work. I don't personally yet know how to-- This just reflects my ignorance. I'm not saying that other people don't know how to do this well, but I don't know how to abstract the principles of neuromodulation in a way that makes them useful for an artificial system. I'm not just going to put-

Paul Middlebrooks
We don't know the principles of how they're useful in biological systems. I think we're still pretty far away from feeling like we have that even close to being tackled.

Anthony Zador
Exactly. The tremendous success of convolutional neural networks, which were inspired by Hubel and Wiesel. That is a great example that not all aspects of receptive fields were stuffed into a neural network. It is the idea that you can now, in retrospect, to any hindsight, we can boil it down and just say, "Look, it's the idea of translational invariance," which maybe could have come from some other angle.

Maybe you didn't need to study receptive fields, but that's how the idea was hatched. That's my attitude toward how these things-- I'll just go and make one other comment though about the appeal of a curriculum as opposed to end-to-end learning and the idea that a curriculum that is rooted in the actual evolutionary path that humans follow to get where we are, why that would be useful.

I think a lot about the example of self-driving cars. Self-driving cars don't work that well. I was surprised. I just recently read an article, apparently even Waymo, which is pretty widely deployed, there's basically literally a room full of people who are helping the Waymo cars out when they get into trouble.

Paul Middlebrooks
They need grounding, right? Is that what they need, grounding?

Anthony Zador
I don't know. No. Waymo, apparently there's a control center and a bunch of people's-

Paul Middlebrooks
Online?

Anthony Zador
Yes. This was a *New York Times* article a couple of weeks ago. There are people sitting in a room somewhere and I think it said every three to four minutes, they are called upon to help out one of their cars, which is having trouble.

Paul Middlebrooks
I said online, what I meant is real-time. They're doing this in real time.

Anthony Zador
Real-time. Waymo is driving around and then, oh no, there's a yellow cone. What do I do? Then there's a guy sitting there saying, "Veer to your left." Okay." I think he has a little mouse or something. I've white-knuckled it on a self-driving Tesla before and it's an exciting experience, but-- Why doesn't it work? The hope has been that just increase the amount of data by another order of magnitude and another order of magnitude, and you'll start fixing all these problems on the long tail.

I guess my argument is twofold. One is that, again, the reason driving is pretty easy for us is that we already understand everything we need to know about how to parse visual scenes. We didn't have to learn how to parse visual scenes by sitting behind the wheel of a car. We didn't have to-- Yes?

Paul Middlebrooks
There's also the frame problem, which is still a problem, and that we have solved the frame problem to know what's relevant when we're driving.

Anthony Zador
Sure. More generally, we have solved the problem of being able to, in a novel situation, figure out what's relevant and what isn't. Then we learned the details. I'm teaching my 14-year-old how to drive and it's--

Paul Middlebrooks
Talk about white-knuckling.

Anthony Zador
[laughs] He's very good. It's definitely a learning experience, but we've gone out a half dozen times and there's been dramatic improvement. He has to fill in some of the blanks, but I guess the other problem in some sense, I would say that this is even a more fundamental problem. Even if we get to the point where self-driving cars on average make fewer mistakes than people, which we may get to, we're going to be very disappointed if the mistakes they make aren't similar to the mistakes we make.

If a self-driving car on average has a much better track record than a human, but it occasionally just runs down a kid in the middle of the street who you're going to review the video and you're like, "What the heck happened there? Anyone should have seen that," that is going to limit our enthusiasm about adopting these self-driving cars. I think this is a general principle that-- What is going on there? If the objective function is don't run things over, you're going to get a system out of many systems that runs things over as little as possible.

In order to ensure that it not only does a good job, but it fails in the same way we do, we need a system that is as isomorphic to the way we do something as possible. It can't just not make mistakes, or it can't just make as few mistakes as possible. When it makes mistakes, they have to be human-like errors. That-- Yes, go ahead.

Paul Middlebrooks
The problem with that, though, is that we're not great drivers. You and I are above average. Everyone's above average, right?

Anthony Zador
I'm not sure what-- We can mostly agree on what are reasonable mistakes to have made. I think that's the key point here. The fact that we do make mistakes, those are mistakes of lapses of attention, et cetera. I expect that those are mistakes that the artificial agents won't make. They hopefully won't get distracted, they hopefully won't be talking on their cell phones. They won't be driving drunk.

Paul Middlebrooks
Romancing and driving.

Anthony Zador
They will not be romancing and driving. That is--

Paul Middlebrooks
That's what we can do while they're driving, right?

Anthony Zador
That's right.

Paul Middlebrooks
Before-

Anthony Zador
Just to that point, when we have robots that drive and romance, they will know that they shouldn't mix the romance and the driving if we can get their objective functions right. If we have a complex multi-objective function. They will know-

Paul Middlebrooks
You're on board with thinking that is a harder problem than implementing an objective function itself.

Anthony Zador
Yes.

Paul Middlebrooks
Tony, what have we missed here? We've gone through those three main things, the reasons that you were maybe the most excited about NeuroAI.

Anthony Zador
Yes.

Paul Middlebrooks
Did that paper, the one that maybe was it a year ago now, the Catalyzing the Next Generation of Artificial Intelligence from Neuroscience Principles, something like that-

Anthony Zador
Oh, wow. You know the titles of my papers.

Paul Middlebrooks
Something. That's close. Did that get much pushback?

Anthony Zador
No. I don't think anyone bothers. My take on it is that-- When I started, like I said, there was widespread agreement among the leading AI researchers, Jeff Hinton, Yann LeCun, Yoshua Bengio, people like that. they were excited about neuroscience, they wanted to learn about neuroscience. They were very clear that these things were important. I think what has happened is there have been several generations of engineers, AI scientists who never actually had direct contact with neuroscience. Maybe in an introductory lecture, they heard that in prehistoric days, AI, machine learning had something to do with neuroscience, but it was prehistoric for them. If you go to NeurIPS today, my guess is that fewer than 1% of the people there have any interest in this potential for Neuro to have any impact on AI. I think that's fine. If my goal is to deploy a commercial LLM for a recommender system or whatever, something that digests law documents, there's absolutely no reason that any of these people should be learning about neuroscience.

Even the next level of optimizing algorithms which takes a lot of work and a lot of theory, it's not clear to me why those people should care about neuroscience. I think the bet that I'm placing, and I think some very small fraction of the overall community is willing to place, is that just continuing on our current trajectories without some really new ideas may asymptote before we get to where we want or where they want.

Paul Middlebrooks
I was just talking with Kim Stachenfeld. One of the things I brought up is didn't DeepMind, in a sense, fail because their original mission was to use neuroscience principles and now they've outgrown that with scale, et cetera? That mostly they've moved on from trying to find inspiration in that sense.

Anthony Zador
Have they failed or have they abandoned their mission?

Paul Middlebrooks
I use the word failed because it's a hotter take.

Anthony Zador
I think there's no world in which you'd say DeepMind has failed.

Paul Middlebrooks
No. I know. I was just being--

Anthony Zador
You're being provocative. Even there, my understanding, and she and others know as well as I do, is that they're under some pressure to deliver

their term, and I think there are people there who are disappointed that they're being asked to-- I think this is happening also at Google Brain and elsewhere. People are being asked to actually get away from the basic research that they were doing and help build a better-- Look, from a corporate strategy point of view, when, your profits are only $40 billion a year, you could see how you might get nervous.

I think these are strategic decisions. I think basically there's a gold rush going on with LLMs. Each of these companies is trying to figure out how to capitalize, how to cash in on that. I don't blame them. If they miss the boat on that, they're cooked. You don't want to fail on that, but I don't think that necessarily has much to say about the medium 5 to 10-year time horizon of potential impact of the intersection of neuroscience and AI.

Paul Middlebrooks
I see that paper getting cited a lot. I've not looked at citation counts. You probably know that, but-

Anthony Zador
No, I don't look at them either.

Paul Middlebrooks
It seems to be a pretty popular-

Anthony Zador
For a small group of people, it helps galvanize their interest. It helps focus their interest, and I don't think anyone cares enough to say, no, this is a waste of time. The people who do think it's a waste of time, and like I say, I think that's a large fraction of the community, they just ignore it. The only reason that people would push back, now that I think of the sociology of it, is if we were fighting for limited resources. By doing this, we were, getting a larger slice of the pie. That pie is so enormous that this is barely even a crumb. Right now it's not in anyone's interest to argue against this research program.

Paul Middlebrooks
Thank you for taking me on that meandering walk through your recent work. Keep up the good work. I'm glad that you're delving into development, and that I'm not.

Anthony Zador
[laughs] Can I just say one final word?

Paul Middlebrooks
Yes.

Anthony Zador
Two words. One is delving is a word that's overused by ChatGPT, and I'm surprised to hear-- Oh, you didn't know that?

Paul Middlebrooks
No.

Anthony Zador
It is the most overused word by-- I'm now suspicious that you are being powered by ChatGPT.

Paul Middlebrooks
Oh. Should I remove it from my vocabulary, or do people say it too much?

Anthony Zador
Certainly from your written vocabulary, because that is the-

Paul Middlebrooks
Shoot.

Anthony Zador
- marker of ChatGPT. Anyway.

Paul Middlebrooks
All right. I will undelve.

Anthony Zador
What I was going to say about development, and this was the one point I wanted to make, is that I realized that the reason that I was turned off to development is that over most of my professional career, development has been a pretty uninteresting list of molecules. I just thought of it as some of the least important. That list of molecules is super important, but unless you're in the field, you don't want to-- at least I don't want to just

memorize a list of molecules or learn any new ones, but there's a previous generation going back even to Turing. Von Neumann and Turing both worked on development.

The question of how you build a system from a single building block, how that thing can make copies of itself and self-organize into a global structure, I think is a really interesting problem and not utterly unrelated to the kinds of problems that neuroscientists and AI people think about. That has been the realization that I've had as I went back to generations of literature before I started watching or not watching talks about development.

Paul Middlebrooks
I didn't know, or I had forgotten that about von Neumann, but for Turing, that's his lesser known, but still really cool work, those cascade instability work.

Anthony Zador
Yes. Anyway--

Paul Middlebrooks
Thanks again, Tony. Have a great conference, by the way. I'm sorry I'm not going to be there.

Anthony Zador
I'm sorry you can't make it.

Paul Middlebrooks
I'll see you in a month or so.

Anthony Zador
See you in a month.

[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]
