THE TRANSMITTER | BRAIN INSPIRED

Xaq Pitkow shares his principles for studying cognition in our imperfect brains and bodies

Pitkow discusses how evolution's messy constraints shape optimal brain algorithms, from Bayesian inference to ecological affordances.

27 AUGUST 2025 | by PAUL MIDDLEBROOKS

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

[music]

Xaq Pitkow
The hope is that you can find some principles that make things understandable. In a sense, the only things that are understandable are the non-cluges, right? The only things that are understandable are the principles.

This is a really important and I think underappreciated element of what we have to do is we're not trying, in my view, we're not trying to come up with the one mechanistic explanation of something. We're trying to find a class of equivalent explanations that have some shared properties.

To me, this is the biggest gaping hole in neuroscience, is that we don't understand how learning works. All of the machines that we use for learning, they're doing gradient descent these days. That's basically what they do. The brain doesn't do gradient descent, maybe it approximates it. What are the approximations? What are the constraints? We don't know, and we don't know because we can't measure it yet.

[music]

Paul Middlebrooks
This is "Brain Inspired," powered by *The Transmitter*. You think you have principles? Xaq Pitkow has principles. Xaq is my guest today, and he runs the lab LAB at Carnegie Mellon University. LAB here stands for Lab for the Algorithmic Brain, and an acronym for that is LAB, which stands for Lab for the Algorithmic Brain. An acronym for that, well, you get the point.

Xaq is a theoretical neuroscientist with a background in some experimental neuroscience as we talk about. I think he actually describes himself as a dabbler. He dabbles in many endeavors, but the main theme of our discussion here is how he approaches his research into cognition by way of principles from which his questions and models and methods spring forth.

We discuss those principles, and in that light, we discuss some of his specific lines of work and ideas on the theoretical side of trying to understand and explain a slew of cognitive processes. A few of those specific topics that we discuss are how when we present tasks for organisms to solve in order to understand some facet of cognition, the organisms use strategies that are suboptimal relative to the task, but nearly optimal relative to their beliefs about what they need to be doing, something Xaq calls inverse rational control.

We talk about probabilistic graph networks. We talk about how brains use probabilities or how brains may use probabilities to compute different ways they could use probabilities to compute. One of his newer projects is ecological neuroscience that he has started with multiple collaborators. These just touch on a few of the many projects that he is running, has run, and is interested in.

You can learn more about his principles and about his work in the show notes at braininspired.co/podcast/219. Thank you so much to my Patreon supporters. If you support the show, you get access to all the full episodes, the full archive. You can join the Discord community. You can access a bunch of complexity group meetings. That is a biweekly-ish discussion group that we've formed around the foundational papers of complexity. Look for my David Krakauer episode a couple months ago if you want to learn more about that.

Anyway, I hope you're doing well out there. I'm going to have a new studio soon in a couple of months. It won't be this tiny, tiny closet that you see before you. All right. Enjoy Xaq.

[transition]

Xaq, I'm going to give it a shot and then you can correct me. I'm going to, in the broadest terms possible, describe what you do, and then--

Xaq Pitkow
Awesome.

Paul Middlebrooks
Or at least a common theme, the broadest possible common theme, right?

Xaq Pitkow
I like it.

Paul Middlebrooks
Okay, so you study normative models under realistic assumptions to discover, or infer cognitive functions, so the realistic assumptions being like metabolic cost, limited resources, the computational cost of cognition, rationality under sub-optimality and so on. That was super brief. Where did I go wrong? How would you correct me?

Xaq Pitkow
Oh, that's a great start. That's a major theme of what we're working on in the lab. I think those things are really fun. I've definitely been motivated by principles and different kinds of principles. It's a normative principles are a pretty natural one, but there's also, and you mentioned some non-normative principles in the sense of constraints that come from inside the brain. How do we end up with those constraints? Some of them are still principled, like in physics. That was how I originally got interested in this whole endeavor.

Paul Middlebrooks
Stop there. What do you mean?

Xaq Pitkow
I saw a talk by Bill Bialek when I was an undergrad and he showed how you could use physics to understand the brain. I was like, really? Yes, that's amazing. That was the beginning for me. Some of those constraints we can see, as well as the dimmest light possible according to physics. That's a constraint that comes from physics. The balance in our eyes between resolution and refractive blur, diffraction, that comes from physics. Those constraints come from physics. Then there's some other legacy things that show up that doesn't-- maybe it doesn't have to be that way.

There's some architectural structures that are there and I'd like to understand those as well. Those are harder to get at because you don't have physics to point to. That's just the legacy of our evolutionary history and the ecological niches that we occupy. How can we understand something of how we end up? For example, one concrete example is that learning and plasticity in the brain is largely local. That's a constraint that comes from our brain wiring.

Other systems like AI systems are not constrained to have local learning rules. That's a particularity of our brains and our biology and our systems, bilateral symmetry that's inherited from a long time ago. Those are things that are not necessarily optimal in some sense, but there are some influences that come in.

Paul Middlebrooks
So-- Go ahead.

Xaq Pitkow
I was going to say another major effort that-- so those are all normative or normative adjacent things that we work on, but we also have some other types of things that we work on. I'm known for doing a bunch of work on correlations of different sorts, like what those do for you, where they might come from, and we now have been working on some really fine scale things like the connectome of a mouse brain and some very large scale things like human language and how that's represented in the brain. It's a pretty wide range of things, but I would say that the bread and butter, like the core out of which these other spokes emerge is indeed this, these normative models, because I really like principles.

Paul Middlebrooks
Let's stick with the normative models for a minute-

Xaq Pitkow
Yes, absolutely.

Paul Middlebrooks
-because you were just describing, you got turned on by physics, and yet these normative principles are built on these things that have happened through evolution, the structures that we can't control. You have these biologically messy things that then somehow you view them, you view the algorithms and the computational processes that they're enacting built on top of them or within them or through them as like normative toward a goal. The normative stuff is built on non-normative stuff in other words, if that's a fair characterization.

Xaq Pitkow
Yes, that's right. The machinery is non-normative and it's pushed in those directions, like the directions of optimality, but whether it actually gets there is a separate question.

Paul Middlebrooks
You would say no, right? Almost--

Xaq Pitkow
Yes, it never really gets there. In some cases it does, there’s the most beautiful examples, but there are plenty of cases where it's not going to be exactly optimal. Then you can say, well, can we understand this anyway as a principle? One of the ways that we've tried to formulate that is through something we call inverse rational control, where we say that the animal is not optimal globally, and it's certainly not optimal for the experiments that they're being put into, but they might be acting in a way that's self-consistent and doing the best they can under their assumptions. Then you can define a set of assumptions, what it thinks it's trying to accomplish, what its goals are, and then say, well, it's optimal within that. You might be mistaken, right?

Paul Middlebrooks
Who might be mistaken?

Xaq Pitkow
The animal. Well, both. I mean the researcher might be mistaken about what's important for the animal and the animal might be mistaken about what's important for the researcher in terms of like the experimental design like, “Oh, this happens this often, these things are independent or these things are correlated,” so whatever

the animal is thinking about the structure of its little world that you've put it in, those assumptions may be wrong.

We can make the hypothesis that the animal is still under those wrong assumptions trying to behave as well as it can, but it's not going to look optimal from the outside point of view because it's doing things that don't make sense according to the task. You have to find the way in which they do make sense, which we call rationalizing, just the same way as like, why did you brush your teeth before you eat your breakfast? Well, now you have to come up with some rationalization of why you do that instead of the other way around. Maybe that's a principle that relaxes the idea of optimality, but doesn't lose-- It doesn't go into, oh, anything goes.

Paul Middlebrooks
It doesn't relax it. In fact, it points directly toward it. It's just not the optimality that the task demands, right?

Xaq Pitkow
Yes, that's right. In some cases it might be the ecology that it is optimal under a different world. If you were actually in the Savannah running around and gathering fruit, this is the right thing to do. Then it would be optimal in a different task, a different environment. It may also be not optimal in its natural environment. There could be some bad assumptions there too. Then it would be optimal in some fictional environment. It's a lot of these things in the normative models, like you're talking about probabilistic reasoning, you're talking about reinforcement learning. A lot of them boil down to constraints, not just optimality.

If you could do absolutely anything, you always have some constraints. Then do you fold the constraints into the principle or the constraints, some side thing? In fact, mathematically there, you can write them as equivalent. I think it helps conceptually to separate them and say, here's a constraint that we have and we're going to work within that constraint and then we'll call the rest of it optimal.

Paul Middlebrooks
I'm trying to understand. In some sense, the history of neuroscience is task based, a large history, right? You design a task, you provide a lot of constraints, you reduce the preparation for the organism, whether it's head fixing or here you have two boxes and you can look under both and with different reward distributions, et cetera, like in the inverse rational control.

What you're saying is, okay, that's fine. Okay, so there's a lot of criticism on this task-based thing because, well, this is not ecological. This is not what organisms were designed to do, to look at these boxes or whatever and see if there are rewards under them. What you're saying is that's okay because they're still optimizing, but they're optimizing for a different thing that evolutionarily they are more prone to do or evolutionarily selected for. We can infer what they're actually trying to do, which is in some sense maybe suboptimal or less related to what we want them to do, but we can still study it.

Xaq Pitkow
Yes, that's right. This whole task-based thing is really critical for neuroscience because we want to control things, but it depends on how-- like the complexity of the task is a knob that we can turn. We've changed over time from the very simplest tasks in the beginning, even without tasks, like the animal is maybe unconscious and you just have the eyes open and you're looking at what the brain does when the animal is out and it still does stuff, starting to move towards simple tasks that give to choose A or B.

As we've gained more data, we've been able to dial up the complexity. Now we're not yet at the point I think where we can do benchmark tasks in machine learning styles where you might have a robot that's going around and loading the dishwasher or swinging from-- I don't know if there's any robots that swing from vines yet, but--

Paul Middlebrooks
Probably, but that'd be cool. I want to see them.

Xaq Pitkow
Someday soon. Certainly running around on rough terrain. Maybe trying to acquire certain goals, or for a real natural case, we would have an animal that would be just in its natural environment, like climbing trees, interacting socially with other animals of the same species, finding food, mating, running away from predators, like playing, all of those natural things are things that we don't have enough data for yet to make that the task of interest.

I think we're always looking for this intermediate level of task, which is complex enough to reveal useful and interesting structure about brain computations, but it's simple enough that we can actually characterize it. We've been shifting this way. I think I like to push a little more in the complex direction than most. Then you need to have a more complex analysis system or framework to interpret that data. Then, the goal is eventually to move really towards naturalism. It's an interesting tension. How do you navigate that?

Some of my collaborators are really trying to collect massive data. Andreas Tolias is building this Enigma Project where he's collecting massive data in freely moving monkeys, like this is the goal, and having them really do all these complicated interactions. If you have massive data, then you can build some massive models like we've seen with large language models, but now of other sources. Sometimes they call them foundation models or frontier models.

Then with those big models, now you have a description, it's a descriptive model. It doesn't say what you should do. It doesn't say how it's done. It just is like, this is what happens. Then you can try to analyze that further. We've played a lot of games with those kind of models as well, trying to see if we had the-- It's a reformulation of the data in a sense. You have this massive dataset that you compress into a descriptive model, but it's still the data. It's just reformatted in a much more sophisticated and potentially interrogatable way.

Paul Middlebrooks
Are you making assumptions about what is being optimized to compress it into the descriptive model?

Xaq Pitkow
Usually, no. Implicitly in some sense, but those are weak assumptions compared to the ones that we were just talking about with normative models. These are much more data-driven models. They're just big neural networks that describe input-output relationships. Then you can hope that they inside under the hood somehow reflect latent variables that are relevant and interesting, but it might not, because there's all sorts of equivalent ways of computing the same thing.

This is a really important, and I think underappreciated element of what we have to do is we're not trying, in my view, we're not trying to come up with the one mechanistic explanation of something. We're trying to find a class of equivalent explanations that have some shared properties, and then understand what those properties are and how they relate computationally to the behavior of the animal and its sensory inputs. When we do this in a big neural network model, you could say, hey, that model doesn't have anything to do with the brain. The neurons in this model are not the same as the neurons in the brain."

Paul Middlebrooks
And yet.

Xaq Pitkow
And yet, exactly, and yet we can still find some shared structure there. That's the challenge. That's the joy. How do we identify, how do we discover things with these new techniques? This is getting to what I would classify as the field of neuro-AI, which is a synthesis of brains and machines where you're trying to use modern AI tools to understand the brain, as we're trying to use AI tools to understand everything these days because AI is so powerful. Neuro-AI has a particularly interesting spin on this because AI came originally from neuro.

We're continually trying to give back new ideas to AI to say, hey, it wasn't just like convolutional networks that inspired AI, but here's some other detailed structures that we could use. Sometimes you can find interesting parallels that way. There's the famous quote by Feynman that a lot of people know, which is, what I cannot build, what I cannot create, I cannot understand. This is a testbed for our understanding of brains. If you really think you are understanding something about the brain, oh, yes, make something intelligent.

Paul Middlebrooks
I just had the thought, so you just mentioned that AI came from very rudimentary neuro. Us neuroscientists are constantly banging on the door, hey, listen to us, you need to do this. They just forge ahead successfully. Then we use their models. It struck me like a different way to go. You're saying that you look at the models, you look at the innards, the inner workings, and you might find some latents. Then you can possibly relate those latent states to the way that organisms are enacting their cognition. You also mentioned that there are essentially an infinite number of ways to solve a single problem.

I wonder if a useful exercise is to solve the problems in ways radically different than neural networks do, although maybe that is the history of neuroscience, which has, I don't want to say failed, but failed to solve the brain. Maybe like these psych math, mathematical psychological models that are fairly simple, accumulator models, drift diffusion models, things like that. Maybe those are sort of what I just posited. I'm just wondering like how far away from brain-like activity can you go to solve to be within that class of that you mentioned, that class of solutions for a given optimization problem? I don't know how one would move forward with that.

Xaq Pitkow
It's a good question. It's a whole family of questions. Exploring the family of, let's say, equivalent input-output relationships is an interesting one. There are two ways that something can be equivalent that I think are critical to differentiate. One is that they are equivalent over everything that we've tested so far. The other is equivalent in all ways, even ways that we haven't yet tested.

Paul Middlebrooks
Well, there's a third category, which is the super narrow equivalent only for this very particular benchmark that we're testing, right?

Xaq Pitkow
Right. Yes. Okay, good. Now we have a spectrum, and the spectrum of how widely are we pushing this system out of the training regime? If we're equivalent only in this narrow one task case, then we might be brittle. We'd have a brittle model and we'd say, “Oh, how well did we do at this, capturing things?” Then somebody comes along and says, “Oh, well, you use gratings, test it with, I don't know, random noise or natural images,” and [makes sound] it breaks. Then you say, “Well, yes, you definitely have the wrong model.”

Paul Middlebrooks
[chuckles] For what? The wrong model for what?

Xaq Pitkow
For the brain. You're trying to fit the brain or you're trying to solve the-- I mean it could be also for machine learning. Your system, it does fine at this one weird task, but it doesn't do fine in general and it doesn't do fine when you deploy it in realistic conditions. We're always looking for those, the test conditions that you really care about. That's a moving target. In fact, in the beginning people were trying to classify binary digits.

Even linear models classify binary digits with, I think it's 88% accuracy or something in MNIST, but pushing it towards higher and higher performance until that benchmark no longer seems like the right test because, okay, we can do that reasonably well. Let's try something that's a better test. Then we move towards--

Paul Middlebrooks
This is Goodhart's law? Is that right?

Xaq Pitkow
I don’t know this one.

Paul Middlebrooks
Goodhart's law states that once a metric becomes a target, it ceases to be a good metric.

Xaq Pitkow
Ah, yes. That's not what I'm talking about, but that's a good one. That's a really important one. Russell Kudinov actually had maybe a corollary of that where he said, whoever makes the benchmark wins.

Paul Middlebrooks
Right. [chuckles] Yes. Okay. That's a good one too. AI wins.

Xaq Pitkow
Yes. AI wins, but we're trying to move towards better benchmarks. The benchmarks are constantly evolving. In fact, major advances were made by developing benchmarks. When Fei-Fei Li developed ImageNet, that was a huge spur for the field. I think a lot of these large scale data collection efforts, like we have it at some big labs, those are going to push the field forward, pushing neuroscience forward, because then you have targets that people can really test things on.

That has not been the tradition in neuroscience for a long time, and now it's becoming so. This is, I think, a major sociological distinction between machine learning and neuroscience. That's really very fruitful when you import the machine learning benchmarking style thing into other fields like neuroscience, for example.

Paul Middlebrooks
I don't know if this is a good time to pivot and ask you about another common theme in your work, which is that, and again, correct me if I'm wrong, that organisms spend more energy or effort actually on meta-cognitive things like discovering. There's a task at hand, but now they have to figure out the constraints and what those constraints and what the probabilities of those constraints are and how much energy they have to spend. There's all these factors that go into solving a given problem. Would you say that organisms have to spend more cognitive effort on arranging and figuring out those limitations and constraints than actually the algorithm to solve the problem?

Xaq Pitkow
That's a good question. I don't know the answer. In terms of sheer brainpower, my suspicion is that we spend a huge amount of our energy at least doing primary sensory processing. The cognitive stuff is, it's lower, it's a lot lower bandwidth. The structures that we're reasoning about are much slower. They're much lower dimensional, but they are-- Right, so you look at an image, you get, I don't know, a hundred million pixels per eye, basically. Whereas the cognitive variables that we have are certainly the output that we have is less than a thousand dimensions. It's just every muscle that we have, basically.

Paul Middlebrooks
You just said a static picture, because in reality, we're looking at static pictures every time we move our eyes as we move through the world. It's that, what, the hundred million every few milliseconds.

Xaq Pitkow
Yes. Although a lot of those a hundred million pixels are the same from moment to moment. You have to be careful about how you're characterizing the information content of these things.

Paul Middlebrooks
We have a fovea and extra foveal. We have a fovea where we're actually only paying attention to a very small, paying most attention or getting the highest fidelity of sensory input, at least in vision of a very small area of an image. Of course, we're talking about vision because that's the history of AI and neuroscience is like almost all vision.

Xaq Pitkow
Yes, that's a big one. It's actually been fun for me that I've gotten to work on a few different systems over the years. Vision, some audition, some proprioception. That's one of the joys about being a theorist is that experimentalists have to invest a huge amount in a particular system with this equipment and everything, but it's just math, I mean the same math applies.

Paul Middlebrooks
You need a computer.

Xaq Pitkow
Yes. Give me a paper and pencil and we're off to the races.

Paul Middlebrooks
I've switched, in my career, so I used to be an experimentalist, neurophysiologist, and I was always super jealous of theorists. We're going to take a little theory sidetrack here. Now, these days I do like, it's all like computational analysis that I do. We're gearing up to do some more experiments in lab with mice. I'm hesitant because I'm like, "Oh, I just want to do the computational stuff.“

I was right to be jealous, I think, of theorists. If I ask theorists and I'll ask you this too, they say, "Well, it takes us a while too," but you don't run in the same kinds of problems. You run into computational problems. You don't run into like hardware problems and organism problems. It's a huge mess. Good for you that you went the theory route.

Do you agree with me that in some sense you not have it easier because you have to think just as hard or maybe harder, but in terms of productivity and you can partner with any experimentalist that's willing to partner with you, and maybe you have to convince them to send you their data. That used to be a bigger deal than it is now. How would you characterize being a theorist knowing experimentalists that you know, are you over there laughing in your office and, ah, I can just do sit on my computer and I don't face the same challenges?

Xaq Pitkow
I'm loving life over here. It's definitely a lot of fun to do this job, but I grappled with the same thing that you were grappling with as a graduate student. I started off-- I did some of everything. I did some experiments, neurophysiology experiments, I did some psychophysics, and I just got tired of when things broke, not, it being some weird, like a wire was loose or the solution was old or things that seemed really out of control, and to be so meticulous that everything was pristine, it didn't suit me.

I found myself, if I looked back and say, where am I spending my time? I was just spending my time in front of the computer or doing some analysis rather than being in the dark room doing some vision experiments, which were interesting. I'm really glad that I did it. I think a lot of the experimentalists that I work with are also glad that I did it because it gives you an appreciation for the difficulty and the messiness of data.

Theorists who come from, let's say, some disciplines where things are more pure, you assume that you can-- physics is a little bit of hacking. Physics is to math as hackers are to computers. It's a complicated dance because like, okay, there's two ways of where that interaction can go. One is an experimentalist comes to you with some data and says, hey, what's this mean?

Paul Middlebrooks
That's rare, right? Isn't that the rare?

Xaq Pitkow
No, that's more common.

Paul Middlebrooks
These days.

Xaq Pitkow
Yes. Assuming that they're coming to you, that you're like in a conversation, because they have the things that they've been working on and they've been thinking about. The other way is that you have a theory you want to test, and then you have to convince an experimentalist, here, would you please dedicate six months or a year of your life to testing this particular wacky idea that I had?

Then when you actually go and work with an experimentalist who wants to, like you can make it a team, and then you can say, let's co-design this experiment. We have these ideas. I know you have constraints, experimental constraints. These things are easy. These things are hard. Let's figure out if we can find the right combination of things. That's really fun. That's really fruitful. It also is a little bit amusing to me that a lot of PIs who are experimentalists are really operating very much as theorists in a sense because they're not doing the experiments. They're designing the experiments. In that sense, when I'm collaborating with experimentalists, I'm working like an experimental PI.

Paul Middlebrooks
Wait, why is this amusing? This is the way the history of neuroscience is, someone is experimentalist who had their ideas that they wanted to test. You set up an experiment and that idea is somewhat theoretical, right? It's not just like, “Hey, let's see what happens.”

Xaq Pitkow
The big boss is sitting in the office and all the graduate students are the hands. That's what I mean. Some PIs, actually, there's a few of them who really still like to go and do the experiments to get a lot of satisfaction out of being there. Of course, you get a lot out of it. You can see things, how the systems evolve. Everybody has time constraints. Those time constraints are often pretty strict.

Paul Middlebrooks
Getting back, you said you often, when you're collaborating with experimentalist labs, the PI maybe is not doing the experiments that the lab personnel are doing the experiments. Then you end up feeling more like an experimentalist. Is that what you were saying?

Xaq Pitkow
Operating like the PI in an experimental group. We're co-designing the experiment. Usually that means that for the PI, the experimentalist needs to know what all the equipment is. They're making sure that things are in the right place, that the resources are available. I don't need to, because I'm not building that lab. I'm not building that equipment, but I like to know it. First of all, the technology that we have these days is so cool. Second, it just helps me understand the constraints of the experiment that much better.

Paul Middlebrooks
Yes, fair. All right, so we went on this big tangent about theory versus experimentalist, but one more thing before we pivot again. Maybe I just have this old conception that I haven't let go of. You have a PI, they have their own lab, they have their own grants, they have their own projects, usually the things that they want to do are overflowing with respect to what they are doing. Along comes some theorist and says, “Hey, why don't you-- Here's my idea.”

Xaq Pitkow
“Here's another one.”

Paul Middlebrooks
“Here's another one. It's beyond what you can do right now, but I need you to order this new microscope and I need you to outfit a new dark room,” things like that.

Xaq Pitkow
That doesn't work like that.

Paul Middlebrooks
No, I know. Because there could be like a certain tension there, right?

Xaq Pitkow
Yes, a fruitful collaboration like that is going to happen where you know what the person is interested in, you know what their technical capabilities are. Every once in a while, you say like, “Oh, it'd be cool if we could do this.” Maybe the experimentalist like, “Oh, yes, that would be cool. Let's do it.” Otherwise, most of the time it's like, “Yes, that would be cool. I wish I could do it,” and so then you just work within the opportunities that you have and you try to make the most of what is usually really powerful technology.

The people I've gotten to work with have incredible skills in measuring stuff and they also come-- I don't want to understate the ideas and interpretive skill that they already have. Coming, there's usually complementarity, they'll know a lot of things that I don't know and they can teach me about this and vice versa. I have a lot of math skills and I know sets of models and theories that we could bring together and it can help synthesize some ideas. They probably know some particular literature way better than I do. They know what kind of signals you might find in what part of the brain and what this animal can do and what it can't do. All of that is critical to making a good collaboration.

Paul Middlebrooks
Many, many episodes ago, I was talking with, I think it was Nathaniel Daw. He's on the theorist side as well. He related to me that he and his colleagues were battling whether they should start a wet lab, whether they should start an experimentalist lab so that they could apply their own theories to it. I was like, "No, no, no, you don't want to do that." Have you thought of that?

Xaq Pitkow
Yes, Vijay Balasubramanian tried to do that too. He did that. It was just tough to get people to test his theories. Sometimes there are some easy experiments that we could run that would be fun to do. The most common of these is just human psychophysics. Here's a game. Let's just collect some data of a human playing a game. That's easy in relative terms. If people are inclined to do that would be, that would be great.

Paul Middlebrooks
Would you say that as neuroscience matures a little bit, it's becoming more like physics, in that physics historically has been happy with, there are theorists and there are experimentalists and they can collaborate but they're separate. Whereas in neuroscience, the past has been the experimentalist is the theorist and it's one person. There had been this tension when I was a graduate student, about 700 years ago, there was this tension between theoretical labs and experimental labs. Is that dissipating where people are more comfortable?

Xaq Pitkow
No.

Paul Middlebrooks
No? It's not?

Xaq Pitkow
I think that the division is indeed becoming stronger. I think it has to do with specialization. Every one of these occupations is subdividing. Even within theorists, you have people who are specializing in, let's just say, deep learning stuff and others who are going to do dynamical models and others who are going to do, I don't know, statistical physics models. People pick their specialties, and sometimes they work together. That could be really cool where you have somebody who has a math way of thinking about things, but it only works in linear models.

Then if you want to go to a non-linear model, you need to either adopt that capability or work with somebody else who does deal with more complicated, let's say, trained models, so the AI style and the math style. They're working together now in fruitful ways. Likewise, the experimentalists are coming up with teams where, here, this person is an expert in molecular biology, and this person is an expert in neurophysiology. Then you can both manipulate the neurons at a molecular level and you can do these, you can record from them at a mesoscale level.

Then maybe you also have somebody who is a cognitive scientist who does a lot of cognitive behavioral experiments. That all their expertise comes together and you can have a much richer set of measurements that are related to each other and a much richer data set that you can connect in different ways and draw insight from. I think it's actually continuing to subdivide.

Paul Middlebrooks
In other words, continuing to get healthier, and if we consider the history of physics healthy.

Xaq Pitkow
Yes, I think so.

Paul Middlebrooks
Did you pick a specialty? You said everyone chooses a specialty, but you're all over the place.

Xaq Pitkow
I'm all over the place. Yes. I've always been a dabbler. I dabble at a lot of different things. I dabble at musical instruments. I dabble in different scientific fields. There's definitely themes that emerge. We touched on some of those themes earlier on. There's a set of tools that I've developed. I keep acquiring new ones because, it's good when things can be question driven, like, "How do you answer this question?" as opposed to, "Here's my hammer, wear some nails."

Paul Middlebrooks
Well, you love to learn also, it seems, right?

Xaq Pitkow
I love to learn. Yes, I really do. It's my favorite thing about this job is that I'm constantly learning.

Paul Middlebrooks
It's amazing, yes.

Xaq Pitkow
The people that you get to learn from are amazing, too. There's so much deep knowledge in the field and the chance to interact with all these other people who have their own ideas and creativity. That's amazing. That's a wonderful experience to have.

Paul Middlebrooks
All right. Shall we pivot to probabilities in brains? It's a hard pivot. We were just talking about the experimentalist and theorist collaborations. You've been on this generative adversarial collaboration, and I talked to Ralph Hefner about this a couple years ago.

Xaq Pitkow
Oh, cool.

Paul Middlebrooks
This is to figure out how the brain computes with probabilities.

Xaq Pitkow
Yes, that's right.

Paul Middlebrooks
There are different theories about how probabilities are represented and used in networks of neurons in the brain. They all have some evidence for them. They all have some evidence against them. It depends on how you look at the data, et cetera, et cetera. You have these, the generative adversarial collaboration is a bunch of people with, maybe opposing is a too strong a word, with different, with alternative views on how probabilities might be represented in the brain who came together.

When I was talking to Ralph with this maybe two years ago, he was really appreciative of it. He was surprised at how well it had gone, how well everyone had gotten along and how productive it had been and how much he had learned from it. Why is it difficult to know how brains compute with probabilities? Then tell me a little bit about the collaboration and what you guys have come up with.

Xaq Pitkow
Sure. Yes, this is a really rich topic that we could go on for hours about. One of the problems is what people call Bayesian just-so stories. If you want to say the brain is doing some probabilistic thing, weighing sensory evidence with uncertainty, extracting latent variables, acting appropriately, you can always construct some probability distribution as your prior, under which your data would be the right thing to do for probabilistic inference.

Paul Middlebrooks
Is this the same as what we were talking about earlier where there's a thousand different solutions to a given problem? Is it related?

Xaq Pitkow
Yes, it's related to that. There are, I guess I would call these degeneracies. This is a very specific one. There are a couple of different degeneracies. One simple one to characterize is, let's say that you're in the framework of reinforcement learning, where you're trying to maximize some utility. You don't know exactly what the real world is. Now you have two things. How important is it if the real world is in one state and you guess a different state, how bad is that error? That has some consequence.

Paul Middlebrooks
This is with a defined utility, not--

Xaq Pitkow
There are different ways of measuring those utilities. The utilities could have different consequence, different utility functions. You can also have different probabilities of those correct or incorrect responses. The state of the world. What we care about is the product of those two. The utility times the probability. You're weighing the utility by how often it happens, and then you take your expected value. If you're taking a wager and you could have a 50-50 chance of $200 or $0, that's equivalent to 100% chance of $100. You just take the average return, average utility that you're going to get there.

You can imagine that there's different ways of changing the utility function and the probability that give you the exact same balance. The product is the only thing that matters. Any ways that you get that product, you could make, divide the utility by half, double the probability, things along that line. That's one degeneracy that is not possible to distinguish in any cases because there's always going to be that.

The Bayesian just-so story is like, “Oh, we found this explanation of the data and it's Bayesian, it's optimal probabilistic inference under this prior.” Well, is that a real reasonable prior? You just made that up. Maybe it was like a very jagged, weird shaped probability that was just the thing necessary to get your data to work out right. How do you resolve that?

You resolve it by testing for generalization. You look for something new and you make a commitment to your model, your probability, the Bayesian prior probability, which says what things are likely to happen. You say, I'm committing to that. Now, my model of how the brain represents probabilities has to remain consistent. I need to still explain the data when I test a new situation that still obeys my committed model.

Paul Middlebrooks
Let me just really strip this down. Two times four is the same as four times two. You get to the same solution. If you say the prior is two, then you need to use that same prior two instead of four. You need to use the two in different domains to get the same answer, and that's your commitment.

Xaq Pitkow
Yes, exactly. Exactly. Every time. You've made a commitment to that two. That two represents the structure of the world, the things that your model assumes that are likely to happen. In all of these cases, when people have said, "Oh, the evidence is favoring this particular interpretation," "Oh, the evidence is favoring that interpretation," they may be using different generative models of what variables they're trying to do probabilistic inference over, and they could be using different probabilities.

We're not comparing apples to apples. In order to do a fruitful comparison, you need to compare apples to apples. You need to be sharing. You need to make a commitment to a model of the way that the world works. Then you can evaluate different models of the way the brain weighs probabilities. You can't do the second part, testing whether these different models are representing probabilities until you make a commitment to a generative model.

That language, that has not been, I think, widely appreciated. This adversarial collaboration, in the beginning, it was like, well, you're making models of orientation, we're making models of additive component features in the world. Yes, those are different generative models. You're going to explain things differently. Ralph actually has shown quite beautifully that you can take one of those generative models and have a sampling-based model for image patches.

Then if you look at it from the point of view of some of his adversaries who are looking at probabilistic population codes, which about the orientation of some line or strike pattern in the input, then you get their data explained. You can explain the same data, different data with the same model in these two different ways. Coming up, reconciling that is hard work. Finding out that what generative model commitments we've made is part of what everybody needs to do when they're describing their own models.

Paul Middlebrooks
Wait, you said that you use the same model to explain different datasets, but in one instance, you're using sequential sampling, which is one theory of how the brain uses probabilities and another one you're using the distributed population probabilities. Those are two different models.

Xaq Pitkow
Actually, well, it's different interpretations of the same data. You can have one underlying distribution that you look at it from this perspective when you ask what is the representation of orientation or what is the representation of the image patches? It's the same mechanism. It's just one thing that looks different in these two different ways.

Paul Middlebrooks
You have to reconcile how those two different stories can end up doing the same thing.

Xaq Pitkow
Yes, there's a third group. In this adversarial collaboration, there were really three different groups that we tried to get represented. These are prominent representations of probabilities. One is sampling, which basically means if you see something out in the world and you're trying to interpret what caused it, you roll a dice, and then some fraction of the time you'll come up with one interpretation, another fraction of the time you'll come up with a different interpretation. That's just constantly happening by our brain. Our brain is rolling dice. It's coming up with alternative interpretations, and the amount of time that you're spending with one of those interpretations is the probability it's right. That's the sampling hypothesis.

A second one is probabilistic population codes, and the third is distributed distributional codes, which are really funnily similar terms for probability representations. In some ways, they have a lot of similarities. They differ by whether you're representing probabilities or log probabilities directly, and by directly I mean linearly through the neural activity. There's some arguments about which one of these is better and which one of these is worse.

They are fundamentally complementary. I think that if you find one, you're going to find the other because they're good at different computations. In doing probabilities, you have two operations that you have to do all the time. You have to multiply and you have to add. The multiplication is when you have two independent things happening, like you roll one dice and you flip a coin. The probabilities of both things happening are the product of each one separately. That's the probability rule.

The other one is only one event happens. If you roll a dice, you got a five or you got a six. You didn't get both on one die. Those, in order to compute the probabilities of that, they have to add up to one. That's the sum rule and the product rule of probabilities, and you have to do that constantly. When you see a new piece of evidence coming in, if it's independent, then you're going to multiply your probabilities, and that's the way you're going to accumulate information.

These different codes are good at different ones of those computations, and so it's natural to jump back and forth between them. In fact, there's a lot of mathematical beauty in this, that they have this complementarity. One of them prioritizes interactions, and this is the probabilistic population code. This is a little bit of a technical thing, but it's fundamental to the representation of structure, not just probability, but structure.

Paul Middlebrooks
What does that mean? What do you mean structure?

Xaq Pitkow
There's a lot of ways of characterizing structure, and structure is really critical in the way that the brain understands the world. I think it would be maybe helpful to talk more about the different kinds of structures that there are. One that I like to work with is called probabilistic graphical models. These represent, in one version of them, they represent causal interactions. Like right now I'm sitting on a chair. The chair is sitting on the floor. The floor is held up by some walls, the walls held up by some foundation, the foundation held up by the earth. There's an indirect chain of causation. I'm held up by the earth indirectly through this long chain.

Paul Middlebrooks
Through those nodes in the probabilistic graphical model, chair, Xaq--

Xaq Pitkow
Each one of those, yes.

Paul Middlebrooks
-foundation, earth.

Xaq Pitkow
Exactly. Each one of those is a node, and the edges between them say what are the interactions. The foundation is not directly interacting with the chimney, or the, I should say, maybe not the chimney, the roof. The roof is held up indirectly by all of these other things. That structure, I hypothesize, this is a key hypothesis that I'm really very interested in testing, that that structure is known by the brain and used by the brain in its computations. It's pretty natural. It actually provides a good way of restricting the possibilities of what computations you have to do. Not everything is possible, because if everything is possible, then you have to consider all those possibilities. Here you can restrict your possibilities in a structured way.

Paul Middlebrooks
Probabilistically structured?

Xaq Pitkow
You can also do the meta-level thing where you have a probability over different graphs. In fact, there's a nice way of doing that in a meta-sense, where you have a dynamic graph interpreted as a graph with hyper-edges. It's really fun. It's beautiful. In fact, I made a 3D-printed model of the natural statistical shapes that emerge out of this. It looks like a rounded tetrahedron.

Paul Middlebrooks
Is this what you're going to show me that's in your office at work?

Xaq Pitkow
Yes. I can send you a picture of it.

Paul Middlebrooks
Yes, send me a picture, and I'll put it up in the video.

Xaq Pitkow
Yes, so this is a little three-dimensional probability distribution. The cool thing about it is that you can have, let's say, two variables, two nodes in your graph, X and Y, that are directly interacting at one time. They're disconnected another time. They're not directly interacting, depending on the value of a third variable.

Paul Middlebrooks
Right, a meta-variable kind of.

Xaq Pitkow
Yes, and it's mutual, and there's all sorts of symmetries, but it creates this little tetrahedron where, from different edges, like the front edge is saying that these are positively correlated or positively interacting. The back edge, if you're at a negative value of your third gating variable, has them the other way. If you connect the dots, you get a tetrahedron, and this rounded tetrahedron is like that. You can definitely get these changing graphs. The graph structure, if you don't have an edge between variables such that they're not directly interacting, that is a valuable restriction. That is a valuable structure that the brain could use to simplify its computations potentially.

Paul Middlebrooks
There's no causal dependency between them.

Xaq Pitkow
There's no causal dependency. Causal representation, like we know that neural networks are universal function approximators, meaning that you can take any of those inferences that you want and do it in an unstructured way, just like a big network. You just throw it at it, and you train it forever, and you'll find the right answer in the end because you can. You always can do that. It might take huge resources. It might take a long time and a lot of data. That's another resource.

Critically, it may not generalize because you're not using the right structure. If you test it in a new situation, if I now put, I don't know, like a yoga mat under my chair, it doesn't change the rest of the structure. If I were just doing a universal function approximator to describe what's interacting with what, I would now need to start over. I would need a whole new circumstance. I can't leverage all the knowledge, structured knowledge that I already have. To me, that graph structure of what is causally influencing what is going to become a really important inductive bias that I would say neuroscience has not really yet resolved. I threw in that term there. We'll come back to it.

Paul Middlebrooks
I was about to bring it up, actually, because, yes, I mean it's--

Xaq Pitkow
Let me just quick finish the thought there.

Paul Middlebrooks
No, no, we're going to stay on. Yes, go ahead.

Xaq Pitkow
That structure of having direct and indirect connections is something which is manifested in one of those probabilistic codes and not the other. That's bringing it back to this generative adversarial collaboration. There are still fundamental questions. That's just my perspective that the natural parameters, which is this basically the non-zero, like the edges on the graph of what's connected are highlighted by one of the representations and not the other.

Paul Middlebrooks
In your representation, right, where you could have the connection or not, you could turn on or off the connection, okay, so you posit that the brain or our minds learn these graphical representations of the causal structure of the world, these inductive biases, which is interesting because it's built on top of an organic neural network that somehow then learns a network that's very meta. Then the other account, does it also posit that the brain has to learn the structure?

Xaq Pitkow
Yes.

Paul Middlebrooks
It does? Okay.

Xaq Pitkow
I think this is why some of the work in distributed distributional codes actually secretly manifests this same structure. It's representing the graph anyway, so it's secretly like a local transformation of a probabilistic population code, but globally it's the same as a probabilistic population code. This is pretty technical. We'd have to go through some math for it.

The idea of a hidden graph, I think, is pretty accessible. We're trying to develop methods to discover those hidden graphs and whether information is actually flowing along those hidden graphs, not just they're representing things that could be present, but what you're guessing about the world as it is right now, what you're looking at. Are those signals flowing along some implicit graph in your mind? Can we find that graph? Can we find how the signals are transformed from representation to representation?

Here's where I think the inductive bias becomes really critical, which is that we're imagining that the brain is good at representing probabilities. One way that it could do that is by just living in a world where that's helpful. This means that every time you're faced with a new problem, the right way to solve that problem, if you practice it a lot, is to use probabilistic reasoning,

Paul Middlebrooks
Bayesian.

Xaq Pitkow
Bayesian reasoning. Every time you're doing an auditory discrimination, you're trying to run down to catch the ice cream truck, all of these different things that you might be doing, for every one of them, the best solution is going to weigh evidence by its probabilities and synthesize them together in this Bayesian way.

Paul Middlebrooks
That's super expensive.

Xaq Pitkow
The question is, is there some motif that lets us do that with less cost? Something which is reusable. That's, I think, a big question. I don't know that we have that, but if we don't, let's say that you're well trained and you end up with good Bayesian solutions for all these different problems. Does that mean that we are actually Bayesian brains, that we use Bayesian brains? That might just be an emergent property of a well-trained network that did not have an inductive bias that favored Bayesianism to begin with. It's just the result of the training. If you just took a generic neural network, and in fact, this was done by Orhan and Ma, they had an earlier version, which I like the name of better.

It was called The Inevitability of Probability.

Paul Middlebrooks
Badass.

Xaq Pitkow
Yes, they had a new title when it was eventually published. They were saying, "Hey, let's just throw a generic neural network at these things." Lo and behold, probabilistic representations emerged. What they would not have because they have no mechanism for this is parameter sharing, such that if it learned to do Bayesian inference for this task, that it would then automatically be better at Bayesian inference in a new task. That element is something that those kinds of networks do not have a propensity towards probabilistic reasoning.

It emerges from the data, not from the inbuilt, the innate capabilities. That means it does not have a good inductive bias for Bayesian reasoning. Though it emerges, I would say that's not a strong case of a Bayesian brain. That's like, okay, just good training.

Paul Middlebrooks
Because it's not normative in that sense.

Xaq Pitkow
It's not using that principle in lots of different cases. It has to relearn that principle every single time.

Paul Middlebrooks
Is that more amenable to a sampling-based approach?

Xaq Pitkow
I think the sampling is exactly the same kind of issue. You need to be able to use samples with the right probabilities. You still need to be able to take the-- Yes, it might be that sampling is an easier thing to have parameters shared. For example, I would say maybe a single neuron representation is easier to have parameter sharing, because you could genetically encode it, than a population-based thing might be. I don't know. This is, I think, a really critical question.

Paul Middlebrooks
Because the population, then it doesn't need to be emergent. It can just be hard-coded.

Xaq Pitkow
Some elements can just be hard-coded. You could have different microcircuits that are really good at doing representations of probabilities.

Paul Middlebrooks
Say a cortical column.

Xaq Pitkow
Cortical column or different brain areas. Another way that you could do it, so here we're talking about parameter sharing over space. You have this group of neurons that does something, and then you copy somehow, which is non-physical. You could copy its parameters to another group of neurons. You can't do that by learning, but you could do that by development. They're both programmed to go down the same developmental path. Then you end up with good probabilistic reasoners at different locations in your brain.

Paul Middlebrooks
Okay, so we've really gotten into the weeds about the probabilistic stuff. I want to move on because I want to talk about your specific work and your reflections on it. Just backing up, there's this generative adversarial collaboration. Everyone has different perspectives on how probabilities might be used or computed in brains to do things. The capacity of the functions of the processing in the brain is so vast. Couldn't you just be using all of them, depending on the context?

You were just saying that the two distributional codes, population codes, had this nice mathematical relationship trade-off, right? Back and forth, that would be useful in different situations?

Xaq Pitkow
Duality there, yes.

Paul Middlebrooks
Yes, inverting. Can't it just use it all?

Xaq Pitkow
Sure, it could. We're looking for universals, but we may not find them. It might be that different things happen at different locations.

Paul Middlebrooks
What's your bet? Do you think of the brain as a kludge, as like lots of different things, just working it out? You're also a normative person. You think of the brain as optimizing in a normative fashion. Where do you land on this? What whole-brain sort of perspective?

Xaq Pitkow
Yes, to be honest, I don't know yet.

Paul Middlebrooks
Is it both? It's always both.

Xaq Pitkow
It's a kludge, and it's principled. I guess that's generally the way that I go is that there's going to be elements of both there. The hope is that you can find some principles that make things understandable. In a sense, the only things that are understandable are the non-kludges. The only things that are understandable are the principles. In fact, this is a point that I like to make when trying to discover one of these graph-structured algorithms in a dataset.

If there are dynamics in the brain computations that proceed along a graph, and if that graph was just every edge did its completely separate thing, then it's not even really very meaningful to talk about that as an algorithm. An algorithm, if the brain has an algorithm, it means it's doing the same thing in different contexts.

Paul Middlebrooks
An algorithm is a defined set of steps that need to happen to accomplish something.

Xaq Pitkow
Yes. In fact, just to give some context here, the name of my lab is the Lab for the Algorithmic Brain.

Paul Middlebrooks
Also abbreviated as LAB.

Xaq Pitkow
Right. Actually, I should say it's Lab for the Algorithmic Brain, and then the first LAB stands for Lab for the Algorithmic Brain, and then the first lab there, it's like--

Paul Middlebrooks
Douglas Hofstadter would really like that.

Xaq Pitkow
He would love it. The GNU Unix people would like it too, because GNU, it stands for, GNU is Not Unix, and the GNU there is, GNU is not Unix.

Paul Middlebrooks
Recursive all the way down.

Xaq Pitkow
All the way down. Looking for those kind of structures, I feel like if we don't have a shared repeatable series of steps, there's nothing to learn. It's just a big hack. Anything that I am going to learn is going to be from some kind of principle that is shared. If every brain does something different, if every part of every brain does something different, it's going to be hard to make any sense of anything. Now, some people actually do believe that, and the place that they look for principles is not in the functioning of the brain per se during, let's say, inference or operation, but rather in the learning.

That there is some underlying learning rule, and that you have a goal or an objective, and that that's what we can understand, not the resulting emergent computations, which are just whatever happens when you learn with this dataset.

Paul Middlebrooks
Oh, it's not that that principled learning rule would result in essentially the equivalent of a shared computation. It's that it learns whatever it needs to learn and the fundamental things.

Xaq Pitkow
Yes, exactly. Some people lean in that direction.

Paul Middlebrooks
That's kludgy. That's the kludgy direction, right?

Xaq Pitkow
It's kludgy in the final result of what computations happen, what inferences happen, but it's not kludgy in how you get there.

Paul Middlebrooks
It's a fundamental learning principle. It's not kludgy.

Xaq Pitkow
Yes, that's the idea. There's also the evolutionary history that we should account for, and I don't know to what degree some of that is kludgy or what degree some of that is core principles. In fact, this brings me to one other major collaboration that we've just started, this Simons Collaboration for Ecological Neuroscience, which is basically saying, like-- Okay, so ecological neuroscience, let me just give a little background on that. Ecological psychology was a field founded by Gibson and Gibson in the '60s.

Paul Middlebrooks
Oh, nice, you included both Gibsons. That's great.

Xaq Pitkow
Both Gibsons.

Paul Middlebrooks
Most people just get the one.

Xaq Pitkow
Yes, the husband and wife team, the husband was more focused on computations and the wife was more focused on development, but they're both critical there, like childhood development, that kind of thing. They were arguing against the idea that the brain creates representations. These Gibsonian psychologists, they really don't like this idea of having-- using the word "representation" is often anathema to them.

Paul Middlebrooks
It's anti-representational, as some people say.

Xaq Pitkow
Yes, exactly. I've found it's a little softer in practice than I expected when talking to Gibsonians.

Paul Middlebrooks
Softer in that they're more--

Xaq Pitkow
They will use the word "representation," like maybe they let it escape their lips accidentally. I'm not going to name any names, but you know who you are. The idea, the contrast would be, here you're picking up a coffee cup here. You look at the scene, you have a cup, it's got some edges, it's a black object, it's got some rounded shape there, and it's got a darker patch in the middle.

Paul Middlebrooks
This thing, we're talking about this.

Xaq Pitkow
You're holding, yes, that thing right there, exactly.

Paul Middlebrooks
Before I picked it up, I figured out all of the planned movements I needed to do, I had a complete mental model of what was going to happen, right?

Xaq Pitkow
Excellent, yes, perfect. That way of putting together objects from pieces, that representation is something that the Gibsons didn't like. They

thought that that was a mistake. Instead, they think that we have-- there's sort of two aspects. One is direct perception, which I'm not a big fan of, but the other is that we interpret the world in terms of things we can do with the world.

Paul Middlebrooks
Affordances.

Xaq Pitkow
Affordances. It's a word that J.J. Gibson came up with. The word "affords" was already there, but not as a noun. He said, "Your coffee cup affords picking up by the handle." He says that that is an affordance. You can grasp the cup. You can fill the cup. You can throw the cup. You can bop the top of the cup and make a boop sound.

Paul Middlebrooks
I like that one. I'm somewhat surprised that I'm hearing you say this because my perspective is that you're more on the representational side because you're learning graphical models, and there's a lot of structure and world models, et cetera. Is this something, is ecological psychology? I guess you'll describe what ecological neuroscience is. It's something that you are coming to appreciate or have appreciated?

Xaq Pitkow
I love the idea of affordances. I think it's a powerful idea, and it helps us structure things in the world in a way that is focusing on the stuff that's useful as opposed to-- This team, which we call SCENE, Simons Collaboration for Ecological Neuroscience, we're really trying to test these ideas through neuroscience and sophisticated behavioral experiments with a whole variety of animals. We have mice and monkeys and humans and babies and bats, baby humans, and looking for different tasks where they have some information that they can't do anything about, some signals they can't do anything about.

They're not affordances. They don't afford anything. They have some things that are useful. You can do stuff, you can act upon them. They're controllable. Then you have some things that are rewarding. Some of the things that you can do are not necessarily experimentally rewarded. They're not part of the task. When you think about the main themes of the big theories of neuroscience and machine learning, they're in these two categories. One is reinforcement learning. You do everything that you can to get your goal, and you learn stuff insofar as it supports your goal.

That's one extreme. Then, on the other extreme, you make a generative model of everything. You try to describe the causes of all of your sensory observations as a compression of your world, even if it's not useful, even if it's not rewarding. We think in between and a little off to the side is this other possibility that you don't learn everything, and you don't learn just the rewarding stuff. You learn stuff that you can do. That's the affordances. That becomes a compelling idea because I think it allows you to use your limited data and your limited resources more efficiently for things that will generalize better.

If you just focus on what's useful right now, the world is constantly changing, and you're going to miss a bunch of things that could have been useful later that you didn't know about. If you try to learn everything, it's a little bit like Sherlock Holmes saying, If the sun goes around the moon, the sun goes around the earth, or the earth goes around the sun, it makes no difference to me. I'm going to promptly forget it. Sherlock Holmes is very practical about that because it's explaining the same data.

Now the generative model would have, you'd be trying to describe whether the sun goes around the earth or the other way around, not because it's useful for you, but just because you're trying to explain everything you can. You're using a lot of brainpower in that model to do things that you might never use. This affordances idea, I think, gives us an interesting potential balance between generalization to new things that you might be able to act upon that could be rewarding later.

Paul Middlebrooks
Total generalization versus--

Xaq Pitkow
Yes. The generative model, which models every cause in the world, that's going to be the best generalization, but you pay for it. A lot of data that you have to accumulate.

Paul Middlebrooks
It's less intelligent, too, because you're learning a lot of things that you are not going to need or use. In some sense, it's less intelligent if intelligence is solving problems at hand.

Xaq Pitkow
Correct. Some ensemble of ecological problems that you actually encounter in nature. This becomes, I think, a useful third theme or third thread that we could explore. We have some neuroscience experiments to test them, which are a lot of fun. Now, the other element to ecological psychology was direct perception. This basically says that we don't have steps of computation. We just directly know things that are there.

Paul Middlebrooks
This has rubbed people the wrong way historically, not just you. This is the main thing that people are like, "What the hell are you--"

Xaq Pitkow
Yes, exactly. One good example of this, which gets back to the kludges, is like Paul Cisek is an ecological neuroscientist, I guess, who is very much in this anti-representation camp. He's written a couple of beautiful papers on the evolutionary history of brains. I love this couple of papers that he's written that are gorgeous and wonderful synthesis. I highly recommend them.

Paul Middlebrooks
He's spending a lot of time on that. My conversations with him, I said, because it's sort of off the beaten path, and it's a passion project for him. It's beautiful work, and I'm glad that he's doing it.

Xaq Pitkow
Yes, I'm grateful as well. He has an example of how you might find shelter as a lizard. You're moving around in the world, and you see a patch of light that is bright on top and dark below. As you move in one direction, the dark part gets bigger and the bright part gets higher. That suggests an overhang. You're getting closer to an overhang, which might be shelter. Now you can imagine direct perception where you just wire this particular visual pattern of a bright patch going up and a dark patch growing.

You just wire that directly to move forward under some context, at least. Then there's loops of loops that are feedback and regulating and all that. That basic movement would be a direct perception approach. You don't have to make a model of the fact that there's a concave area there. You just do this. You connect this sensory input to that motor output, and you're done.

Paul Middlebrooks
That's probably why it's hard to swat a fly.

Xaq Pitkow
Because they're really good at those kinds of things. They have very quick, close reactions. Whether that describes-- Gibson thinks that describes everything. It seems just false on its face that that happens in the brain.

Paul Middlebrooks
I agree.

Xaq Pitkow
It's just false. There was a whole article of a *Brain and Behavior* journal that I actually took out of the library [chuckles] last year.

Paul Middlebrooks
Wow.

Xaq Pitkow
I know, I was holding this book. It's a real book here, not online.

Paul Middlebrooks
Did you get calluses? Did you develop calluses from that?

Xaq Pitkow
Oh my God. No, they got really sore because turning the pages was a muscle I hadn't used in a while. There were a bunch of responses to-- I think it was Shimon Ullman who was critiquing direct perception in this way, and a lot of luminaries giving their responses to it. One of them was, I thought, pretty interesting from Jeff Hinton, which was like, Gibson couldn't possibly have meant that, that it's just there are no intermediate steps. I think instead what he was probably arguing against was more the good old-fashioned AI sense of, like, first you extract edges.

Then you put them together into contours, and then you do this thing, and then you do that thing in a very computer science-y way.

Paul Middlebrooks
Had Hinton read the Gibson-- I have not read the main text of Gibson. Yes, I wouldn't know.

Xaq Pitkow
I don't know what Hinton had read, but he was trying to reconcile these things. I think his connectionism was along those lines. He was saying, "Yes, it's just an emergent behavior of all these neurons working together." That could be perfectly consistent with this idea of non-step-by-step algorithms that Gibson may have had in mind. Now, from my perspective, it might be possible to take one of those connectionist architectures and actually interpret it as, "Hey, look, here's some sequence information about contours."

The information flow actually prioritizes along the contours, and some other information is off the contours. It's not like here's a contour neuron, and here it's-- There's some balance here. Saying that these neurons don't interact with those neurons in this context, that would be structure to the computation that I suspect is still going to be there.

Paul Middlebrooks
You would still find single neurons. You mentioned Horace Barlow earlier, right? The-- What is it? The neuron hypoth-- No, the single neuron doctrine, which he was partly responsible for. You would still find neurons that correlate with the contour, right? You wouldn't just think of it as a contour neuron. [chuckles]

Xaq Pitkow
Correct.

Paul Middlebrooks
You could decode like that it probably has to do with the contour from that single neuron. As opposed to the historical single neuron doctrine of Barlow et al., you wouldn't call that a grandmother's cell, a contour cell, for example.

Xaq Pitkow
Neurons will have more and less specificity. It'd be perfectly fine calling that a contour neuron. The real question is, what is the range of generalization that it has? If it generalizes that over a wide variety of contexts, backgrounds, contrasts, then, yes, you might say that that neuron has localized some information about contours. You can always find such information, like whatever neuron-- a neuron response selectively to whatever turns it on.

Paul Middlebrooks
I think by calling it a contour neuron, you hearken back to like, if you kill it, you don't see contours anymore. That's like-

Xaq Pitkow
No.

Paul Middlebrooks
-your mental representation of a contour is due to that neuron. That's the sort of implication that people rail against.

Xaq Pitkow
Yes, absolutely. I think even people who are thinking about the single neuron doctrine weren't necessarily saying that it was only that neuron.

Paul Middlebrooks
I agree.

Xaq Pitkow
I think that the evidence shows that these pieces of information are more widely distributed. I think that finding the right basis, the right pattern-- We're looking for patterns and how the patterns relate to each other. This brings up, I think, a fundamental point that people who think about representations often neglect, which is that representations are useless by themselves. They need to be connected. You need to think about the brain in terms of representations and transformations.

Paul Middlebrooks
I just had a panel on to talk about representations, and it got siderailed because Jon Krakauer related everything to mental representations. In this case, you mean the structure of the neural activity? Is that what you mean?

Xaq Pitkow
I mean the relationship between the neural activity and the external world. I listened to that episode that you're talking about, and I talked to Jon Krakauer.

Paul Middlebrooks
It's just a contentious term.

Xaq Pitkow
It's a contentious term. I actually have another generative adversarial collaboration that I've been working with.

Paul Middlebrooks
Is this what makes a representation useful?

Xaq Pitkow
Yes, exactly. Different people use the word "representation differently." It's fine. Just say what you mean. Then we can go on and say, "Does that thing happen? Does this thing happen?"

Paul Middlebrooks
In your case, it's some relation between the neural activity and something happening in the world.

Xaq Pitkow
Yes, I'm going to use the word "representation" right now in this joint sense of having information. It's the simple, typically neuroscience style. It's actually the style I started with, the meaning I started with, before I was convinced otherwise. If you take the view that something is a representation, if it has information about the world, then it is--

Paul Middlebrooks
Shannon information or meaning information.

Xaq Pitkow
Any information.

Paul Middlebrooks
About this intentionality information or something?

Xaq Pitkow
I think it doesn't-

Paul Middlebrooks
Really matter. Okay. [chuckles] It's another contentious term.

Xaq Pitkow
Let's just use the term "Shannon information" for now.

Paul Middlebrooks
Sure.

Xaq Pitkow
Then it is not possible to have a misrepresentation.

Paul Middlebrooks
Oh, okay.

Xaq Pitkow
This is something that other people have made as a point before, and I didn't appreciate it, but some philosophers actually pointed it out to me, and I was like, "Oh, okay, so that makes sense." Now you need to use that information in a way that is consistent with some rules of transformation in the world for it to be a correct representation. If I see an image on my left and I turn to my left, then I am correctly representing. If I want to be directed towards it, then I have a representation that is used appropriately.

If I see something on my left and I actually move to the right, then that's a misrepresentation. If I'm wearing prism glasses or something. Representation by itself, it's just sitting there. It needs to have some function. It needs to do something. We always want to be thinking about information that's there in the neurons about the world, but also how it gets transformed either to behavior or to other neurons. That joint representation and transformation is what we should be studying, because you can have the same representation transformed in different ways, which would mean that it might be a misrepresentation instead.

You can have one transformation applied to two different representations, and one of them would be useful and one of them would not be. For example, let's say the transformation is just adding things together. That is the correct thing to do for probabilistic inference, bringing back that other idea. If you are representing log probability and then you multiply probabilities by adding log probabilities. That's a match between the representation and the transformation that is helpful for that particular thing.

If you were trying to add together probabilities directly in order to synthesize evidence, that would be a mistake. You would not get the right answer that way. You need to have this alignment between how things are, like the patterns of neurons and how they relate to the world, and how you change those patterns over time or over space to get to new formats of things that matter for the world.

Paul Middlebrooks
How did we get here from ecological neuros-- Did we describe because--

Xaq Pitkow
Yes. Anti-representation, right? I think you asked me like, "Hey, how could you like--"

Paul Middlebrooks
We're talking about the middle ground where there is representational structure, but it's not completely general, and it's not completely brittle. There's direct perception.

Xaq Pitkow
Yes, it's against direct perception, but it's something like I think that there are some cases where you find the right combination of features, and it does the right thing. You can say, "Oh, that's direct perception." Or you could say that "Oh, this is a useful computation for this particular task." Now it becomes a matter of semantics. It depends on what direct means. I don't think it's all that important to go into tremendous detail of what Gibson might have meant by direct perception, but I think we can find that there are spatio-temporal patterns in neurons that relate to spatio-temporal patterns in the world.

That the way that those patterns evolve over time and space are the computations that we do. This is computation by dynamics. This is the thing that couples the patterns of input, which we can call representations if we want, to the way that we use them and behave upon them, which is the thing that lets us know whether in Krakauer's sense, we have a mental representation of predictions about the future states of the world are reacting in a way that's consistent with where we want to go.

Paul Middlebrooks
Mesh this with the idea, your work that we were talking about way earlier on inverse rational control, where an animal is not necessarily optimizing for the task at hand. It is optimizing for something that's suboptimal relative to the task, but it's rational in the sense that it's optimizing for something that it has evolved to to optimize for.

Xaq Pitkow
You could say that if they're behaving rationally in this false world, then they're misrepresenting their sensory evidence. They're getting one form of sensory evidence.

Paul Middlebrooks
That would be an error. That's what I wanted to bring it back to. The error in representation, you would consider that an error.

Xaq Pitkow
I would consider that an error in the sense of related to the external world, but it's self-consistent.

Paul Middlebrooks
Rational.

Xaq Pitkow
Yes, it's rational. That's why I like to distinguish rational from optimal. Other people use the word rational differently.

Paul Middlebrooks
That's what I was going to say. Some people confound them, right? Or they could be confounded in some people's definitions.

Xaq Pitkow
Yes. Some people talk about bounded rationality, which are things where you might have some superstitions or some other approximations that you make. In the sampling context, Ed Vul has this paper, One and Done, where basically, there are some cases where behavior is consistent with taking a single sample of a probability distribution and just acting upon that. That's rational, but under the bound that you have extreme time constraints or some other kind of resource constraints.

I think in a lot of these-- We have a shared understanding of what some words mean. When we find some conflict, now we have to go through and say, "All right, do we mean the same thing by these words? Is it a real conflict?"

Paul Middlebrooks
I know, but we don't want to be philosophers. We don't want it all to be about that. We want to move forward.

Xaq Pitkow
Yes, exactly. I think we do move forward. We assume that we know what each other is talking about until we don't. Then we try to reconcile the way we're using terms. Then maybe we don't. Somebody says, "I don't want you to use that term that way." I say, "Okay, fine. We can agree or not. Let's go on to the substance."

Paul Middlebrooks
Right. Yes, exactly. Okay, Xaq, we've spent a lot of time already, and we haven't even talked really, except for the inverse rational control about some of your projects. Just to list some off, some of the recent work that you've done is you've studied how attention fluctuates, like the pattern of high and low attention based on the constraints of the task and the probabilities, et cetera. How we control what we do. What is the phrase? Move. I have it in my notes here.

Moving more to think less. You have work on the recurrent graphical probabilistic models. We can't go through all these, but is there something you want to highlight? Is there something you're most proud of or most joyous of? Because you are a person of many pursuits. I want to leave it up to you to sort of highlight what you think is most fun or interesting.

Xaq Pitkow
It's like asking you to pick what's your favorite child.

Paul Middlebrooks
I know. What's your favorite color? Yes, your child.

Xaq Pitkow
Oh, I can say my favorite color is green, for sure.

Paul Middlebrooks
Okay, I can say favorite child easily, too, but I won't reveal it. No, I'm just kidding. I can't say it.

Xaq Pitkow
Yes, which is my favorite project lately? I'm really happy doing this ecological neuroscience stuff.

Paul Middlebrooks
I was super surprised that you-- I didn't know about that, and I was super surprised that you were going down that way.

Xaq Pitkow
Yes, we have zero publications on it. It just started in July. It was a long, slow process of putting this together. The team is big. It's 20 people in this team, but there are six theory teams within this group. It's very theory-led.

Paul Middlebrooks
How do I join?

Xaq Pitkow
[laughs]

Paul Middlebrooks
All right, too late.

Xaq Pitkow
Stay tuned. I think there will be opportunities for broadening these--

Paul Middlebrooks
Fair enough.

Xaq Pitkow
It's going to be a 10-year project.

Paul Middlebrooks
Holy cow.

Xaq Pitkow
Yes. This will take us a while. That one's a lot of fun. We've already talked about it. I'm really excited about this dynamic graph thing that we talked about, where-- I'll send you that shape, the little tetrahedron.

Paul Middlebrooks
Cool.

Xaq Pitkow
In fact, if you make a graphical model of this where you have one circle for each variable and then a square for how those variables are interacting, you actually end up with a picture that looks like the flux capacitor from *Back to the Future*. It's like these three things coming into the middle. I am excited about using that motif to explain a whole bunch of things. I call it the statistical transistor.

Paul Middlebrooks
Nice.

Xaq Pitkow
Because once you have that third variable gating whether the other two are interacting, the expressive power of that structural graph becomes vastly larger. It explains some interesting properties just in low-level visual perception. I think it can explain a whole lot of structures in the way

that we're changing cognitive patterns. Interestingly, the fundamental math equation there that you use for these three interacting variables is x times y times z. Remarkably, two things show up out of this.

First, that is what you get in transformers. Transformers, the very popular machine learning architecture that underlies large language models at the moment. This multiplicative gating has ancient histories. If you go back to the '60s, that's ancient. In computer science sense, people were using sigma pi networks, where this product is-- sigma is the sum and pi is the product. Multiplying these x times y times z actually gives you a lot of interesting expressive power. Transformers use that in a slightly specific way.

Fundamentally, the unit is that you're multiplying these different elements together and using that as what they call attention, which has a lot of similarities with biological attention in terms of gating things. The other thing that I really like about this that I find exciting is that that motif emerges naturally when you give neurons more capabilities than traditional neural networks. Traditional neural networks, artificial neural networks, you take one neuron, you take all of its inputs, you take a weighted sum of all those inputs, and then you pass it through a nonlinear function. That's your neuron's response. This has been stupendously valuable.

Paul Middlebrooks
It gets the same one every time it passes through, as long as the weights are the same. It gets the same input.

Xaq Pitkow
Yes. Every other neuron is going to be the same nonlinear function that just takes a different weighted sum of maybe a different set of inputs. The weighted sum is what we call a projection of the input. If you allow it to learn a nonlinear function, mapping inputs to outputs, but you learn it based on not just one weighted sum, but two weighted sums, now automatically emerging generically is a product, X times Y. That very operation that shows up as the critical new ingredient in transformers, this product, and was something that people dabbled with in the 1960s, it emerges naturally when you give neurons the power that they automatically have in biology.

In biology, you have neurons that don't just-- like the dendrites don't just come in and give one input to the neuron. You have an apical dendrite. You have a basal dendrite. There's more structure there, but at least you have that. If you give neurons that very simple biological structure, all these things emerge. You get attention. You get this gating. You get these statistical transistors just popping out for free generically. This, to me, is connecting the low-level microcircuit, like a small neuron-scale thing that you could share across lots of neurons easily by just giving them apical and basal dendrites, and all this stuff, all these good computational properties emerge out of that.

That, for me, is really beautiful because it connects something very simple and low-level to something very powerful, abstract, and computational.

Paul Middlebrooks
I'll tie this in. Something you said at the NeuroAI conference in Washington, D.C., a few months back surprised me. Now you're saying this, it doesn't surprise me as much in terms of what's missing. What do we need? What do we need funding for? What do we need to accomplish? You said we need a synaptome, essentially. We need to learn all the connection strengths of synapses, which is such a low-level detail. There's resistance to-- we don't want to go all the way down to electrons. We don't want to go down to ion channels, but you want to go down to--

Xaq Pitkow
How about quarks?

Paul Middlebrooks
Quarks. I was avoiding saying it because I always say quarks. You want to go down to the synapse strength level, which what you just said, being excited about some of these low-level implementation processes, different weighted sums essentially separated into what are biological compartments, apical and basal dendrites. We know it's way more complex than that. Even just with those two subsets, and people like Matthew Larkum do work on apical versus proximal, distal.

That makes a little bit more sense to me, why you would want then-- are those two directly related? Is that why you're wanting that sort of synaptic? Because these low-level ways in which the higher-level, let's say attention in this case, higher-level functions, you are connecting those in your pursuits?

Xaq Pitkow
There are connections there, but that's not what I had in mind when I said that we need the synaptome or whatever. What I really had in mind was that one of the fundamental things that our brains do is learn, but we can't measure it. The thing that is taken from-- not at scale. The thing that has taken us from the single neuron doctrine to the population doctrine and understood at least some elements of how patterns of neurons are related to driving behavior and understanding the world is that we can measure them at scale.

We can measure now-- I think the world record is a million neurons at the same time in one animal doing one thing. That gives us much richer understanding of what is there, right, what the brain computations are doing. When we talked earlier about, well, maybe the only thing that we can really understand about the brain, although I don't agree with that claim, but those people who think about that is that the only thing we can learn is the learning rule and the objectives of the brain. We will not be able to see that learning rule operating in its natural context until we can see the synapses changing.

I actually did a small project with somebody trying to infer the learning rule from just neural activity in a simple case, where the learning rule was really simple. Inferring synapses, synaptic weights from activity is really hard. In fact, in many cases, it's impossible. Let's say you have causal perturbations. It's really hard to figure out what the connections are. Now you have to not just learn what the strengths are, but how they change and how that change depends on other activity. It's super hard.

If we had direct measurements of the synapses, as people like Ehud Isacoff are starting to study, like starting to measure, then we have the chance of understanding what the learning rule is doing at scale. The kind of rapid learning that I would love to measure is one-shot learning. You know what a car is, the first time a convertible. What all of a sudden changes? Bobby Kasthuri, who was involved in some of the earliest connectomes in the modern age here, he was suggesting that we could use bird imprinting as a model system to find really rapid, huge-scale synaptic changes that cause major computational consequences.

All of that requires looking at the synapses. That's not something I've done in the past, but to me, this is the biggest gaping hole in neuroscience is that we don't understand how learning works. All of the machines that we use for learning, they're doing gradient descent these days. That's basically what they do. The brain doesn't do gradient descent. Maybe it approximates it. What are the approximations? What are the constraints? We don't know. We don't know because we can't measure it yet.

Paul Middlebrooks
Why do we need to know the low-level implementation details if we know the learning algorithms or even approximates that have the end result be the same?

Xaq Pitkow
No, then we could. If you can find a way that we don't need to know the low-level details, that would be fine. We do think that the machinery depends on the-- The synapses are the things that do a lot of the learning. I'm definitely an advocate of taking a step more abstract.

Paul Middlebrooks
Yes, that's why I was surprised.

Xaq Pitkow
Yes, I would be very happy trying to understand that. I'm not sure I have a lot of confidence that it's going to be easy to do. Let's say we have a population of neurons that are physically connected with all these synapses. Then what do you do? You abstract from that some graph-structured inference algorithm that's there at a more abstract level. Now, in parallel, you have the low-level synaptic updates, and you have the updates in that graph, that abstract graph that you extracted. If you could just do the abstract graph updates, that would be great.

Then you could understand maybe that's a good way of understanding the way that things work, or maybe you need to go down to the low-level synapses. I don't know. One thing I do know is that when we measure new things, we find new insights. That's why I was saying this is like, I think we're right on the cusp of being able to do this. I'm sure that it's going to provide a lot of new insight.

Paul Middlebrooks
Maybe we'll end with this thought, and you can reflect on it as well, because I have to actually go get doing some neuroscience here in a minute. We'll have to have you back on because we didn't even talk about your work. We talked about a lot of the principles which undergirds all your work. I'll point to all your papers and stuff in your lab in the show notes. The interesting thing, so the arguments against studying the implementation level in the past-- How do I want to phrase this? I want to see what you think about this.

In modern neuroscience, we have these neural networks to work with, these probabilistic graph models, these theoretical tools, which, when we look at, let's say a synaptic strength changes and measure them, now we can actually relate them to these theoretical entities that have been developed in the past couple decades, like better developed. Whereas in the past, you're measuring synaptic strength with the assumption that they lead to learning, and you can measure it experimentally, but then there's still this huge gap.

Maybe we're in an era where we're being able to actually go across levels and close those gaps a little bit better and understand how the low-level implementation effects and details matter for the higher-level properties. How would you reflect on what I just said?

Xaq Pitkow
I would say that would be a great cause for celebration.

Paul Middlebrooks
Ah, all right.

Xaq Pitkow
Yes, absolutely.

Paul Middlebrooks
Do you agree with it, though?

Xaq Pitkow
Yes. Progress is moving gradually. I think we're gaining more insight. There's some really fundamental things that we don't understand yet. I think we understand some things, and whether that's a lot or a little is going to depend on the judgment. I was having a discussion with Konrad Körding the other day, and he was saying we don't understand anything. I was like, "I think we understand some things." Then, it was understand the brain. Do we understand the brain?

Paul Middlebrooks
No, what does that mean even?

Xaq Pitkow
I think we do understand some things about the brain, right? I think we understand. It's a mystery in a huge number of ways, but it's not a total mystery anymore, the way that it used to be. It's not like just a big ball of tangled fat and string. We know that there are patterns. [laughter] We know that there are synapses and synapses change. The patterns have influences on our behavior. There are feedback loops. We don't understand symbols. We don't understand language. We don't understand many of the dynamics.

We don't understand some fundamental models, like memory or most of the computations, but we have some hints. I would say we're on our way, and it's been facilitated by massive data. I think this is a great time. We have such an amazing confluence of factors right now that makes this a really good time. One is that we have analysis tools that have very high power from all the AI stuff. We have incredible neurotechnology, which is giving those models something to chew on. It's a good time to be a theorist and a good time to be partnering up with people collecting some of this epic new data.

Paul Middlebrooks
All right, Xaq. We carry on into the mystery of the brain and its functions with some less mysterious things along the way. Thank you for your time. I'll see you around campus, and we'll have to have you back on. I appreciate it.

Xaq Pitkow
This was great, Paul. Thanks so much.

[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 hear is a little slow, jazzy blues performed by my friend, Kyle Donovan. Thank you for your support. See you next time.

[music]
