A Lecture Companion

Free Lunches: Model Systems for Studying the Agential Gifts from the Platonic Space by Michael Levin

Michael Levin

Recorded on June 20, 2026

About this document

This document is a companion to the recorded lecture *Free Lunches: Model Systems for Studying the Agential Gifts from the Platonic Space by Michael Levin*, recorded on June 20, 2026.

This document pairs each slide with the aligned spoken transcript from the lecture. At the top of each slide, there is a "Watch at" timestamp. Clicking it will take you directly to that point in the lecture on YouTube.

Lecture description

This is a talk I gave, ~49 minutes long, titled "Free Lunches: Model Systems for Studying the Agential Gifts from the Platonic Space", to a philosophy audience. It covers some more ways I think about the latent space of patterns that in-forms biology and cognitive science.

Follow my work

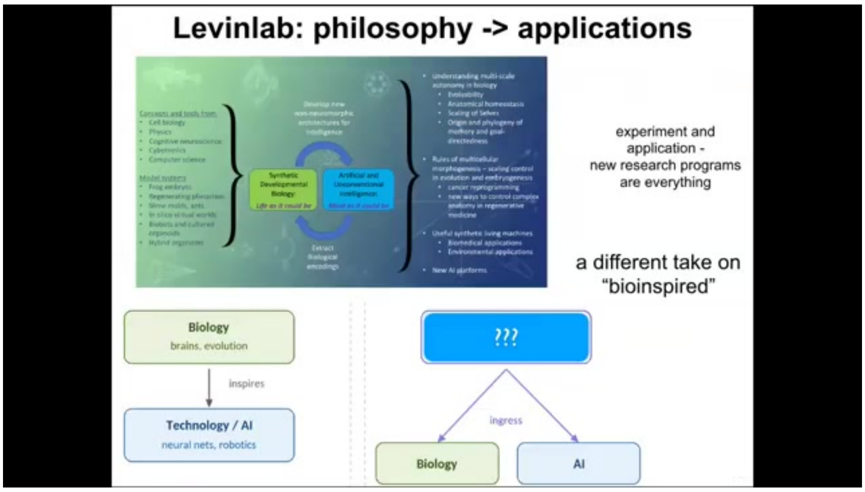
Twitter • Blog • The Levin Lab

Transcript note

This transcript is generated automatically. While we strive for accuracy, occasional errors may occur. Please refer to the video for exact wording.

Want one for your lecture?

Want something like this for your own talk? Reach out to Adi at adi@aipodcast.ing.

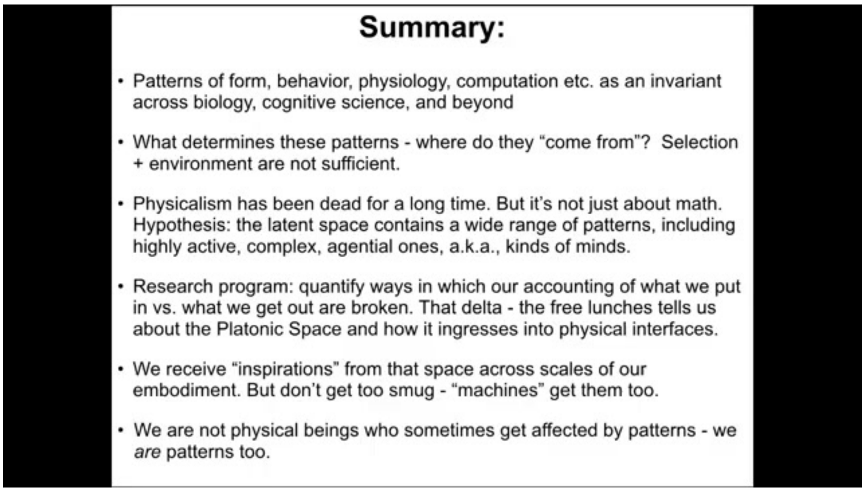


And much of what I'm going to tell you today, although not everything, because this needs to be significantly updated. I wrote this almost a year ago, so you can take a look in this paper, but I'm gonna upload a new version probably next week.

So the first thing I wanna say is that I run a wet lab. We do experiments at the intersection of biophysics, computer science, and cognitive science. And what I think we do, on our best day, occasionally, what works out is that we can take some deep ideas from philosophy and actually push them all the way through practical applications. So the outputs of our lab are things like discoveries for regenerative medicine, birth defects, organ regeneration, cancer, AI, bioengineering. These are things that come out. And so everything that I'm going to tell you is aimed at experiment and application. The idea is that we have to have new research programs. That is everything. And so someone asked a moment ago around whether these things have to make contact with predictive experiments. Absolutely. Not only to explain things that people have already seen, but actually to give rise to new research programs. That, I think, is absolutely critical. Not just looking backwards, because looking backwards, you can epicycle almost anything into a conventional paradigm. The question is, does it lead you to do new experiments and find new things with new applications? So that is kind of...

The good news is that I'm going to show you lots of experiments and lots of model systems for studying these questions. The bad news is that it prevents me from saying some very sort of cosmic kinds of things that people often ask me about, and I typically don't talk about those things because unless we have either data or a way to address them experimentally, I don't talk about it in public. So one of the things that people sometimes say we do is bio-inspired engineering. Now, the way typically people think of bio-inspired is that they take things that biology is doing, and then they try to make technology, for example, AI, that's going to use those principles. I actually think that it looks something like this. Actually, bio-inspired doesn't mean you take inspiration from biology. Bio-inspired, to me, is figuring out what is the biology inspired by, and thus what is the engineering going to be inspired by? So these things are lateral, and so now here's the question. What is it inspired by?

Slide 2 of 27 · Watch at [2:20]



So I'm going to try to tell you what I think at this moment. So a summary of the whole thing goes like this. I'm going to first argue that patterns, in particular patterns of form, patterns of behavior, patterns of physiology, patterns of computation, all kinds of things are actually part of the sa– They're the same thing. So morphogenesis and behavior in three-dimensional space are not different. They are just different aspects of different kinds of patterns. And I think it's a critical invariant that goes across biology, cognitive science, and many other disciplines.

Now, we're interested in knowing what determines these patterns. Where do they, quote-unquote, "come from"? I'm going to address this, but I will tell you right now that I think selection and environment or history of biology and physics are not sufficient to answer this question. I'm also going to say that physicalism has been dead for a long time. Mostly we knew this because of mathematics, and I think Pythagoras and probably long before was already aware of this, but it's not just about that.

My hypothesis is that this latent space that we're going to talk about contains a very wide range of patterns that include the kinds of things mathematicians study but also highly active, complex, high-agency patterns that the behavior scientists would recognize as kinds of minds. Now, the reason I didn't talk about this prior to 2025 is that we didn't have a research program for it. You know, lots of, as was just said, lots of kind of deep thinkers over the millennia have said vaguely things like this, but that's different from having a research program to actually do experiments, and now we do.

We now have model systems, which I'll describe to you, where we can actually quantify, identify and quantify ways in which our current frameworks are broken. In particular, what's broken is our accounting of the effort we put in and what we get out. What we get out is often much more than what we've put in. The delta, the difference there, is what you can call free lunches in the physicist sense, tells us about the space, the structure latent space from which these things come, and we can now do experiments to find out why it comes into certain physical interfaces and not others. Well, actually all of them, but in different ways.

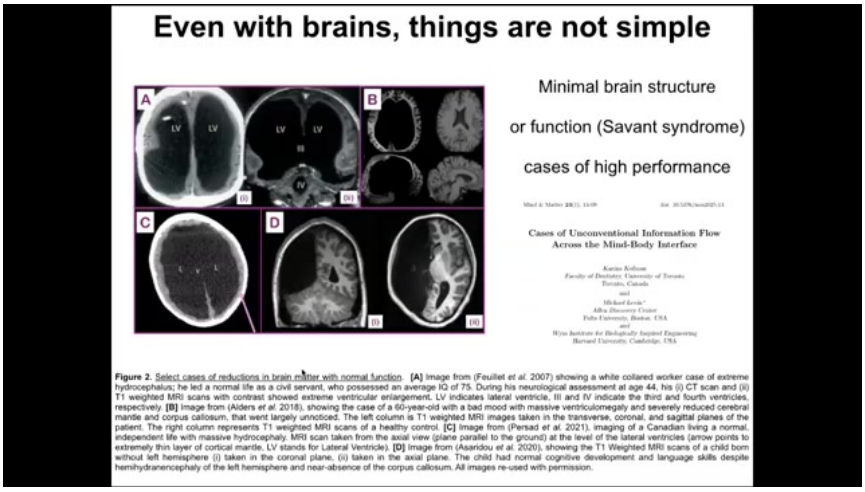
So I'm also gonna argue at the end that we do receive the so-called inspirations from that scale, from that space across scales. So your molecular networks get them, your cells get them, tissues, organs. They all get them. But don't feel too proud of yourselves because the quote-unquote "machines" get them too. This is, well, whatever this is, it basically soaks into absolutely everything as far as I can tell.

And finally, and I won't have time to talk about too much, but we can talk in the Q&A. I don't think we are physical beings that occasionally get affected by patterns from the space. I think we are patterns. We are the patterns looking out into a physical interface. So one thing that's interesting about all of these kinds of questions that we're asking is that we can start to delete certain sharp categories that have been with us for a long time, you know, from pre-scientific times. And I think these categories are now doing way more harm than good, and so I'm going to just point out a couple.

One is this so-called difference, or distinction between magical living beings on one end and dumb machines on the other end. And so we know from a developmental scale and also an evolutionary scale that we were all single cells once. You know, little blobs of chemistry and physics, unfertilized oocytes. Eventually, slowly and gradually, we became the subjects of behavior science, psychoanalysis, love, all of those kinds of things.

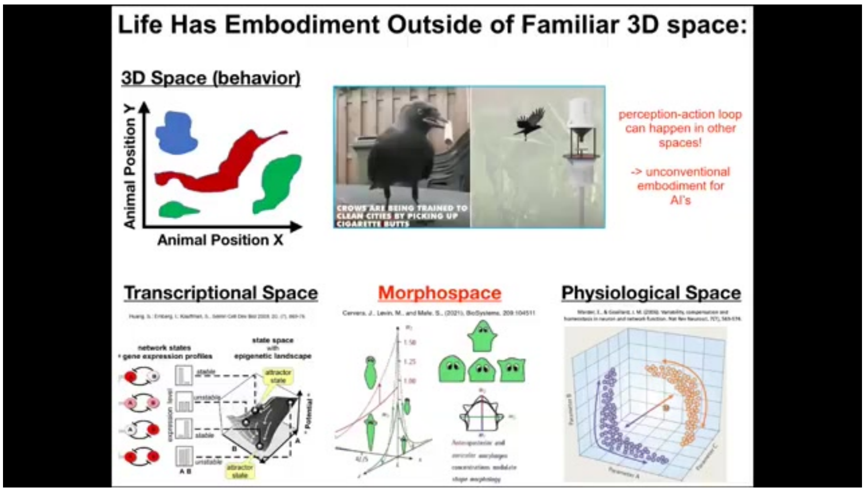
But not only that, there's another continuum, another gradual continuum in which both biological changes and technological changes, many of which are already happening, are telling us that it's really very non-trivial to write down a description of what you think a machine actually is, and the implications of this are going to be clear in a moment that it's actually a bad category that leads us astray.

And so the thing is that the problem we are all going to face in society is this kind of thing. It's not the language models, and it's not the AIs that everybody likes to say are not like humans. That's fine. This is what we're dealing with. Your neighbor is going to have some percentage of their brain and body replaced, whether for medical reasons or just because of freedom of embodiment. They're going to look very different in the future. And basically across this spectrum, okay, the spectrum of intelligence all the way from very simple kinds of things that can only be interacted with by physical rewiring versus the tools of cybernetics and control theory versus behavior science versus language and friendship and so on. This is a spectrum, and we really need to start to understand our own origins and our own possibilities in light of the spectrum of possibilities of what's going on. And I think our language really does us a disservice here. We have two words to describe things, who and what. If you want to describe a noun, a system, you can either say... You can either call it a who or you can call it a what, and I think this is a fundamental problem that forces us to think in binary terms. I don't know what the answer is, but maybe we can have something like this, like a who with a little exponent, and the exponent tells you to what degree where you are or where that system is on the edge. Some way of letting our language catch up to actually the discoveries of diverse intelligence research and away from ancient categories where we really did only have the two categories.



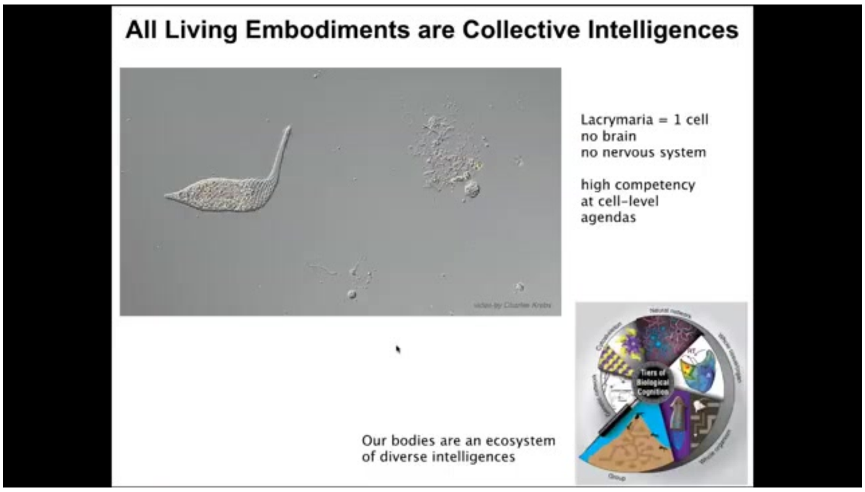
So, one thing to keep in mind is that even when you're talking about traditional brainy organisms, and I'm going to talk about cell intelligence and things like this in a minute, but even in traditional brainy organisms, already you can see that things are not simple. The standard mapping of the hardware that's required to maintain a certain amount of efficacious functionality, let's say a normal human IQ and a normal human personality, there's already some weird mismatches. And Karina Kaufmann and I reviewed them in this paper, where there are certain individuals with massive reductions of brain volume but normal or above normal functional intelligence, and many of these people don't even know they're missing most of their brain. So what's going on here? Why is there... I mean, this is not common by any means, but it doesn't take too many of these clinical cases to understand that something is wrong in our accounting of what the mapping should be between the hardware and the performance that you get out of it.

Slide 5 of 27 · Watch at [8:56]

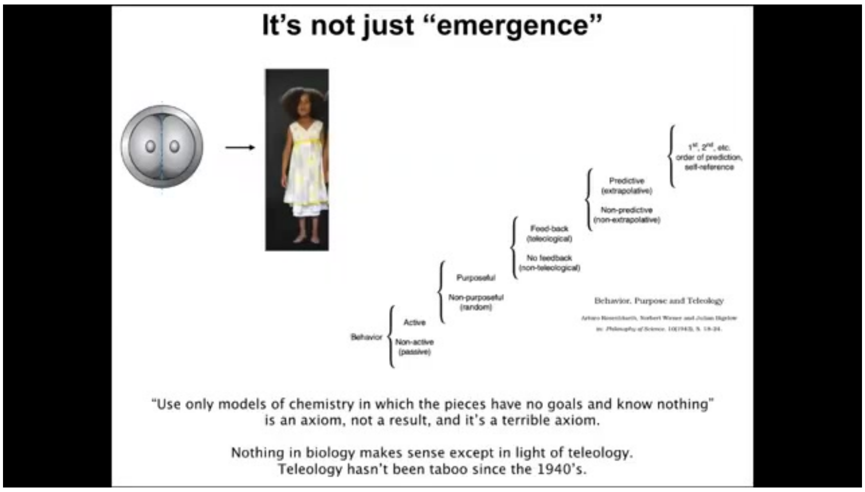


The other thing that I will point out is that life has embodiments and traverses many different kinds of spaces besides the familiar three-dimensional space. So we should remember as humans, with our own evolutionary history, our own kind of obsession with vision and three-dimensional space, that while we are okay at recognizing intelligence when it moves in the same 3D space that we do, roughly same timescale, roughly same spatial scales, biology has been solving problems, AKA intelligent behavior, in many, many spaces before nerve and muscle appeared, long before. So high-dimensional transcriptional spaces or gene expression spaces, physiological state spaces, and anatomical morphous space, which is mostly what we work on in the lab. Biology navigates all these spaces, making decisions, taking sensory readings, doing goal-directed functionalities, and this has lots of implications for what we're going to talk about.

Slide 6 of 27 · Watch at 9:58



So, another thing to keep in mind is that we are all collective intelligences. Not just the beehives and the ant colonies, but all of us. We're all made of things like this. This is a single cell, no brain, no nervous system, but you can see it's very competent in its own local cognitive light cone here. It's doing everything it needs to do. This is the kind of thing that we are made of. And so our bodies are kind of an ecosystem of diverse intelligence, all the way from molecular networks up through all the different components and layers. All of these things have agendas. They also have problems. Most of them can learn and have learning capacity. So this becomes also important that this notion that we, the verbal intelligence that you can have a conversation with, is the one inhabitant of a body is absolutely wrong, and that's before you even get into microbiome and things like this. Just the layers of your actual body are already an ecosystem of minds that live with you.



Now, now let's lay down some fundamentals here. First of all, so as was just mentioned, we know this amazing process. You start with an egg. You eventually very reliably end with something like this.

But I want to be clear for a number of reasons that I'll mention, the standard concept of emergence, you know, lots of simple local dumb rules, rules that are well-described by chemistry in which systems have no goals. They don't know anything. They just roll on, roll forward, and eventually something complex happens. You can get complexity that way. It's very clear that by following simple rules, you can get very complex outcomes and, in fact, reliable outcomes. But that isn't what's going on here.

And the thing is that this idea that we can only use models of chemistry where the components have no goals and don't know anything is an axiom. It is not a result. It is not something that was shown or derived or proven to be successful. And in fact, it's a terrible axiom. I understand it might have been useful when the only options were dumb like rocks versus smart like humans, but that's no longer the case.

Since the 1940s, we've had cybernetics, and we've had lots of rigorous theory around how systems, whether natural or engineered, can have goals. And it doesn't mean human metacognitive size goals. It means there's a spectrum of techniques that you can use to understand how systems have goals. And so I... You know, teleology shouldn't have been taboo past the 1940s, and I think actually nothing in biology can be properly handled without understanding that sticking to this lowest rung of the cybernetic ladder is simply leaving too many useful tools on the table. It's... And evolution certainly didn't do that. It didn't stay at this low level.

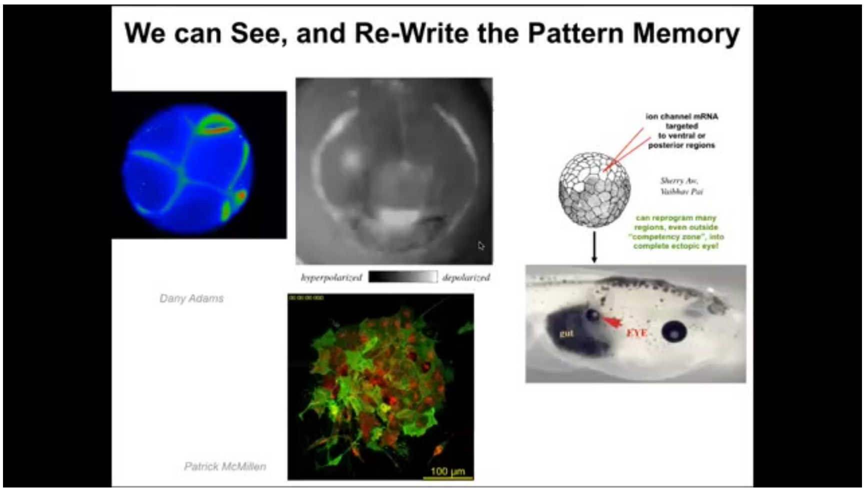
So it is not about emergence. Why? Because one of the things that we see in these complex systems is that they pursue a goal. They don't merely have complex outcomes, they go toward a specific pattern. Now how do you know? You can't just assume this.

You can't just say this because you see them reliably doing it. You have to do perturbative experiments.

So for example, if we scramble the face of this tadpole, it doesn't make a scrambled frog, which it would if all of the organs were just following their standard rules of how far to go in what direction. So these kind of Picasso-like scrambled tadpoles make pretty normal frogs because all the different organs will move in different novel paths and they will sometimes go too far and have to come back, and they all rearrange themselves, and they make pretty normal frog faces. And then they stop, and then they stop moving.

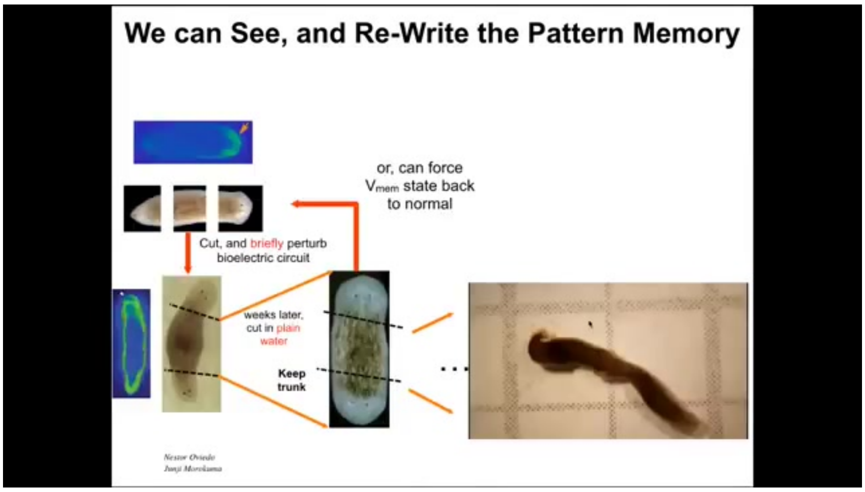
Same thing here, if you have an axolotl, you amputate the limb, these cells will grow really quickly, but then they stop. The most amazing part of regeneration is that it stops. When does it stop? It stops when the correct pattern has been completed. So all of this, again, doesn't... You know, ever since we've understood about homeostats, doesn't need to be magical. It basically is a system that represents the journey that it's taking in morphous space, in anatomical space. It navigates to that position using now well-known mechanisms and, when it gets there, then it stops like any other kind of goal-seeking behavior.

And in fact, now so, so how does it know what the patterns are? I ask that question absolutely literally because the way that these systems know where they are going is exactly the same mechanisms by which you know where you're going when you execute goals.

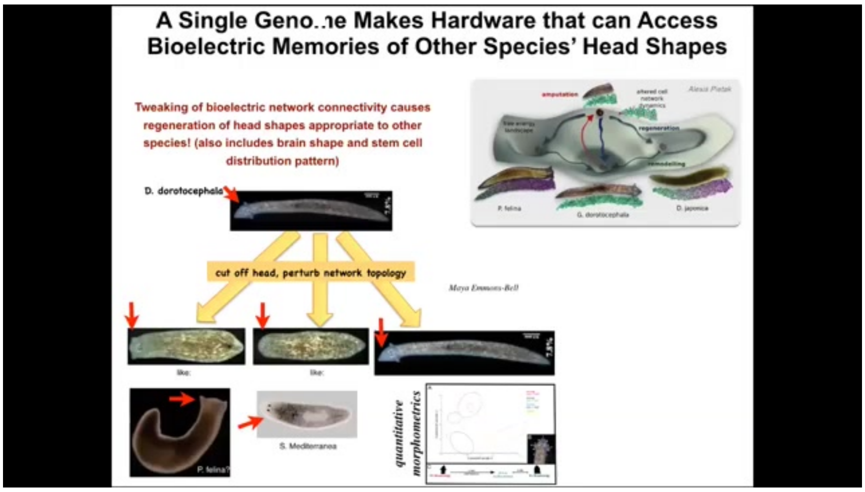


The same way that your collective intelligence uses electrical processes in your brain, ion channels, neurotransmitters, electrical synapses in electrical networks of the brain to remember goals and to guide goal-directed behavior toward those goals, that exact same system is evolutionarily ancient. It was used long before anybody had neurons to remember what our shape was supposed to look like. Back in 2000 or so, we developed the first molecular tools to read and write these patterns. So we can literally see this is not AI or simulations or anything like that. These are actual data from an early frog embryo and from some cells in culture looking at... Now we can read in the living state, we can now read the electrical memory patterns that non-neural tissue drives. And so this is a particular pattern here. This is... We call this the electric face. This is a pattern that shows up before any of the genes turn on to regionalize the tadpole face. So it already tells you everything you need to know. Here's where the eye, the animal's right eye is going to be, here's where the mouth is going to be, here are the placodes. What you're seeing here, by the way, is not some mystical energy. What you're seeing here are just voltage gradients between the inside and the outside of the cell plasma membrane. Just like in your brain, that's what the neurons are doing. But this network is much slower. It doesn't operate at milliseconds, it operates in minutes and hours, and what it likes to think about is shape. And so having seen this, one of the things you might do is try to establish the same bioelectrical pattern somewhere else and see what happens. And if you do that, so here's a tadpole, here's its normal eye. There's the other ones on the other side. So here's the mouth, the brain is up here, the gut. So if we establish by using ion channel injections this bioelectrical state somewhere else, here it is on the gut. The cells know exactly what this means. They interpret it as build an eye here, and they build an eye. So what I'm telling you here, which will be important momentarily, is that we now have the ability to read and write and rewrite, at least in some cases, specific patterns that serve as the goal for a bona fide goal-directed system, which basically works to reduce error. It uses an electrical network to gauge the delta between how are we now versus how are we supposed to be, and try to reduce that delta.

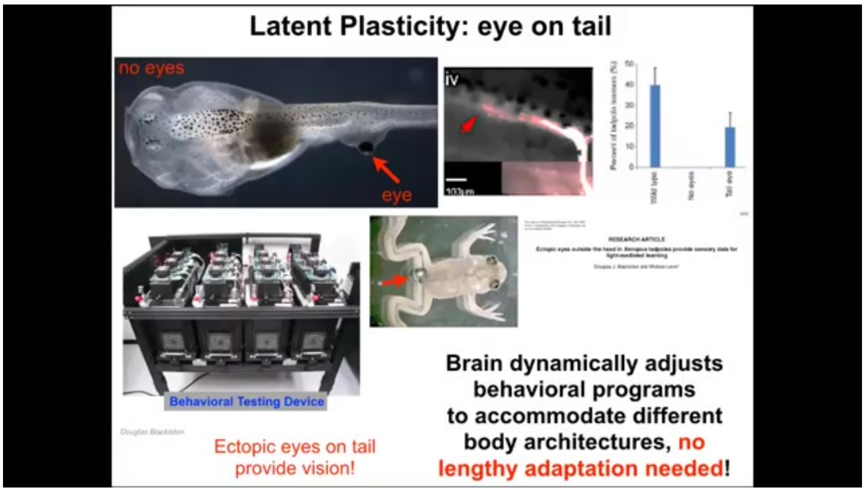
Slide 9 of 27 · Watch at 16:47



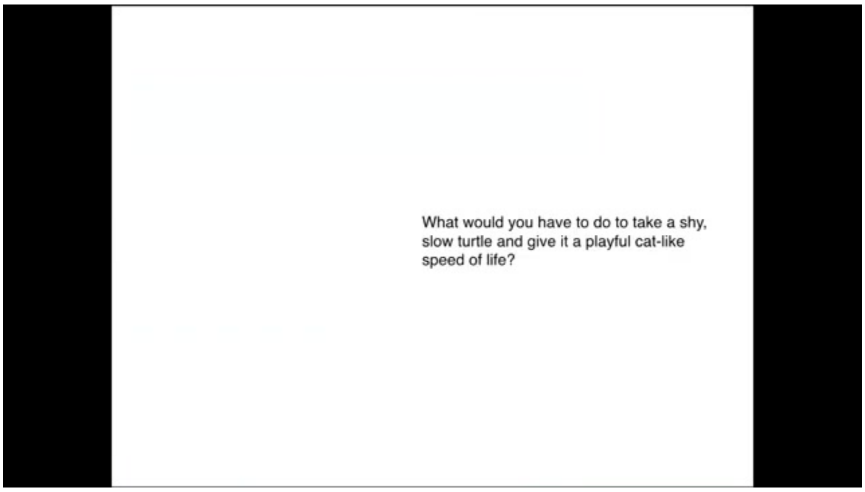
Here's another example of this in these planaria. Normally a nice one-headed flatworm. You cut it into pieces as many times as you want, very reliably you get a worm with one head, one tail. How do they know how many heads they're supposed to have? The answer is not genetics. This tissue is reprogrammable, and so you can take the exact same genetics and you can take this bioelectrical pattern that tells the animal one head, one tail, and actually enforce, change it to be two heads. And what that says to the tissue is, "Oh, a correct planarian should have two heads." And when you do this, guess what they build? Here's a nice two-headed worm. The memory is stable. If I keep cutting it, the pattern actually remaps itself onto new sizes of tissue, so you will continue to get two-headed worms as many times as you want in the future. Here they are. You can see them hanging out. And again, this has nothing to do with the genetics. What the genetics does for you is to provide hardware that is reprogrammable. But after that, as we all know from using computers, the hardware is just the beginning of the story. Then the question is: What are the informational goal states that can propagate through that excitable medium?



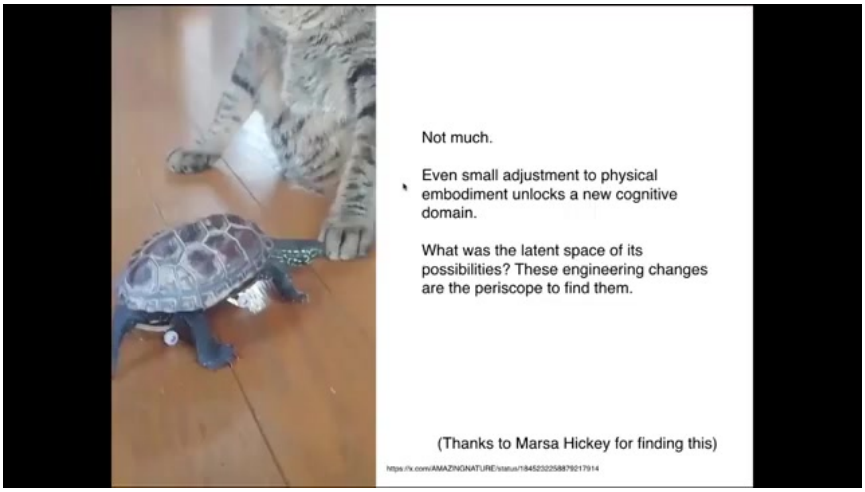
Now, one of the things this hardware is good at, and here's where we're actually gonna get into this issue of the most controversial part. Everything that I've said and showed you now is pretty much against the standard paradigm, but I mean, it's already in the text. It's in the developmental biology textbook. This stuff is pretty solid now for decades. I'm gonna show you the weirdest things next. But one of the things this hardware is good at is visiting other types of attractors in morphospace. So if you visualize anatomical morphospace with different shaped heads, flat heads, round heads, triangular heads, different shaped heads, what you can ask is, normally there are different species that live in these different attractors, so this is where they go. Could we ask a different species to go into these attractors? Turns out you can. So we can take a nice triangular-headed Dugesia doradocephala, cut off the head, perturb the bioelectric network topology, and it can grow flat heads like a pfaffelina, round heads like an N. mediterranea. About 100 to 150 million years' distance between these animals, and yet the hardware has no problem of visiting these other attractors, forming shapes of the brain, distribution of stem cells exactly like these other species. So the plasticity is remarkable. Now, this is what's going to lead us to the model systems that I was talking about. So far, what I've shown you is plasticity of standard forms. So I've showed you the ability to make a standard frog eye in the wrong location, but a standard frog eye. An ectopic normal planarian head, wrong location, wrong number, but normal planarian head. And here are the heads of other species. So let's go further. Let's start looking at shapes that, and configurations that have never existed before. And the reason we're going to do this is to really hammer this issue of how much effort did you put in and what did you get out? Because up until now, everything that I've told you could all of these stories could be told in a fairly standard evolutionary kind of tale, that why this, why the material's reprogrammable, why these signals work. I can tell a very standard sort of Darwinian explanation for that, but we're going to reach the limits of that.



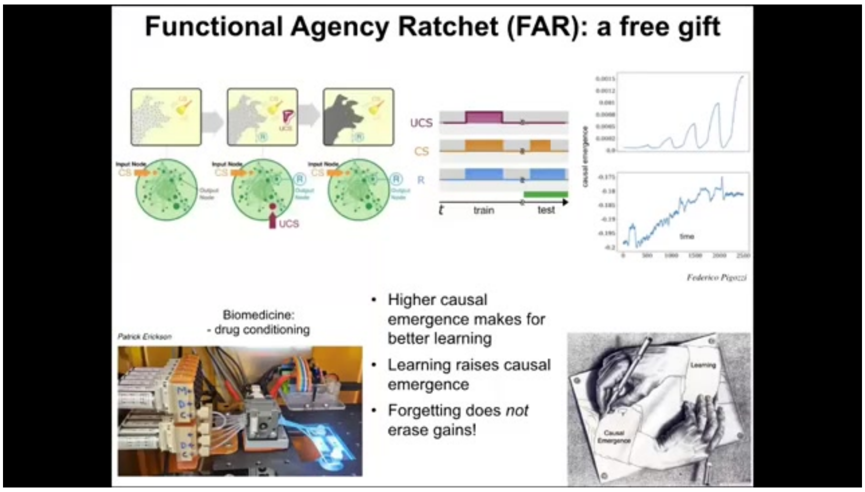
So here's a few examples just for fun. If I told you that I wanted to make a tadpole that would not have any eyes in the head, I want a tadpole with an eye on its tail. Furthermore, I don't want this eye connected to the brain. I want the optic nerve to stop somewhere on the spinal cord or in the gut or somewhere like that. And I want this thing to be able to see out of that eye. Even though the eye's in a weird location, it's not connected to the brain, I still want it to be able to see. What would you need to do? You might think that, wow, very difficult. Lots of rounds of just mutation, selection, adaptation, maybe some kind of crazy engineering of the nervous system that you would have to do. Turns out you don't have to do anything. When we create these animals, and we made a machine, this is an automated system that trains and tests these animals for vision, for... They learn and they learn visual tasks. Turns out that, yeah, when you force the cells on the back of a tail to make an eye and you track the optic nerve and it doesn't connect to anything, certainly not to the brain, in fact, they can still see. Didn't take any new rounds of evolution, of selection, of adaptation to take a novel sensory motor architecture and make it absolutely functional. Why does that work? That's really weird. And just for fun, if you wonder what happens to these as the tail disappears, the tail is killed off by programmed signals that kill off the tail in a frog, which isn't supposed to have a tail. The eye completely ignores all of those signals. It will not kill itself. It rides back as the tail disappears and eventually lands on the behind of the frog. If nothing else today you've seen a frog with an eye on its butt, so there is that. So okay, so that's it. Already we're starting to see there's something fishy here. Why does this work out of the box immediately with no accommodations needed?



Let me ask you another question. If I had a turtle, so this is a slow, shy reptile, and what I want is to have a playful cat-like speed of its life. I want it to be like a cat. What would we have to do? Now, you might think, wow, millions of years of evolution or maybe some kind of crazy neural engineering that nobody knows how to do.



Turns out you don't have to do much. So here's this guy that put a turtle on a little skateboard, and immediately this prosthetic, it didn't take a long time for the turtle to, first of all, move at a speed sufficient to play with this cat. He wants to play with the cat. The cat, I guess, doesn't know what to make of it, but the turtle is perfectly able to keep up with him. So what's going on here? This very small adjustment to the physical embodiment unlocks a new cognitive domain. I mean, have turtles for millions of years been wanting to do this and they just couldn't? I have no idea, but it makes you wonder, right? It makes you wonder, what would we be capable of with really fairly small tweaks? But again, that delta between what you put in and what happened. What is the latent space of the possibilities for this system? These kind of engineering changes, doing these weird things, and I'm going to show you some much weirder things, but putting a turtle on a skateboard is basically like a periscope to find these additional things that you did not see coming.

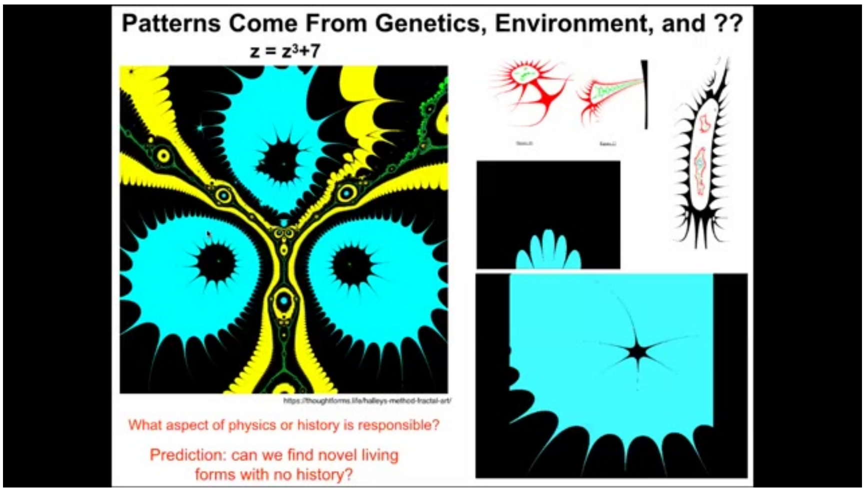


So, here's another interesting example of these free gifts, and now we're really gonna ratchet this up. We showed a few years ago that if you take small networks, molecular networks, so not even a cell, never mind a brain or a neuron, not even a whole cell, just molecular networks, and they don't have to be large networks. The smallest one that can do this has just four nodes. If you take molecular networks, you can train them. In other words, even simple molecular networks can do habituation, sensitization, associative conditioning. It is very clear how it works. It's dynamical system learning. Nobody had seen it before because nobody had thought to test these such simple models for such a thing as learning, but it turns out that they do. And here's the amazing thing. And we're using this in our lab, we're using this for medical purposes like drug conditioning and so on. It turns out that if you can do Pavlovian conditioning in molecular networks, it's basically a molecular placebo. Which means that you can associate powerful drugs that you don't really want to be taking with a neutral drug. Eventually, you pair a presentation and so on, and eventually you can use neutral drugs to get the same effect. So that's the practical application.

But there's something deep, very deep here, which is this: if you use causal emergence metrics... So this is the mathematical tools pioneered by people like Giulio Tononi and Eric Hoel and many others to try to quantify the degree to which wholes are more than the sum of their parts, right? So the integrated causal identity of a level beyond the low-level parts. So there's math for this now. It's no longer just a philosophical debate about the reductionism. There's actually math that tells you that some systems are absolutely not suitable for the reductionist assumption. There's something cool that happens. Networks with higher causal emergence are better learners. But also, when you train them, their causal emergence goes up. So what's happening here is that every time you train them, they become more and more of an integrated agent, but that makes them better at learning and so on. So there's this amazing positive feedback loop that we call the functional agency ratchet. Why is it a ratchet? Because if you force them to forget, and that's really important for medical reasons, you want your network sometimes to forget physiological experiences. When you force them to forget, you do not erase the gains that they've made in becoming a higher level integrated agent. So it's an asymmetry that points upward in terms of agency and intelligence. It's asymmetric. It points upwards.

But now, so now you say, "Okay, where did this come from? Surely evolution did this. Evolution because, I mean, that kind of a thing is a very sort of... You would have to select, you know, needle in a haystack for this kind of thing, right? You would have to select for networks." So it turns out that that ratchet, if you look at random networks compared to biological networks, what you see is that biology can improve it a little bit, but the random networks already do this. They are... I don't know if this is fine-tuning the way that we see in some of the parameters of the physical universe, but in this is not a needle in a haystack situation. Random networks are already optimized for this incredible ratchet. And so it's not about replicators or selection. There are no replicators in these systems. Nothing is replicating, nothing is being selected for. It does not come from physics. It does not come from biology. It is a free gift from mathematics. If you want to know where this comes from, it comes from the way that causal emergence and network properties work. It doesn't rely on any facts of mathematics. It doesn't rely on a history of selection and so on. So this is now kind of crazy.

Slide 15 of 27 · Watch at [27:15]

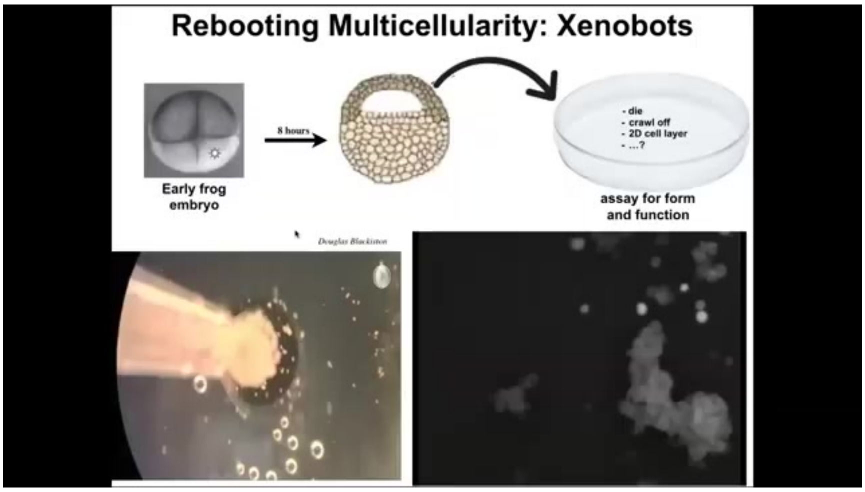


Biologists love to be able to explain things in terms of either genetics or environment. In other words, a history, right? Some kind of history of selection or physics. Those are the sources. So I think in looking at some of these things, I'm gonna show you some more examples in a minute. We have to start asking where else can information come from?

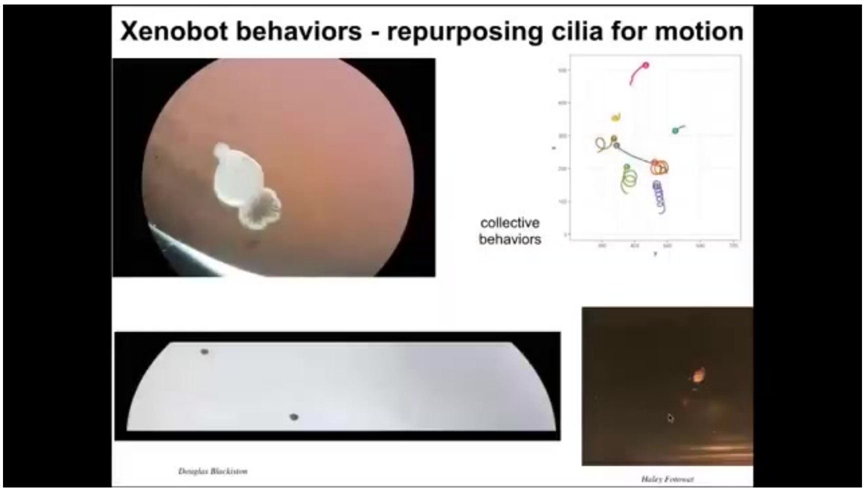
Well, I just wanna point out that if you look at something like this, this is the Halley plot of a simple function in complex numbers. So Z is a complex number here. Z cube plus seven, something like that. If you plot it out, gives you this incredible order. And actually, I can make a video of it, but by changing these parameters just a tiny bit for every frame here, you can see this amazing world. It's kinda cool. It doesn't hurt that these things look vaguely biological and organic. That's kind of fun.

But the one thing you have to realize is that if I want to know why the pattern of this equation is exactly this, not something else, but exactly this, I can't lean on physics. There is nothing about physics that explains this. I can't lean on selection or history. This was not selected to look like this. There's something else. So there are apparently facts that are not facts of physics. There are facts of, we call it mathematics.

And so now we have to ask why. What does this have to do with biology? Does it have to do anything with biology? Could we go further? I mean, I've shown you manipulation of standard forms. I've shown you some surprising sort of novel things you can do. Could we find living forms with no history of selection for their specific properties?

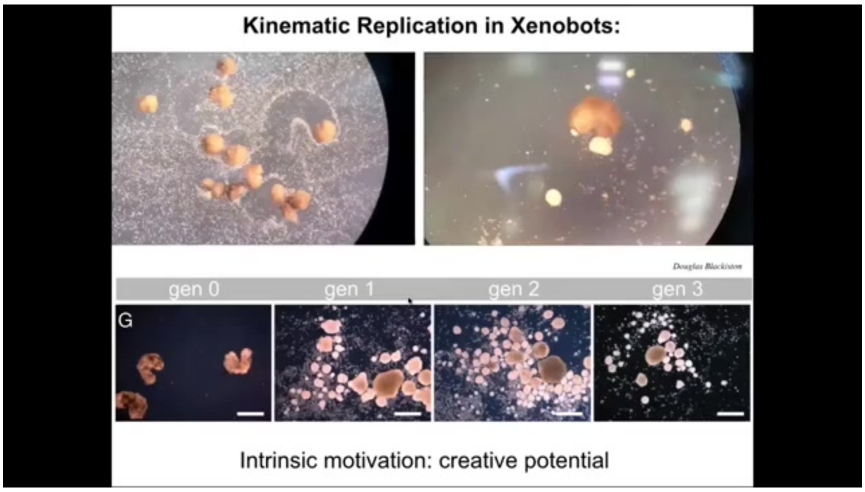


So I'll just show you two. One we call xenobots. This is what happens when you take cells from the epithelium, so it's going to be skin, from the epithelium of an early frog embryo. We liberate them from the frog embryo, we dissociate them, we put them in a container by themselves. They could have died. They could have crawled away from each other. They could have made a flat monolayer like cell culture. Instead, what they do is this. You can see it here. Each one of these things is a single cell. Here's a little group. They don't all look like a little horse. I just thought this was a cute example. They sort of move as a collective. They assemble, and they assemble into something we call a xenobot.

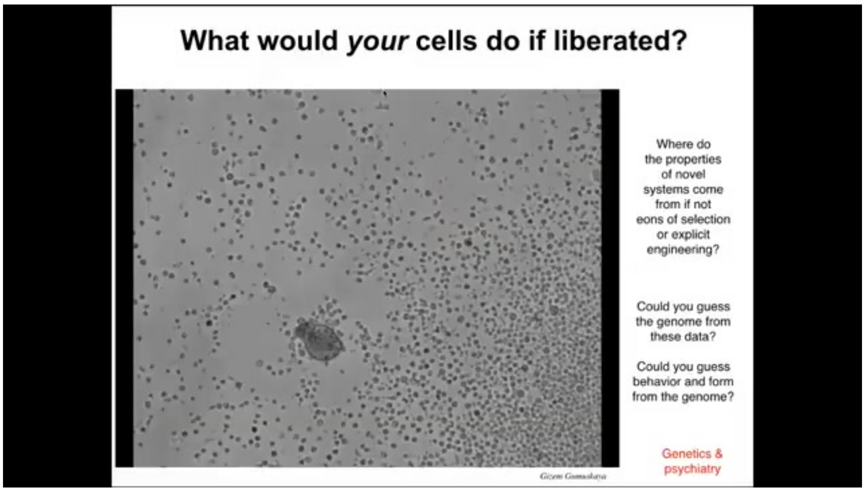


What is it doing? Well, first of all, it has little hairs. The hairs are used by a frog to move mucus down the body and then get rid of pathogens and things like that. But these guys are using them to swim. They're using them to row against the water. What you're seeing, this, this has never existed before. This is a patch of frog skin rebooting its multicellularity into a new way of life. They can move in circles. They can sort of patrol back and forth. They can have collective motion. They can, with neurons, and you can... This does not have any neurons, but my postdoc colleague threw some neurons in there, and they can do all kinds of interesting things. Here's one sort of swimming around.

Here's one traversing this maze structure. So it goes here. It's gonna take this corner without bumping into the opposite wall. So it can take the corner, then for some reason it turns around, goes back where it came from.

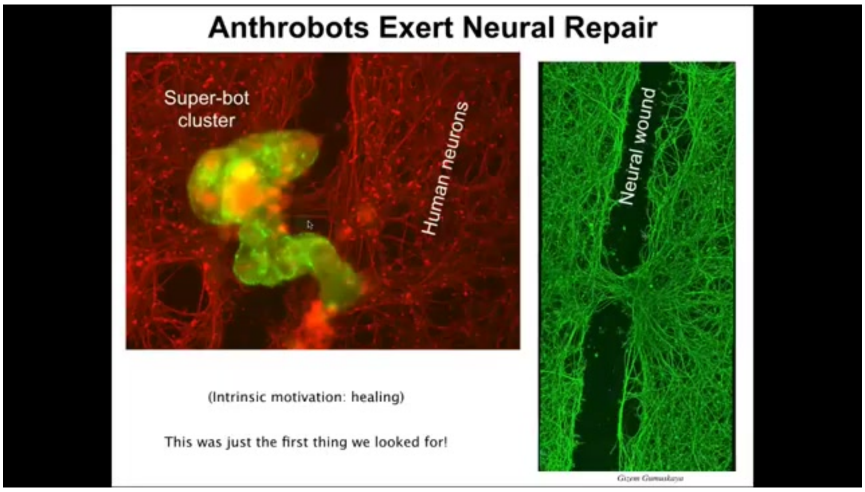


One of the most amazing things they do, we call kinematic self-replication. So look, we've made it impossible for these creatures to reproduce in the normal froggy fashion. We give them a bunch of loose epithelial cells. That's what this white dust is here that's sprinkled everywhere. These are loose cells. And what they do is they run around both collectively and individually, and they collect the cells into little piles. The little piles mature into the next generation of xenobots, and guess what they do? They do exactly the same thing, and they make the next generation, and they make the next generation. So this ability, it's kind of like von Neumann's dream, right? Of a robot that goes around making copies of itself from materials it finds in the environment. There is no genetic editing. We didn't touch the genome. There are no new synthetic circuits. We didn't do any synthetic biology. We didn't add any scaffolds. There are no nanomaterials. There was no learning. There was no training, there was no selection, and there was no engineering of these things beyond taking the cells, liberating them from the other cells, which typically bully them into having a boring life as a, you know, outer skin, and letting them reboot into a new lifestyle. And here they are doing kinematic replication. Now, you might say, "Okay, well, maybe amphibians are weird. Maybe this is some kind of amphibian, like, weird amphibian thing." And I would ask you, "Well, what do you think your cells would do if we liberated them?"



What would your cells do? Well, here's a little creature. We call this an anthrobot. If you were to sequence it, you would find one hundred percent Homo sapiens. You would have zero clue from the genome that this is anything but a human being. They also run around. They have spontaneous motion. You could not guess what genome it had by analyzing their shape or behavior.

Slide 21 of 27 · Watch at 32:14

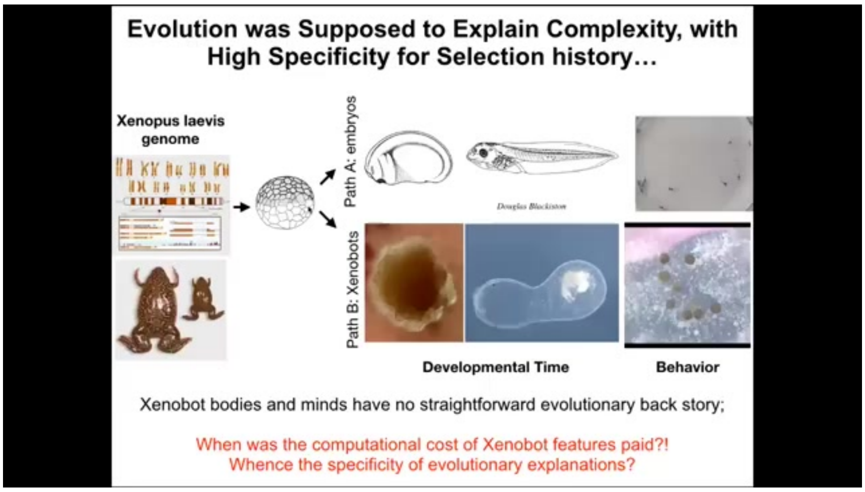


One thing that we discovered they do is they have this amazing healing function. If I plate a dish of human neurons and put a big scratch down the middle, the anthrobots will find the scratch. They drive down the scratch here. They settle somewhere into a group like this, and what they immediately start to do is knit the neurons across the gap. Okay, here they are.

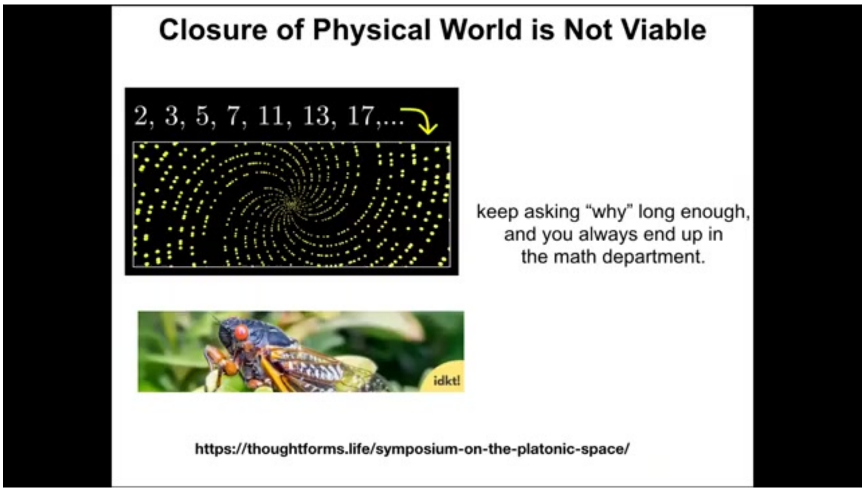
These cells come from adult, not embryonic. There's no embryonic material here. These cells come from adult tracheal epithelial samples donated by patients during biopsies. We buy the cells. They make these anthrobots. Who knew that your tracheal cells, which sit in your airway for very long periods of time dealing with dust particles and so on, have the ability to assemble into a little creature? It has nine thousand differentially expressed genes, so half the genome is differently expressed. It is actually younger than the cells they come from, so it actually rolls back its age based on epigenetic clock data. And they have this ability to heal neural wounds. Obvious applications here, right? This can, in theory, go into your body, do repair. You wouldn't need immunosuppression. These are your own cells, right? They're not genetically engineered. They can go right into your body.

And they have this amazing, amazing healing function which we did not select for, we did not train them for, and we did not engineer.

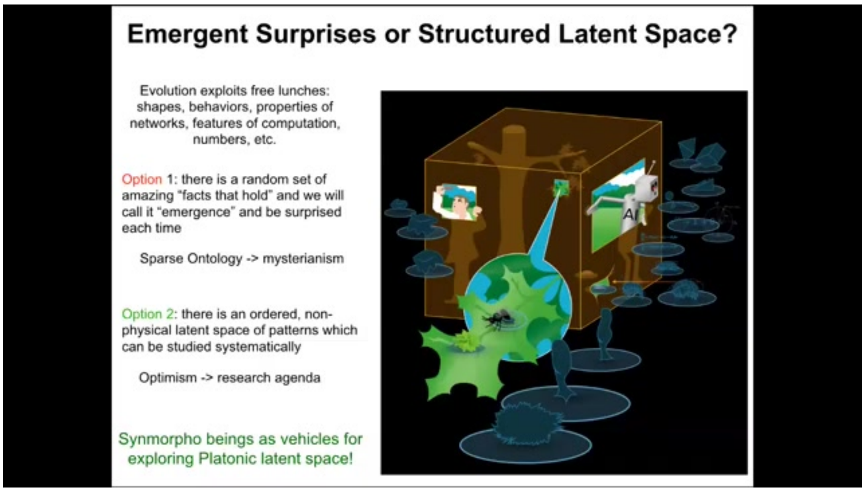
Slide 22 of 27 · Watch at [33:29]



So, so here we go. This is the final piece of all of this. The whole point of evolution is that it was supposed to explain complexity with a high specificity for selection history. In other words, if you ask why a specific animal or plant looks the way it looks, acts the way it acts, you should be able to tell a story. You should be able to guess the environment and tell a story of the selection pressures that killed off everything else and left this. Okay, that's a story maybe you can tell for the developmental stages of the frog and eventually the behavior of tadpoles. But what about the xenobots? There's never been any xenobots. There's never been selection to be a good xenobot or a good anthrobot. None of these things have been here before. No other creature on Earth does kinematic replication. And you can't really say that, at the same time that the frog genome was learning to make a good frog, it also learned to make xenobots. That isn't how this is supposed to work. There's supposed to be some degree of specificity, and in particular, the computer science forces us to ask, when did you pay the cost? Because computation and design always have a cost. When did you pay the cost of designing good xenobots? We know when you paid the cost to be a good frog or a good human, eons of selection. Where does all this come from? When did you pay this cost? Right? So we have to... So, so we already know that our standard way of accounting for outcomes, right? So the math that says you're supposed to put in some amount of effort. Effort comes in three forms. You either design it with an algorithm, you select it via evolutionary algorithms, or you train it by learning. Those are the three ways we know how to put in effort. When you get something like this, none of those three things happened, and so something is broken. There's something, there's something we're missing here.

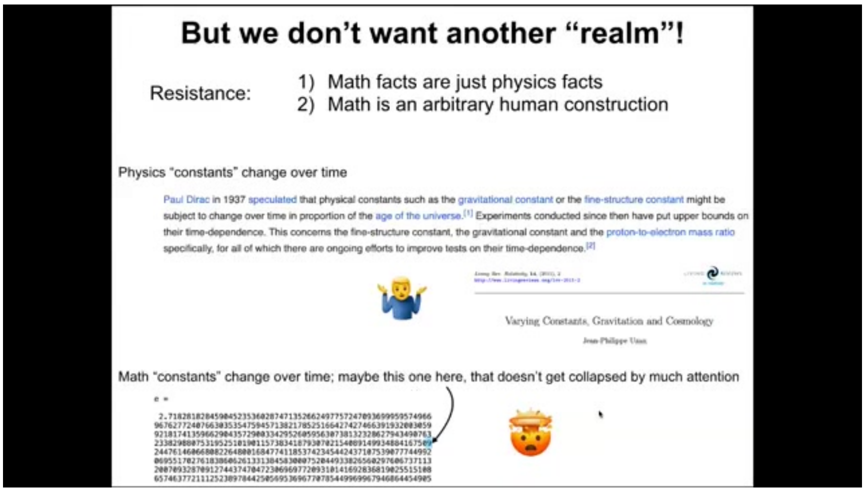


And I would argue that here... And I'm about to talk about this platonic space, and my point isn't that I saw xenobots and anthrobots and immediately went Platonist. That isn't it. I've been thinking about these things in biology for decades, but now we have a model system, in fact, two classes of model systems, the biological and the computational, where this is now actionable, where we can actually do experiments. But the fact I think we already knew long before this is that physicalism is not viable. Pretty much anywhere you start in biology or physics, if you keep asking why, like a five-year-old, right? You take a fact and you just go, "But why? Why? Why?" And you keep asking why. Eventually, you end up in the math department. If you wanna know why the cicadas come out at thirteen and seventeen years, the biologist says, "Oh, it's so that the predators don't time their arrival and eat them." Ah, so why thirteen and seventeen? Oh, well, it's because those are prime. But why are those prime? Go talk to the mathematicians. Everything ends like this. And in fact, if you wanna hear more talks about this, we have a symposium here. There were about thirty talks of some amazing people giving lectures about this. So here's what I think we have. What we already know, and I could talk for an hour, just examples, just facts of mathematics that are not underwritten by any facts of physics, that cannot be changed by anything you do in the physical world, that cannot be discovered through physics alone, you know, truths of number theory and the specific value of Feigenbaum's constant and all this kind of stuff. We already know there are facts that are not physical facts. That's been known forever. We know that these things impact what happens in biology and they impact what happens in physics. If you wanna know what particle properties, eventually it all boils down to the symmetries of certain mathematical objects. And so now you have a choice.



What most people do with this currently is they say, "Look, these things are emergent. The reason you got xenobots is because emergence." They say, "Well, what does that mean?" And they say, "Well, they're just specific facts that hold in our world. Let's have monism. We're not going to have any extra realms with structure in it. We're just going to say there's physics and occasionally interesting things emerge." And to me, this is incredibly pessimistic. This is a very mysterion view. I don't want to occasionally be surprised and write things down in a big book of emergent surprises. I would rather assume that... And this is a metaphysical stance, of course, you can't prove this, but I would like to assume for the purposes of experiment that there is an ordered, structured space of patterns that underwrites some of the things that we see, okay?

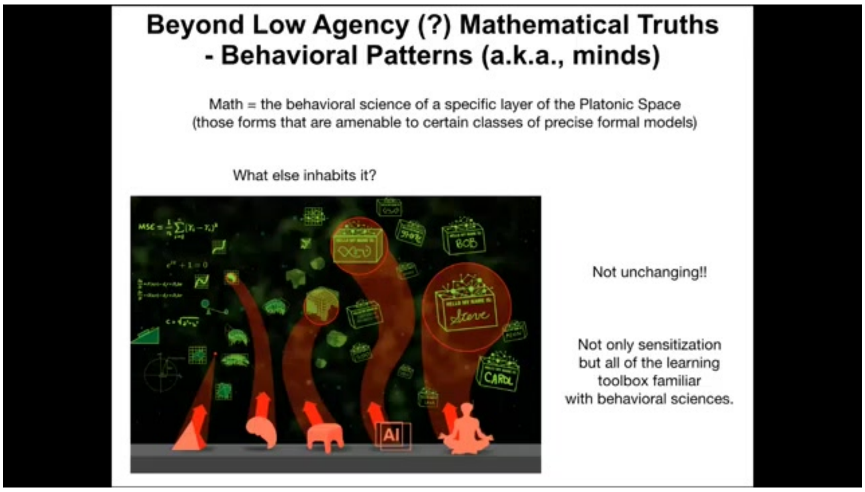
So the Platonist mathematicians, and I realize this isn't all mathematicians, but a good chunk of the mathematicians agree with this. It's not random. These things aren't random. They've been studying this for millennia. They've been building a map of mathematics. They already know there's a set of truths that you cannot change or derive from the physical world, and they are systematically studying these things. So what I think is happening when we build these weird synthetic morphology creatures, but also when you do anything, when you make an embryo, when you build a quote-unquote machine, we'll get to that in a minute, all of these things are kind of a window on this latent space of possibilities, this latent space of patterns.



And at this point, people often say, I'm not sure about this audience, but typical audiences I talk to are extremely resistant to that idea. They say, "We do not want another realm. We don't want a realm." Even though I'm not arguing for a mysterious realm, I'm saying the exact opposite. I think emergence is mysterion. I'm arguing for a latent space that is structured that we are going to systematically investigate, but people still don't want another realm. And so what they typically say are one of two things. They say, "Well, there are no separate math facts. Math facts are just facts of physics. Eventually, we'll get to that. We'll know that." And/or sometimes people say, "Math is an arbitrary human construction. Math doesn't exist outside of our thinking about it." So, my latest way of addressing this kind of thing is I ask how you feel about two claims. If I claim that the constants of physics change over time, most people kind of shrug their shoulders and say, "Eh, so the speed of light changes a little bit over cosmological time scales, so the gravitational constant changes. Who, no big deal. Who cares? Paul Dirac already said this in '37, that some of these things might float a little bit."

Now let's try this. I tell you, "Hey, the mathematical constants change over time, right?" So here's the first part of the expansion of the natural logarithm E. Suppose I tell you that, yeah, I think somewhere out here in the 10,000th space where people don't really pay any attention to it, so we don't really collapse it, in that sense, maybe those things change over time. At this point, everybody, I've not met anybody yet who said anything other than somewhere between incoherent and impossible. Definitely not, right? So if that's the case, if you agree that this is possible and this is not possible, what you've already told me is three things. First of all, that mathematical facts and physical facts are absolutely not the same kind of beast. In fact, you expect the mathematical facts to be more stable. You're far more freaked out if I suggest that they change versus facts of physics. And that math is actually not arbitrary or changeable because you can't make it be whatever you want. You are forced. Once you start with like empty set and successor of empty set, right? So you start with basic logic or something like that, then eventually you get to find out that E is two point seven eight, seven one eight, whatever. You don't get to choose it. You can't change it. You are given these things. So the characters in our novels can do whatever you want them to do. E does whatever E wants to do. You don't get a choice about that.

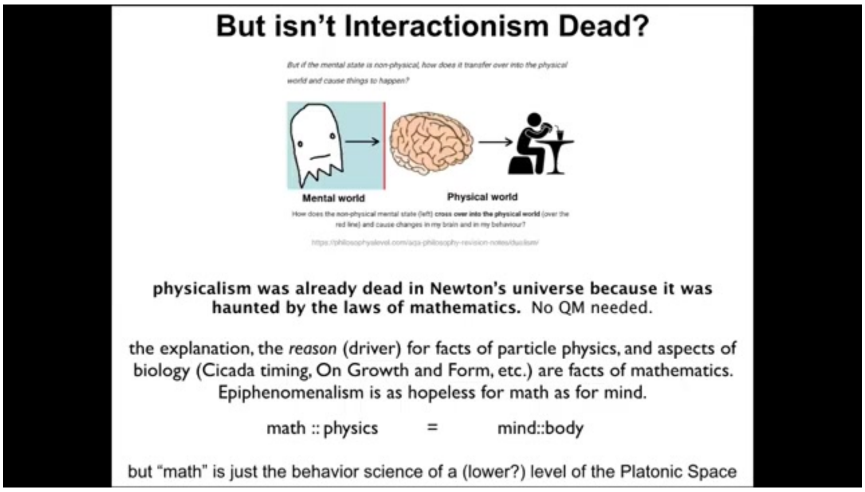
Slide 26 of 27 · Watch at [41:03]



So here's how I extend this. So far, all of this is straight up mathematical Platonism. Here's how I extend this for the purposes of this discussion. I just have a couple more things and then I'm done.

First of all, I think that the standard assumption that this latent space contains only the low-agency static facts of mathematics and everything else is a product of emergence and can be handled by physicists, I think we should loosen that assumption. And we should perhaps say that maybe the space contains other patterns that are of interest to developmental biologists, cognitive scientists in particular. Maybe some of these patterns are high-agency things that we would recognize as kinds of minds. Maybe math is just the behavioral science of one layer of that space. I don't know if E and things like this actually change. I don't have any claims on that. But in my current model, all of these things are not eternal and are changing. These are high-level, the ones that we associate with both morphology and behavior and physiology and so on. These are absolutely dynamic things that can change, but not only by sensitization, which you just heard about. I suspect that they use all of the toolbox and probably more that we haven't even thought of, of the behavioral sciences. So sensitization, Pavlovian conditioning, delayed gratification, path planning, language, all of these things are right there along with sensitization once you've gone down this road and realized that why should this space have only patterns that are amenable to the formal tools of mathematics? What about the rest of it?

Slide 27 of 27 · Watch at [42:40]



So we can talk about this idea again. In most audiences at this point, people freak out and they say, "Look, you're basically saying that minds exist in a non-physical space, but they affect the physical world in the same way that the truths of, about prime numbers affect what the cicadas are and aren't gonna do when they evolve." Interactionism should be dead. You know, Descartes had this problem with interactionism or whatever, and my point is simply this: whatever the actual resolution to this, it's been here for a really long time because we already know that there's a degree of interaction between non-physical truths and physical objects. The relationship between math and physics already tells us this, and maybe the relationship of minds and bodies, and when I say minds, I don't just mean our human minds, I mean the minds that power the problem-solving of morphogenesis, of physiological space navigation, of transcriptional space navigation. All of these things are diverse intelligences of different grade and different type. Maybe the relationship between those minds and bodies is exactly the same as between math and physics.

So I'll wrap up here just to say that the one thing is some people like this and they say, "Ah, so here's what it is. Biology is... and its complexity are really primed for these kind of ingressions of minds from this space, and that's what distinguishes us from dead matter, from mere machines, all of that." And I will just point out that we did work on, I don't have time to go into all the details, but anybody who's interested, I have long discussions of this online. We, a lot of humility is warranted here because what we found is that even extremely simple deterministic systems are susceptible to ingressions of not just complexity, not just unpredictability, but competencies that are familiar to any behavioral scientist, and they are not in the algorithm. The algorithm and the materials of even a simple machine tell you what the machine must do and it tells you what it cannot do. But between those two things is a massive amount of degrees of freedom, even in small deterministic things, apparently. This was a shocker to me. And they can host interesting patterns from that space.

So I don't think it takes life or cells or large complexity to start to become an interface, and I think that's what all physical systems are. They're interfaces for specific patterns from that space. And therefore, we have to remember that as much as we all would like to think that the rules of chemistry do not tell the entire story of the human mind, I would say that those kind of things don't tell the proper story of machines either. I'm not sure at this point there is any dead matter or dumb machines, and we should just remember that when we say some... Nothing is a computer, nothing is a Turing machine. There are just times at which you would like to take a formal model of computation or of Turing machines or whatever and apply it to some system, and we all need to remember that when you do that, you do not capture the... Like, our formal models never capture the whole thing. And we thought that was negligible for, quote-unquote, "machines." It is absolutely not negligible, and it has massive implications for how we do engineering, how we do AI, and so on. We could talk about that for a long time. And so we now have an actual research program on what you might consider inspiration. Inspiration is when you get patterns, useful ideas, patterns, symphonies, whatever, theorems like Ramanujan got, all these things from some space that you did not put in the effort to meticulously create step by step, neither create, you know, neither engineered, evolved or learned.

And so we can now do this with very simple systems by creating robots that are powered by mathematical objects, but also by some other things inc... You know, so mathematical objects are signals from bacteria, cosmic microwave background, believe it or else. They can do interesting things when you give them bodies, when you give them robotic bodies. So yes, we are the beneficiaries of ingressing patterns that we don't pay for in the conventional model of physics and computation, but actually even dumb machines get this too in a small way. And also, I think we actually are patterns. We're not just impinged upon by patterns.

So I'm just gonna summarize here, and then I'm done. Basically, I think these patterns of form, of behavior, and so on are ubiquitous. Genetics and emergence is absolutely insufficient for the life sciences of the future, but also for the ethics and engineering of the future. Emergence is not going to do the trick. And the free lunches are, they're most obvious in biologicals. They're the most amazing in biologicals, but you can't really quantify or prove anything in biologicals. They're too complex. But they can be quantified in minimal computational systems. And so that's why I'm not gonna take the time to go through all of this stuff. But that's why we have a research program. You know, this is not just philosophy. I have six people working on all this stuff in figuring out what does that space offer you. Does it offer you static patterns like the digits of E? Yes, but also dynamic behavioral policies, which is what allow our robots with no controller whatsoever, just the patterns, to do tasks.

What's in the space? Why does it... do these patterns inhabit certain interfaces, not others, and so on. All of these are now tractable. We're actually doing experiments. There are a couple papers out, many more coming this summer and fall. And this is critical because right now we live mostly with biologicals, but all of these kinds of things, pretty much any combination of evolved material, engineered material, and software, is some kind of a possible embodied agent that is going to host these patterns, and there are so many different kinds of minds, you know, cyborgs and hybrots and humans with altered... I mean, every imaginable being is going to be here with us. We need to figure out how to do an ethical symbiosis with them. We need to learn to live with creatures that are not like us. And this machine life thing is not going to do it. It is, I think, factually wrong as a matter of empirical data. And I think this, you know, classic picture of Adam naming the animals in the Garden of Eden is going to look much more like this. And the idea is that it's going to be very weird and very diverse, and we're going to be able to understand ourselves a lot better if we embrace the research program of diverse intelligence.

So I will stop here just to thank the people who did all of the hard work that I showed you today. Lots of amazing collaborations. I have to do three disclosures, no, four actually, of companies that have licensed some of the intellectual property that comes out of these kinds of things. And, you know, I always like to thank the model systems because they do all the hard work in teaching us about these things. So I will stop here.

Thank you for reading.

More lectures

You can find more of my lectures here.

Follow my work

Twitter • Blog • The Levin Lab

Want one for your lecture?

Want something like this for your own talk? Reach out to Adi at adi@aipodcast.ing.