

A LECTURE COMPANION

“Unconventional Selves - a diverse intelligence perspective on consciousness” by Michael Levin

Michael Levin

Recorded on July 7, 2025

About this document

This document is a companion to the recorded lecture "*Unconventional Selves - a diverse intelligence perspective on consciousness*" by Michael Levin, recorded on July 7, 2025. You can watch the original lecture or listen in your favorite podcast feeds — all links are on the page [here](#).

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

"Unconventional selves in novel spaces, scales, and embodiments - a diverse intelligence perspective on a continuum of cognition and consciousness" is a ~1 hour 12 min talk I gave at a conference on consciousness (<https://www.hardproblem.it/>), as it relates to our work on diverse intelligence and unconventional cognition.

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.



Thank you to the organizers for allowing me to participate, even though I can't be there in person. I would love to be there.

What I'm going to talk about today are some ideas around embodiments of diverse kinds of minds and maybe the implications for consciousness. I myself do not work specifically on consciousness, but many of the things we do have implications for it. Today I'm going to go out on a few limbs and try to introduce some data to you that may be relevant.

All the primary papers, the software, the data sets, everything is downloadable here at this laboratory website. This is my own personal blog where I talk about what some of these things mean.

I Do Not Claim:

- to have a new theory of consciousness, or
- to have definitive data supporting one specific ToC

I Do Claim:

- Not linguistics or philosophy: use of cognitive frameworks outside brain leads to new discoveries, capabilities = empirical fertility
- If we use cellular mechanisms, problem-solving behavior, and evolutionary relationships as evidence of consciousness in addressing the problem of other minds, then for the same reasons we associate consciousness with complex brains, we need to
- **consider seriously the possibility of consciousness in many body structures.**
- AI and many ethical problems will not be resolved with a restricted focus on "human", 3D space for embodiment, and "brain" as privileged



What I want to be clear today, especially for this audience, is that I do not claim to have a new theory of consciousness, nor am I going to try to support one specific theory of consciousness by data. But here are the things that I am going to claim.

The use of cognitive frameworks, mentalistic terms and all of the tools, both practical and conceptual of cognitive and behavioral science, are highly applicable outside the brain. I do not think neuroscience is about neurons at all. I think a lot of what we have gained from behavioral and neuroscience can be used far outside the brain. This is not a linguistic, philosophical, or poetic kind of claim. Over the last 25 years, being able to do this has led to many new discoveries and new capabilities, both in our lab and in other labs. My defense for this kind of weird idea is simply empirical fertility.

I think that if we use any of the following criteria for addressing the problem of other minds—cellular mechanisms that underlie cognition in the brain, problem-solving behavior, evolutionary relationships, or even metrics such as phi and causal emergence—if you use any of those things as evidence of consciousness in addressing the problem of other minds, then for those exact same reasons, you need to consider very seriously the possibility of consciousness in many body structures.

I personally would go significantly farther than that, but for today, we're going to focus on biologicals mostly, with a little bit at the end of some other things. I think those criteria are telling us that we have to take this very seriously. I think that outstanding problems of AI and ethics and many other things are not going to be resolved if we insist on maintaining a restricted focus on humans, on three-dimensional space as the only kind of arena for embodiment, and on brains as privileged. I think all three of those things have to be widened quite significantly.

What I'm going to do today comes in four parts: a little bit about my framework, then I'm going to show you what I think is a really interesting model system for these kinds

of studies, which is the collective intelligence of morphogenesis. We'll talk about some new creatures that have never existed on Earth before, and then I'll make a couple of conclusions and say some fairly outlandish things.

Slide 3 of 44 · Watch at [3:19](#)

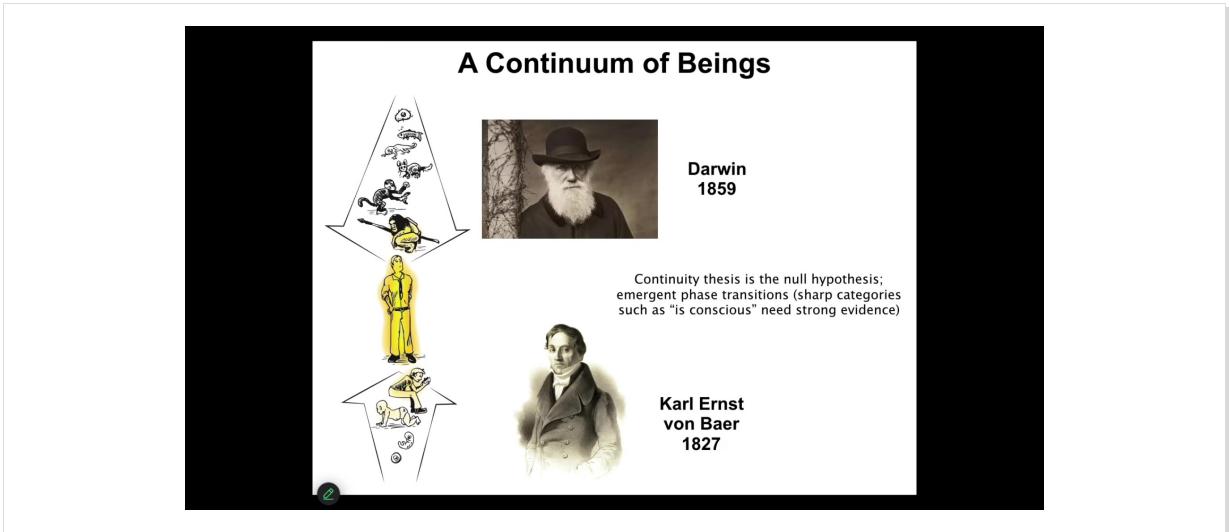
Beyond Static, Discrete Natural Kinds



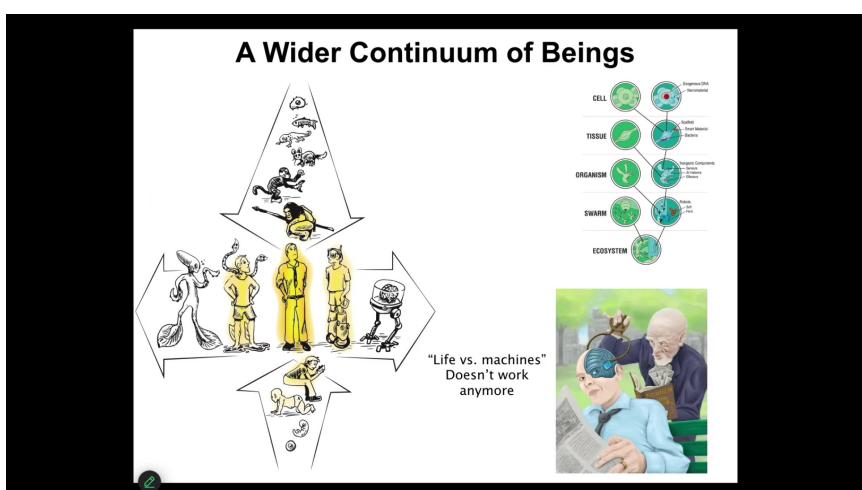
The first thing I want to point out is that I think a lot of discussion in the philosophy of mind and in this field in general has been driven by this kind of picture. This is a famous piece of art called "Adam Names the Animals in the Garden of Eden." And there's a couple of things here that I think are deeply wrong with this picture, but also something very deep and interesting.

The thing that's wrong with it is that it gives this idea, which I think a lot of people have internalized, that there are discrete natural kinds. There are different kinds of animals. We know the differences between them. And so here's Adam. He's different from all the others. And I think this needs to go for reasons that I'll show you.

But what I think is really important here is that in this ancient biblical story, it was on Adam to name the animals. God couldn't do it, the angels couldn't do it, and it had to be Adam that names the animals. In these kinds of mystical traditions, naming something means discovering or even somehow bringing forth the true inner nature of something. By giving it a name, it means you've understood its true inner nature. This, I think, is profound because we are going to have to understand the inner nature of a number of other minds that are very different than ours.



The first thing I would like to claim is that ever since an understanding of evolution and an understanding of embryonic development, meaning that we arise through a set of slow and gradual changes from a single cell on these time scales, the continuity thesis should be the null hypothesis. That is, it's not the case that these categories that we all like to think about in terms of the conscious and the cognitive are binary. When people say, why do you think that they are continuous as opposed to categorical, the onus is on those who do think that these are sharp transitions and categories to show why that makes sense. I think the background assumption should be that it's continuous and gradual. And if you think there are phase transitions, you have to say what those are and how they exist. So I take continuity to be the null hypothesis for all of this.

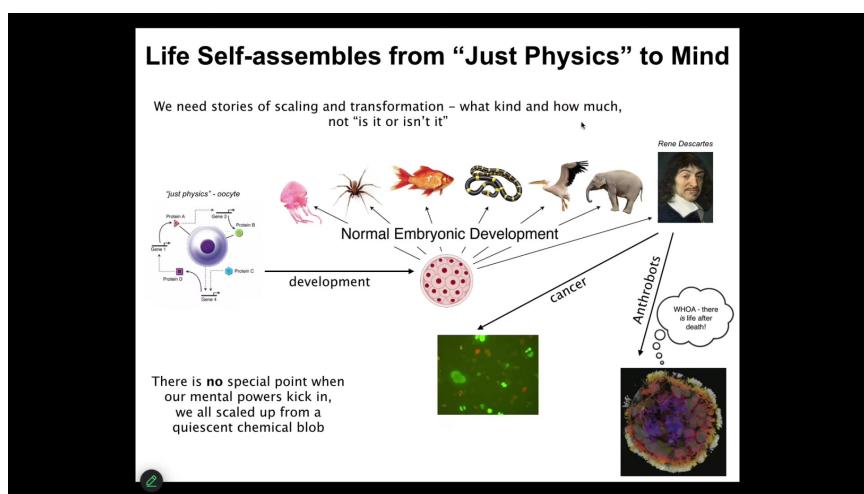


Now, we are at the center of not just this continuum, but actually another continuum where both through biological changes and technological changes, the notion of a human, which figures very prominently in discussions of philosophy of mind, which they typically mean the mature, modern adult human here with his agential glow of responsibility and moral worth.

It is no longer going to work to talk about humans versus machines or living things versus machines, because all of this forms a very tight continuous spectrum. When we have things like this, it is not going to be feasible to start issuing proof of humanity certificates to figure out if he's got more or less than 50% of engineered components, because life is highly interoperable at every level of organization; you can introduce the machine-like parts. This is another distinction that I think is going to disappear.

I've been working on a framework whose job it is to enable us to recognize, create, and ethically relate to truly diverse intelligences. This means not just the primates and birds, maybe a whale or an octopus—the kinds of things that people often think about. But also on the exact same scale, I want to be thinking about all kinds of weird creatures such as colonies and swarms, synthetic engineered new life forms, AIs, whether purely software or with robotic embodiment, potentially someday exobiological agents, and some very exotic things that I won't have time to talk about today, including patterns within media and some even stranger things that I've talked about in other places.

What I'd like to do is to develop a framework that allows us to think about what do all of these things have in common? Very critically, it has to move experimental work forward. It has to give us new capabilities at the bench, and it has to enable better ethical frameworks. I'm not the first person to try for something like this. Here's Rosen, Luth, Weiner, and Bigelow, who tried a cybernetic scale all the way from passive matter to human metacognition. But this is the thing that I've been trying to flesh out with newer data that these guys didn't have access to. The details are in this paper.



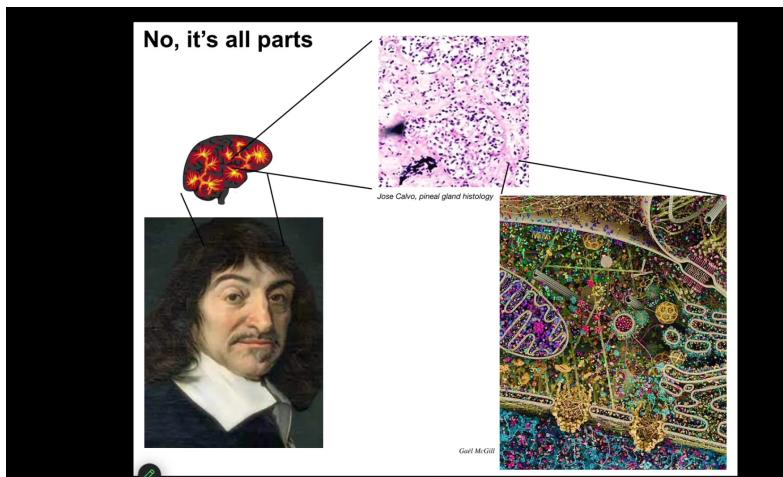
Let's just talk about the biological material that underlies examples of cognition and consciousness of which we are aware. This is where it comes from.

We all start life as a single cell. This thing, the unfertilized oocyte that people will look at and say this is subject to the laws of physics and chemistry. And then gradually it becomes something that is subject to the rules of behavioral science or even psychoanalysis and things like that. Developmental biology tells you quite clearly that there is no magical place to draw a bright line where you can say, okay, you were physics before, but now you're a mind. What we actually need are models of scaling and transformation. What happened to the competencies of this material in order to get to this kind of thing? And the question should be what kind and how much of cognition does it have, not binary categories—is it or isn't it? There's a few interesting things that can happen even after that.

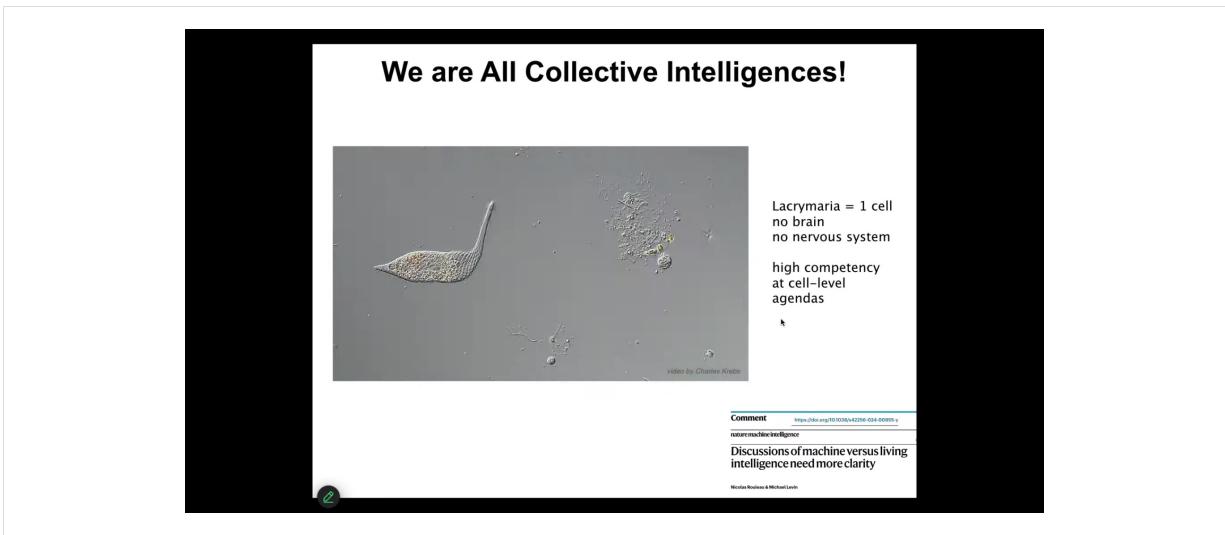


Okay, but at least some people will say, sure, you can talk about ants and beehives as a collective intelligence, but maybe that's just metaphorical talk.

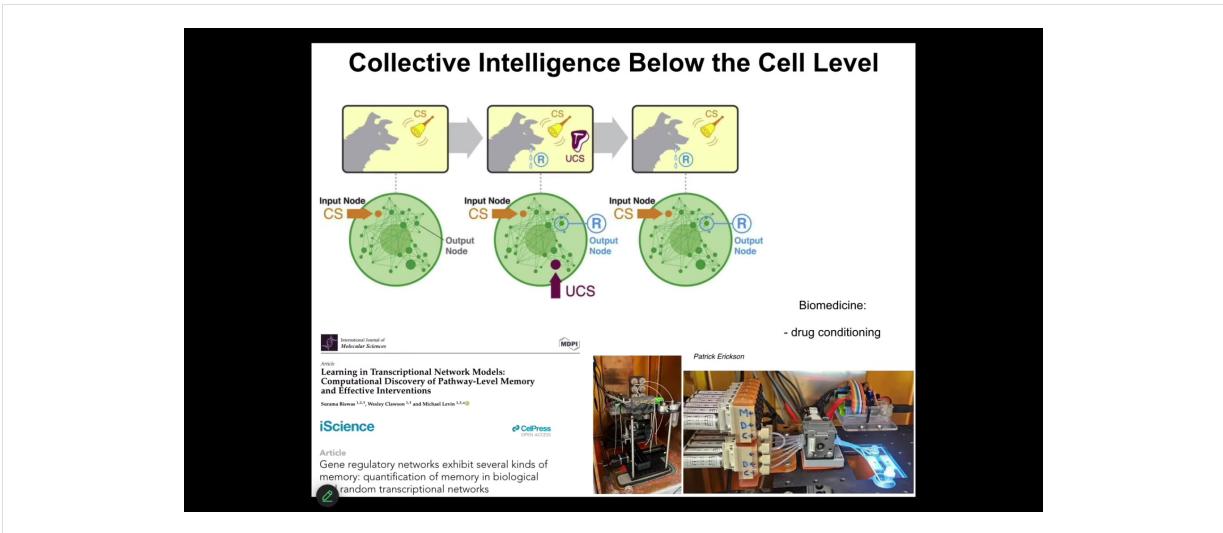
Surely we are a true unified intelligence. We have this nice centralized brain, and the human experience of being a unified mind is underwritten by this kind of singular organ.



Descartes really liked the pineal gland because it's only one of them, and he felt that was a good place for the unified human experience. But he didn't have access to good microscopy. If he had looked in the pineal gland, he would have found all of these things, tons of cells. Inside each one of those things, he would have found all of this stuff. Nothing is unitary. We are all made of parts.

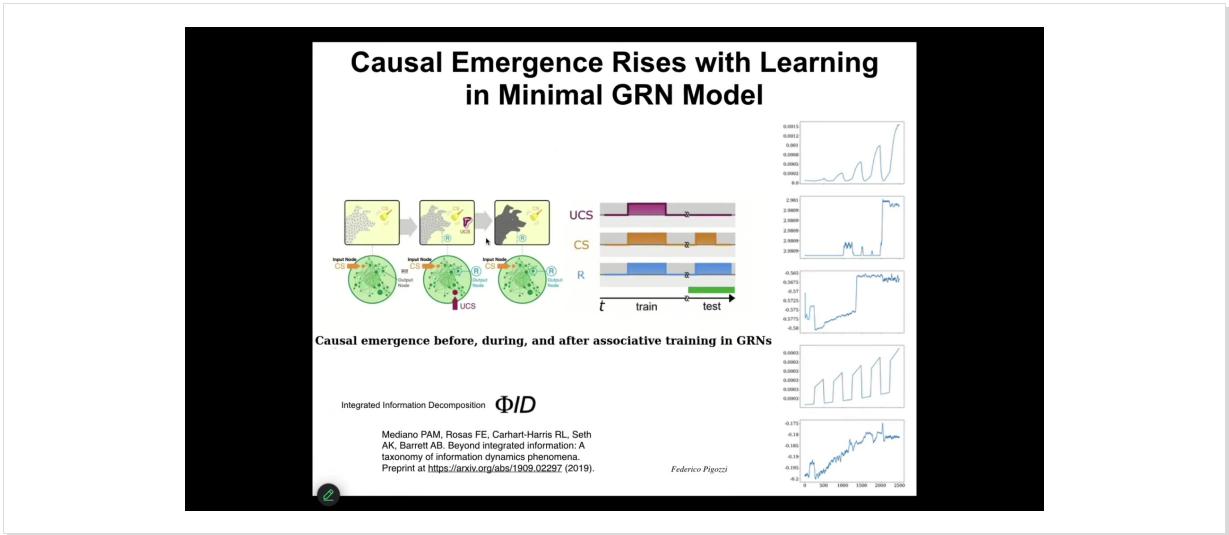


I would claim that all intelligence is collective intelligence. I'm not aware of any agent that is simply not divisible into other parts, that does what it needs to do in a monadic form. And so this is the kind of thing that we are all made of. So this is a free-living Lacrymaria. It's a single cell, no brain, no nervous system, and handles all of its needs in ways that make soft matter physicists and roboticists very jealous, the degree of control it has over its body as it pursues its tiny local goals. If we want to say things about our own consciousness and that of even weirder kinds of novel engineered constructs, we need to be able to say something about this kind of creature and what happened to it when it joined into networks to enable it to become the kind of thing that we are. Nick Rillot and I tried to make those questions very sharp in the form of a flowchart in this paper. Now, the interesting thing is that while this little guy has all kinds of behavioral properties, it is even made of material that already has some of those properties.

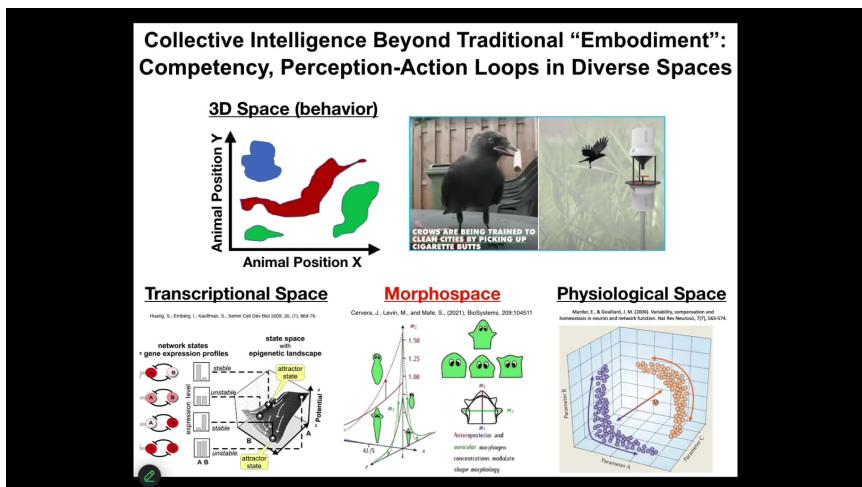


For example, it turns out that molecular networks — never mind cells, the nucleus, the synapses, never mind any of that — just molecular networks, represented by a small set of ordinary differential equations that describe how each chemical turns other chemicals on and off. It's just a little network. Even these things are capable of six different kinds of learning, including Pavlovian conditioning. We're taking advantage of this in the lab for all kinds of medical purposes, such as drug conditioning. You can train the material inside the cells.

When we think about ingredients to cognition and the kinds of things that go along with significant consciousness, such as memory and learning, we have to understand this did not wait for brains and synapses to be formed. This was in many ways a free gift from mathematics because the properties of these networks actually enable the ability to learn from experience in precisely the ways that are recognized by behavioral science in terms of habituation, sensitization, anticipation, associative conditioning. Those things emerge from the mathematics of these networks. They are not dependent on evolution of the substrate.



We discovered quite recently, and this paper is going to come out next week, that we found that if you compute causal emergence, and in particular we're using this one called Phi-D. If you compute causal emergence of these networks during the training process, what you find, not in all of them, but in many of them, is that the process of being trained causes their FID to go up; that is, instances of training and then testing reify the network as being more than the sum of its parts. This idea of being able to use causal emergence metrics to understand how much causality is in the higher level system versus its parts, again, is not a brain-specific thing. You can find these phenomena very early on. I need to move on.

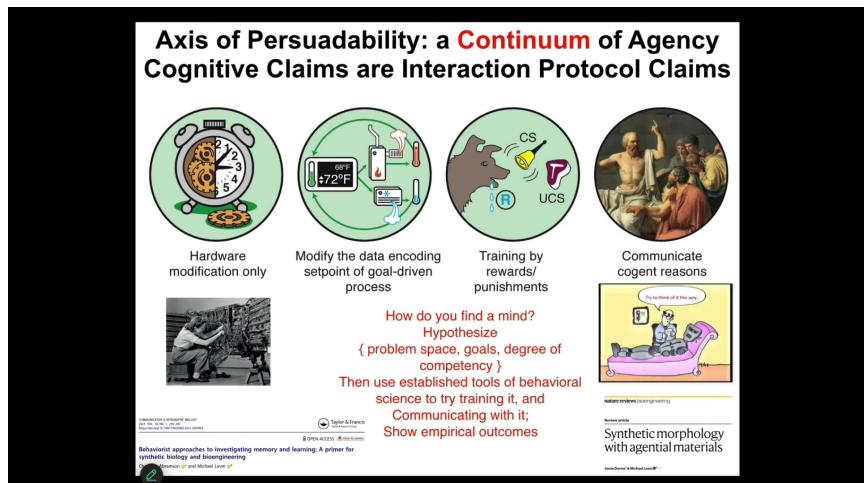


But one of the most important things to realize about these kinds of systems is that they remind us that as humans, given our own evolutionary history, we are obsessed with the three-dimensional world. In other words, when we say behavior, or more importantly, when we say embodiment, what we typically mean are medium-sized objects moving through three-dimensional space at medium speeds. Those are the things we are primed to recognize as intelligent. This is a problem for this field because we find it very difficult to visualize things that biology has been doing long before brains and motility came on the scene. Biology has been navigating all kinds of other problem spaces.

Embodiment is really critical, but embodiment is not what we think it is. It is not just the ability to move around through three-dimensional space. Because this kind of loop, this perception–decision–action loop, happens in all kinds of spaces.

Living things navigate the space of gene expression states. They navigate the space of physiological states. What we'll talk about the most today is anatomical morphospace. They navigate the space of possible shapes. In all of these cases, you can, it turns out, use the tools of cognitive and behavioral science to understand how they navigate these spaces and the degree of ingenuity they deploy, and if we had some sort of inner sense where we could directly sense, the way we do with taste, either gene expression states or physiological parameters in our blood, I think we would have no trouble realizing that our liver and our kidneys are these amazing symbionts that intelligently navigate these other spaces to keep us alive every day.

So we need to widen this notion of what embodiment really means and be broader about what kind of spaces intelligence exerts itself in.

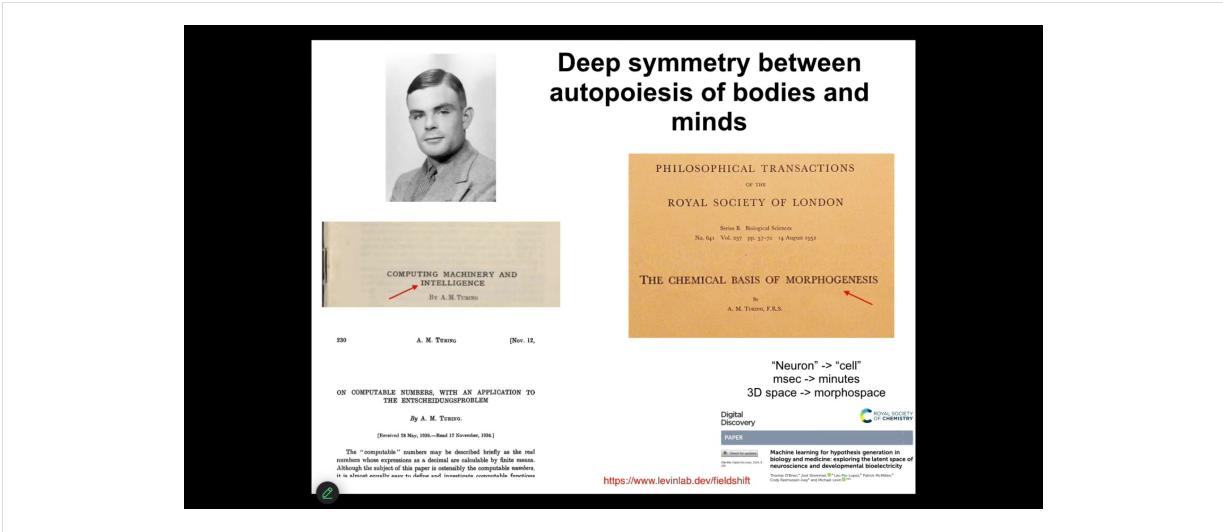


Now, just the last thing I'll say about my philosophical cards on the table is that what I think cognitive claims are basically interaction protocol claims.

In other words, when somebody says this system has a particular degree of cognition, intelligence, I don't think we're talking about what it really is in its essence as some ground truth that we can argue about for a bit and then eventually settle on the correct answer. I don't think that's what's going on here. I think what you're really saying when you make those claims is that here is a set of tools that I bring to interacting with the system, and those can be hardware modifications. It might be cybernetics and control theory. It might be behavioral science. It might be psychoanalysis, love and friendship, and what happens at this end. But what you're saying is I'm going to bring this bag of tools and then I'm going to interact with the system. Then all of us, as a matter of empirical discovery, can find out how well that worked out for me.

If you assume that cells and tissues, as modern molecular medicine does, are somewhere down here, they're mechanical machines that need to be rewired with genomic editing, you're going to leave a lot on the table because actually testing the applicability of some of these tools makes many discoveries that are not accessible at this level. I think all of these things are empirical questions. I don't think they can be settled from a philosophical armchair. We cannot just decide. We should not be tied to ancient philosophical categories; those kinds of things have to advance with the science. The question simply is this. If we assume a continuum hypothesis, then we are open to testing tools from other disciplines, typically reserved for minds and especially humans in other cases, to see what kind of empirical purchase we get on this.

At this point, I'm going to talk about some very specific examples using a model system. The model system is the collective intelligence of cells making decisions, learning, and meeting goals, and failing to meet other goals in the process we call morphogenesis.



I think Turing was on to this. He didn't write about this directly, to my knowledge, but he saw clearly a deep symmetry between the self-assembly of bodies and the self-assembly of minds, because he not only talked a lot about different embodiments of mind and reprogrammability and intelligence in diverse implementations, but he also wrote this really interesting paper, "The Chemical Basis of Morphogenesis." Why would somebody interested in intelligence be looking at chemicals in early development? He realized this is fundamentally the same problem.

I used to have my students do this by hand and then we developed a tool to do this that you can all play with. You take a paper in neuroscience, you throw it into Microsoft Word, and you do a find-and-replace. Anywhere it says neuron, you say cell. Anywhere it says milliseconds, you say minutes. Anywhere it says something about three-dimensional space, you say anatomical morphospace. You usually end up with a very interesting developmental biology paper. The symmetries are quite striking, and you can try this yourself, and it's described here.

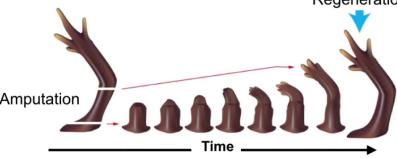
Goal-directed Navigation of Problem Space:

- get to the same outcome (maintain set point)
 - despite perturbations
 - from diverse starting positions
 - via different paths
 - stop when goal is achieved



Amputation

Anatomical homeostasis:



Time

Perfect Regeneration

it stops when the correct large-scale setpoint (target morphology) has been reached

What I'm going to talk about next is to give you some specific examples and then show you some discoveries that this idea has led to. The first idea is that morphogenesis is not hardwired. It is not the case that you start with an egg, then you follow some rules that the DNA laid out, and then by emergence, you get a complex thing, this feed-forward process. What does happen is a goal-directed navigation of problem space.

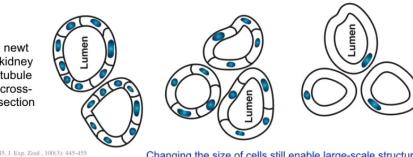
Here's an example. This is an axolotl. Axolotls are very good at anatomical homeostasis, which means that if you amputate some part of their body, for example, the limb, you can amputate anywhere you want, the cells can immediately detect that they've deviated from their correct position in anatomical space. Anatomical space has many dimensions. The cells will very quickly try to reduce that error. And then the most amazing thing about regeneration happens, then they stop. How do they know when to stop? Well, they stop when the correct structure has been completed. This is an example of anatomical homeostasis. The system tries to reduce the delta between where it is now and where it wants to be. This means that it has to store a set point, which I'll show you exactly how that's stored. It has to store a memory of what is the correct shape. And the other thing is that you can do this again and again, and it will keep doing it about five or six times. Eventually it stops. Whether that's a true instance of learning, we don't know yet, but it may well be that eventually it just gives up and no longer wants to do this anymore. One important thing is that this whole process of anatomical homeostasis is not about repairing damage. It's much more interesting than that. If we take a tail, and this is old work from the 50s, if you take a tail and surgically attach it to the flank here of the salamander, between the two limbs, what will happen is that over time it will get remodeled into a limb. Now think about what that means. The cells sitting up here at the tip of the tail are fine. They're tail tip cells sitting at the end of the tail. There's nothing locally wrong with them, but they turn

into fingers. Because what the system is doing is not only does it have a stored set point of what a correct layout should be, it should have a limb in the middle, not a tail, but it's got this interesting top-down causation where all the molecular steps that need to happen in order to turn tail structures into limb structures are in the service of a larger scale goal, of which they probably know nothing about, but the system as a whole filters down this very high level design spec of having a limb instead of a tail, filters that down to the behavior of cells and molecules, exactly as happens in cognitive systems, where your very abstract social and financial goals have to be transduced down into moving ions across your muscle membrane so that you can move around and execute on those goals. So this idea of local order obeying a global plan is very interesting. If you ask why the cells are doing what they do, you run into the same kind of thing as with explaining instances of behavior. You can try to give molecular explanations, well, it's because this molecule signals to that receptor and so on. But that's not the deeper answer as to why it's happening. The real answer as to why it's happening has to do with the target state that it's trying to achieve.

Slide 16 of 44 · Watch at [21:54](#)

Creative Problem-Solving by Cell Groups

- get to the same outcome
 - despite perturbations (external **and** internal)
 - from diverse starting positions
 - via **different molecular mechanisms!**



newt kidney tubule cross-section

Franklin, 1945, J. Exp. Zool., 100(3): 445-459

Changing the size of cells still enable large-scale structures to form, even if they have to utilize different molecular mechanisms = top-down causation

INTERFACE
of communication

Top-down models in biology: explanation and control of complex living systems above the molecular level

Integrate

Science Research and Application

Volume 1, No. 1, 2013

INTEGRATIVE BIOLOGY

Re-remembering the body: applications of computational neuroscience to the top-down causation of the organization of cells and entire complex organs!

PERSPECTIVE

Volume 1, No. 1, 2013

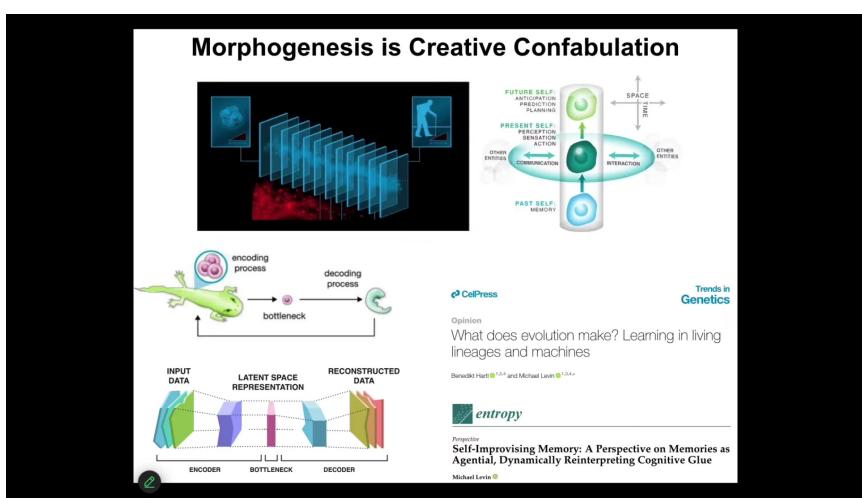
To take it even one step further, and to link it further to problem solving and intelligence, that kind of remodeling has some amazing examples. This is one of my favorites. This is a cross-section through a kidney tubule. You'll see 8 to 10 cells are making this kind of structure. One thing you can do is make polyploid newts that have extra copies of their chromosomes. If you do that, the cells get bigger. As the cells get

bigger, you find that it still makes an animal of exactly the same size. How can that be? Because it's scaled the number of cells to the new cell size.

If you make truly gigantic cells, and these are like 6N polyploid newts, only one cell will bend around itself, abandoning the molecular mechanisms used here, cell-to-cell communication, and use cytoskeletal bending to solve the problem and create the same structure. That same journey in anatomical space, the system is using different tools in its genetically specified set of affordances to get to where it needs to go.

This is a standard component of many IQ tests. Here's a bunch of objects that you have to use to solve this problem you haven't seen before. Think about the plasticity here. If you're a newt coming into the world, never mind uncertainty about your environment, you can't even trust your own parts. You don't know how many copies of your genetic material you're going to have. You don't know how big your cells are going to be. You have to use whatever tools you have creatively to solve the problem that you have, which is to get from point A to point B, from where you are in the anatomical space as an egg to where you're going to be as a good newt. There are many examples; we could do this for hours. There are just remarkable examples of this.

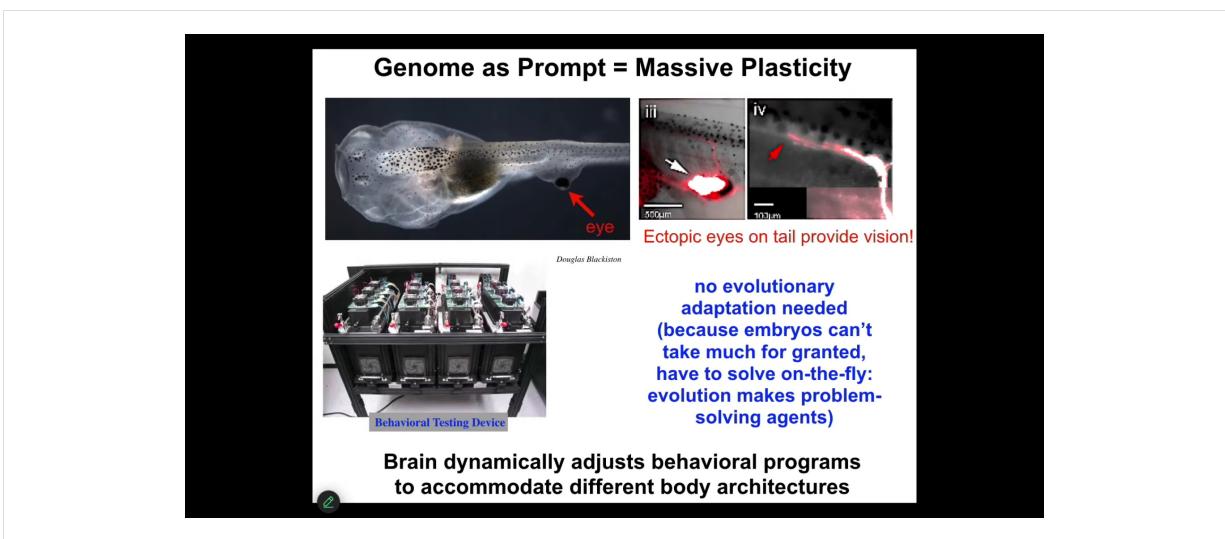
Slide 17 of 44 · Watch at [23:34](#)



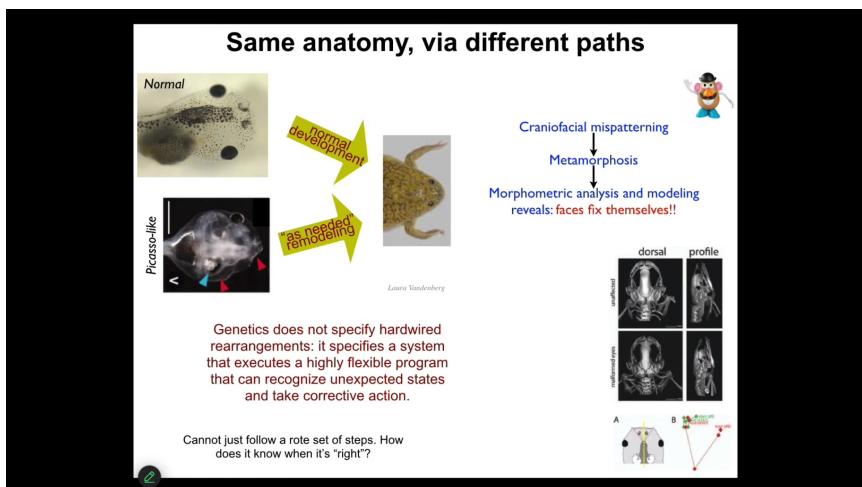
So, morphogenesis does exactly the kind of creative confabulation that I think goes on in cognition, where at any given time slice, an agent doesn't have access to the past. What you have access to are the memory engrams, the messages left by your past self. Your memories have to be interpreted. They don't need to be taken literally. They need

to be interpreted. And there's some examples that show that they can't be taken literally. For example, memories that pass from a caterpillar to a butterfly, and they get remapped into a novel embodiment. And this process of continuously reinterpreting and telling the most adaptive, forward-looking story you can is exactly the same for cognition and for morphogenesis, where you have this now moment and everything that you've experienced before, be it behavioral memories or genetic information from your lineage, is squeezed down into a very sparse representation, the genome or memory engrams, and then has to be creatively decoded. And all of this is described here. These parallels are really deep, I think. And in particular, this idea that morphogenesis, just like behavior, is a creative storytelling using whatever information you have, gives biology amazing plasticity.

Slide 18 of 44 · Watch at [24:50](#)



Just like for behavior, morphogenesis has it too. Here's an example. Here's a tadpole. Here's the mouth. Here are the nostrils, here's the brain, the gut. What you'll notice is that we prevented the primary eyes from forming, but we put an eye on its tail. Turns out these animals can see perfectly well. How do we know? We built a device that trains them on visual cues. These eyes form; they make an optic nerve. It does not go to the brain. It goes sometimes to the spinal cord, sometimes to the gut, sometimes nowhere at all. These animals can see. It does not take new rounds of mutation, selection, adaptation. This works out-of-the-box. You've changed drastically the sensory motor architecture of this animal. I think that's because of this incredible plasticity that the tissue has, that it interprets its local situation to tell the best story that it can. It is not committed to whatever its ancestors were doing with that information.

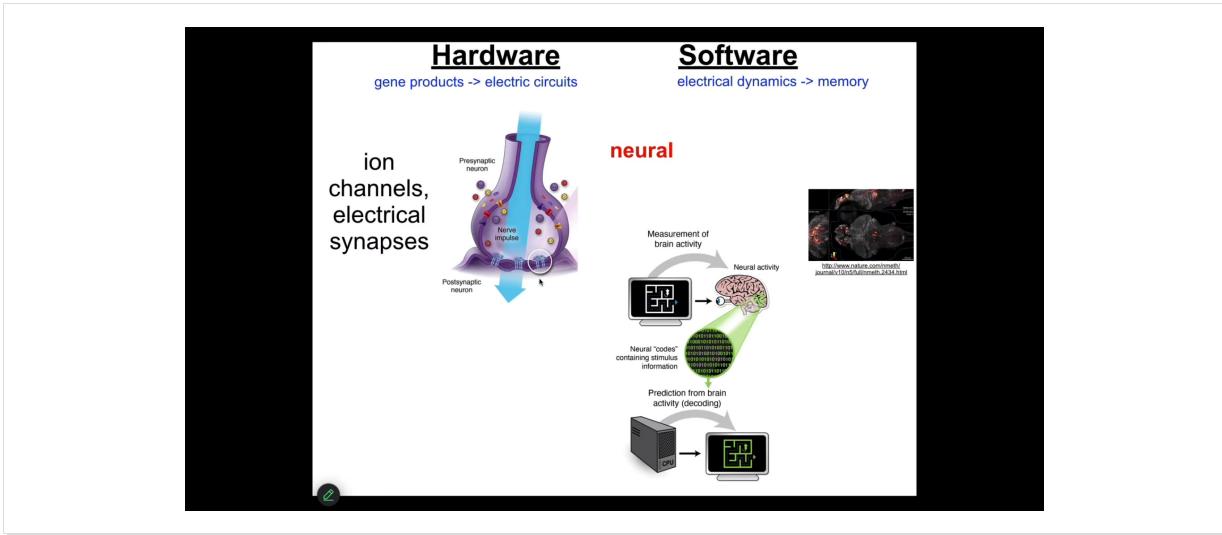


Before I show you how all of this works, I want to give you one simple example to remind us of the idea of the importance of storing pattern memories. This is something we discovered a few years ago: these tadpoles, in order to become a frog, have to rearrange their face. This is metamorphosis; the eyes have to move, the nostrils, the mouth, everything has to move.

The standard assumption in the field was that this is a hardwired process, that the genetics somehow tells every structure which way to go and how much. As I said, you can't just assume these things. You have to do perturbative experiments to find out how much competency the system has at problem solving.

What did we do? We created these so-called Picasso tadpoles. We scrambled all the craniofacial organs. The eyes are on top of the head, the mouth is off to the side. It's like a Mr. Potato Head after you scramble all the parts.

It turns out that those tadpoles make largely perfectly normal frogs, because the genetics does not give you a hardwired set of rearrangements. What it actually gives you is a system that can execute error minimization, and all of these things will move through novel paths to get to where they're going, despite the fact that they're starting off from weird positions, and sometimes they go too far and they have to come back. This tells us several things. The system has to have a mechanism for knowing what is the correct position. How do you know you've gotten to the right place? It has to have mechanisms for actually computing the means-ends analysis of how do I reduce error? What am I going to do to reduce this error?



Well over 25 years ago now, what I did was to simply say, what is the one example that we know of where groups of cells store goal states and try to execute towards those goal states. And that is the brain. I don't need to go into all of this, but just to say that there's ion channel and gap junction hardware that underlies this electrical network, also neurotransmitters and many other things.

There's this project of neural decoding, the idea that if we could just read out the electrophysiology and if we knew how to decode it, we would be able to read out the cognitive content, the mental content of the animal. We would know the goals, the preferences, the competencies, if only we knew how to read out this kind of electrical activity.

So it turns out that this idea—that the ability of this kind of hardware to do these amazing things—is absolutely ancient. Evolution discovered the benefits of electrical networks for integrating and processing information long before neurons, in fact, long before multicellularity showed up. Around the time of bacterial biofilms, bioelectric networks were already being used to create proto-individuals that were more than the sum of their parts.

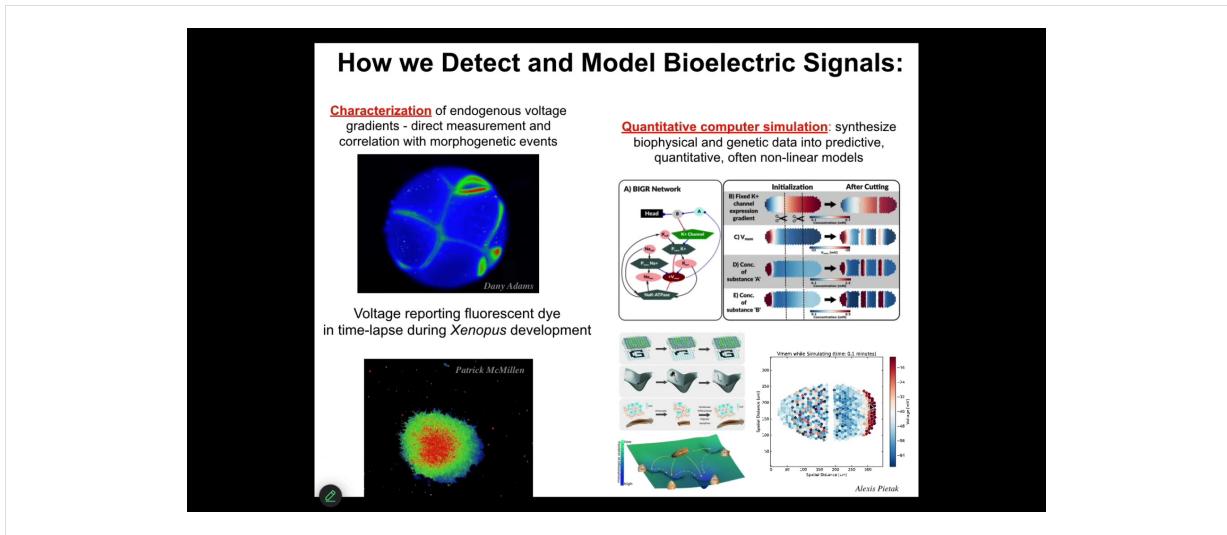
Every cell in your body has ion channels. Most cells have these electrical synapses known as gap junctions. Lots of neurotransmitters are part of this network outside the brain.

This is critical to my claim that if you think consciousness is associated with brains because of the amazing hardware that the brain has, that hardware is ancient. It long predates neurons and brains. It is present throughout biological bodies.

That raises the interesting question: before there were brains and before there's a brain in the embryo, what do these networks think about? In the embryo, what do they think about?

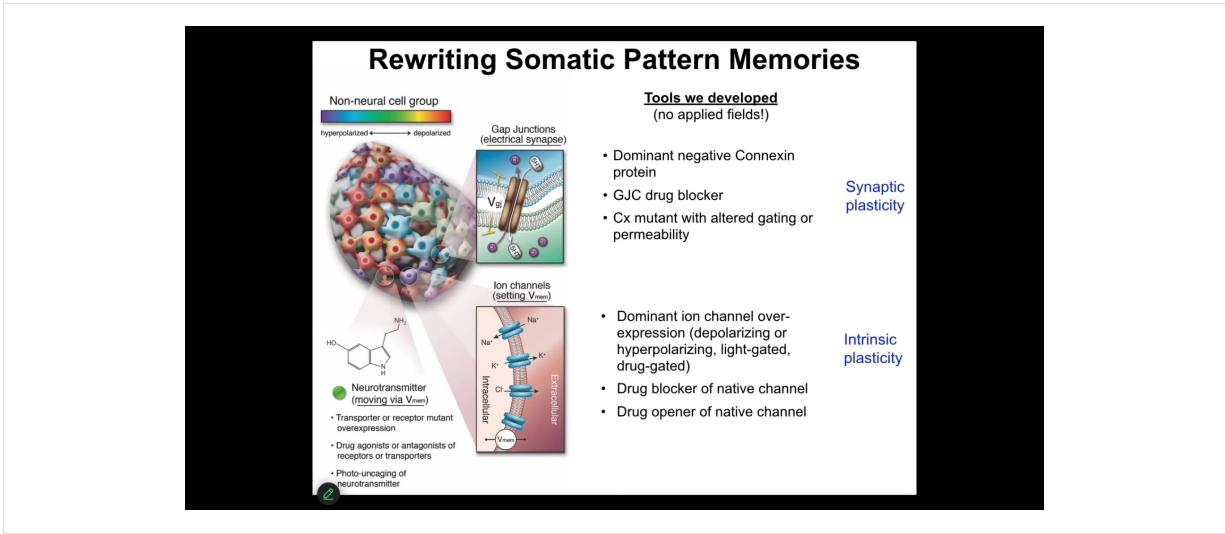
We undertook the process of stealing as many tools and concepts from neuroscience to ask very similar questions. Could we do a kind of decoding here to ask in these embryonic examples what these networks are doing?

Slide 21 of 44 · Watch at [29:46](#)



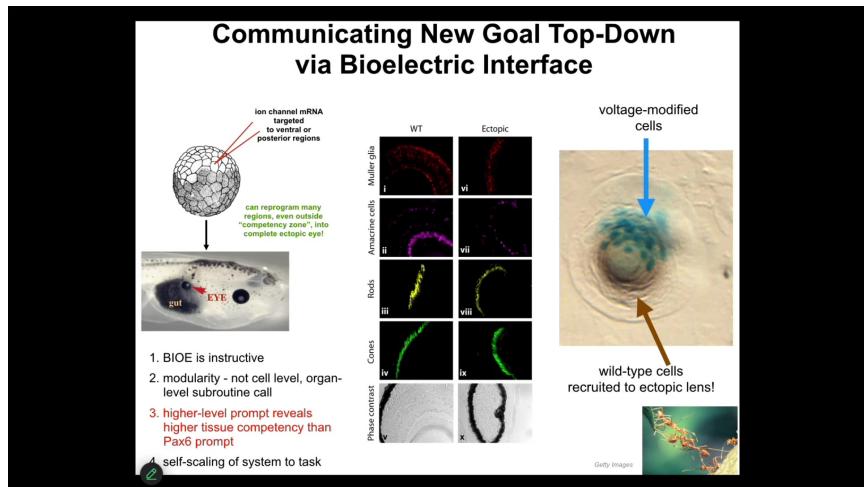
We developed a number of tools. First of all, voltage-sensitive fluorescent dyes and genetically encoded reporters, and these are not models, this is real data. This is an early frog embryo and these are some explanted cells to be able to read out all the different voltage patterns and all the conversations that these cells are having with each other, and then use all kinds of modeling integrated with the molecular biology of what ion channels underlie these gradients.

And then tissue-level dynamics of what's happening with the voltage gradients as a function of time. Continue scaling up so that we can integrate that with some ideas in dynamical systems theory and connectionist ideas around networks that store memories and can do pattern completion. Regenerating a limb that I showed you is an example of pattern completion. We can integrate across the levels of organization here. But that's just reading the patterns.



More importantly than that, you want to be able to write patterns because it's not enough to just do observations, you have to do perturbative experiments.

I really think that what we're doing here is not simply trying to micromanage physiological states on some kind of machine. I think we're actually communicating with an unconventional decision-making, memory-forming intelligence here. In order to do that, we use the electrical interface. No magnets, no electromagnetics, no applied fields, no frequencies. What we do is we hack the interface that the cells use to become a larger level entity, the cognitive glue. This is just like in our brain: the electrophysiological processes are the cognitive glue that make us more than a pile of neurons. We can open and close these gap junctions. We can manipulate the synaptic and the intrinsic plasticity of any tissues in the body. Then we can ask what happens when we write new memories of new voltage patterns into the system.



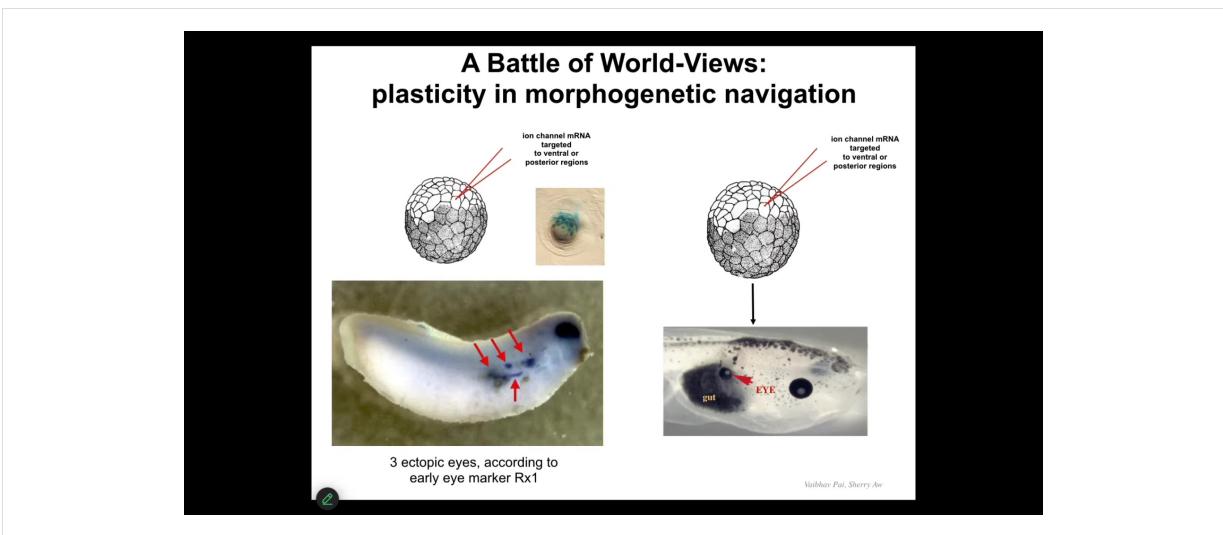
I'll show you one of my favorite examples. We have a particular voltage state that we can induce by injecting RNA encoding certain ion channels. If we inject that RNA into the early embryo, this is early frog embryo, into a bunch of cells that are going to become this gut region, we can set up an eye-specific voltage pattern, and those cells get the message: they make an eye. If you section those eyes, you get the same lens, retina, optic nerve, all the stuff that eyes are supposed to have, so it's a proper eye.

This teaches us a few things. First of all, these bioelectric patterns that we're putting in are instructive for behavior. In other words, we can communicate a message that will guide the way that the collective navigates morphospace. Again, remember the claim. The claim is that groups of cells are a collective intelligence. They use electrical networks, just like our brains do, to be the cognitive glue of that larger scale intelligence. The decisions they make are about movements in anatomical space. Here we can provide a message that causes them to go somewhere else and build an eye instead of a gut. It also tells us that much like when I'm communicating with you now, I don't need to worry about adjusting the synaptic proteins in your brain. I'm giving you a very thin information signal through language, and then I rely on you to adjust the molecular biology and biochemistry in your brain to make it make sense. Same thing here. We're not communicating with the genes. We're not telling the individual stem cells what to do. We're providing a very high-level abstract signal that is actually an arbitrary symbol that, because that's how the cell interprets it, means make an eye here. That is all that we said. The material takes care of the rest.

I want to show you something interesting that happens here, which has interesting parallels to what happens in the brain. If we take a cross-section through one of these eyes, what you'll see is that the blue cells are the ones that we injected. But all this other stuff, we never touched it. Why are they participating? What is happening here is that the cells that we inject are recruiting their neighbor cells, the way that other

collective intelligences like ants and termites recruit their nest mates when they find something too heavy to lift. They recruit them to build this eye. We didn't teach them to do that. They already know how to do that.

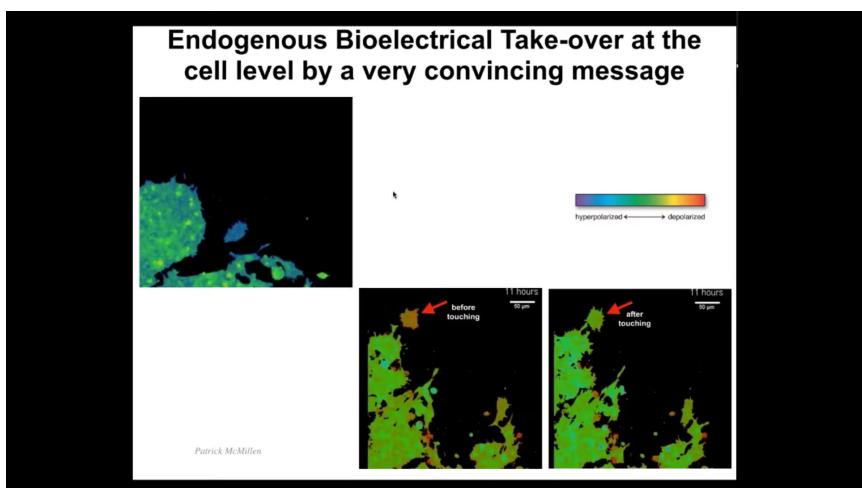
Slide 24 of 44 · Watch at [34:17](#)



But here's something interesting that happens. If we do this, we inject this message and then we look in the early embryo—this is a much earlier stage embryo than this one—what you see is you see four or five nascent ectopic eye fields. These are the cells that got the message and are starting to flirt with the idea of making an eye. They're starting to express early eye markers like Rix1.

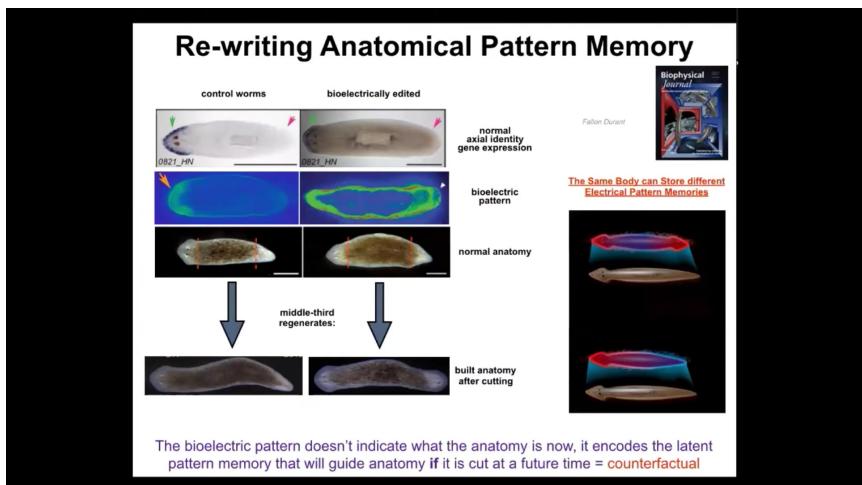
But when you let this embryo grow up, one of the things you will notice is that most of these resolve and don't actually become an eye. That's because what's actually going on here is a battle of patterns for the future. These cells, the ones we injected, are saying to their neighbors, you should be an eye with us. You should help us become an eye.

But there's a cancer suppression mechanism at play here where the cells with abnormal neighbors, the first thing they do is try to tell them to normalize and be like them in terms of voltage. We have a whole oncology program centered around that mechanism where the local cells are saying, no, you shouldn't be an eye. You should be skin or gut. What you have there is a battle of patterns of who is going to dominate the future. Which pattern? Are we really going to give in to this new message and become an eye, or are we going to stick with our priors and be skin or gut? We can actually watch this happen.



Now I want to show you what it looks like at the single cell level for a convincing message to be passed along from one subunit to another. I'm going to play this video in a moment. What you're going to see is this cell, which migrates around here. You can see that the color here indicates the voltage. The voltage here before it encounters this whole mass is different than the rest. It has a more depolarized voltage. But as soon as it touches—this is all it takes, a tiny little touch—it feels the voltage of this cell and that voltage takes over. The cell basically becomes hacked by the bioelectric information that is present in this collective. After that, it becomes the same and you'll see it joins.

Here it is. The colors are unfortunately inverted here, but you can see it's quite different than the rest. There it is. It touches the rest, and now it's got the same voltage, and now it merges with this collective. What you're seeing here is a very convincing bioelectric state taking over this subunit, which then joins a collective with a larger cognitive light cone. This network has greater capacity to pursue morphogenetic goals than single cells do. This is what it looks like in a single cell. But these kinds of pattern memories, these bioelectric pattern memories, really come into their own in the multicellular context.



What I want to show you is how we read and then rewrite the goal states, the anatomical set points for this tissue, again, as an example of a collective intelligence.

The whole point of collective intelligence is that you make individuals that are composed of parts where the higher level individual knows things and can do things and can pursue goals, larger goals, than any of the parts can.

What we see here is a set of experiments that we did over about a decade with these flatworms. These are called planaria. They have a number of features, but one of the features is that if you cut them into pieces, each piece regenerates exactly what's missing, and then it stops.

Every fragment of this body knows exactly what a correct planarian is supposed to look like, and it does what it needs to do to reduce the delta between the injured state and the final goal state, and then it stops.

What you see here is that a single normal planarian will have a molecular marker expression of anterior state up at the head, not at the tail. When you amputate the head and the tail, the middle fragment extremely reliably generates a worm like this, one head, one tail.

This happens every single time correctly. One wonders how it knows how many heads to have and where to put those heads. Because this region of this piece here grows ahead, but so does this one back here; the anterior portion of the tail is also going to grow a head, but the posterior portion of the mid fragment is not. It's going to grow a tail.

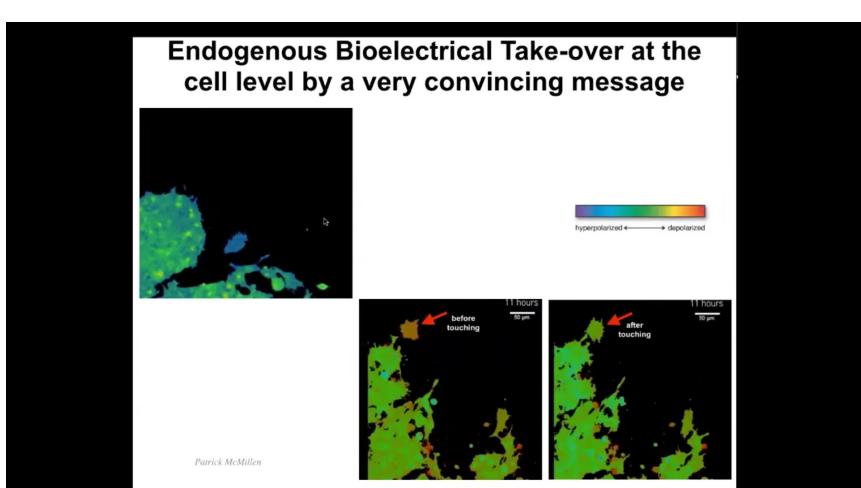
Yet all of these cells were direct neighbors to each other, meaning that the positional information is the same, but the fates are quite different. The anatomical fates are quite different. You see here, this decision is made holistically. It cannot be made locally

based on just the position. It has to be made in the context of what else is here, which way is the wound facing, that kind of information.

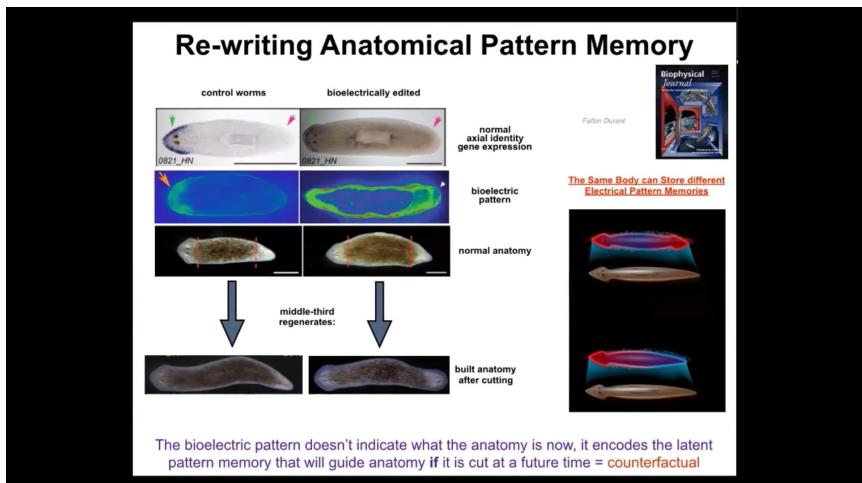
One way that these fragments know how many heads is that they have this bioelectrical pattern. These are voltage-sensitive fluorescent dyes that we use to determine that there's one region here that says one head and one tail. What we were able to do, and the technology is still pretty messy and still being worked out, was to change that pattern by exposing these animals to ionophores, which allow specific ions to enter the cells, changing the voltage in a predictable way.

It only takes three hours. The regeneration process takes about eight days. The voltage change happens in three hours. I just showed you in a single cell it can happen very quickly.

Slide 27 of 44 · Watch at [39:36](#)



In this kind of context, it happens literally in a matter of minutes. You touch and that's it.



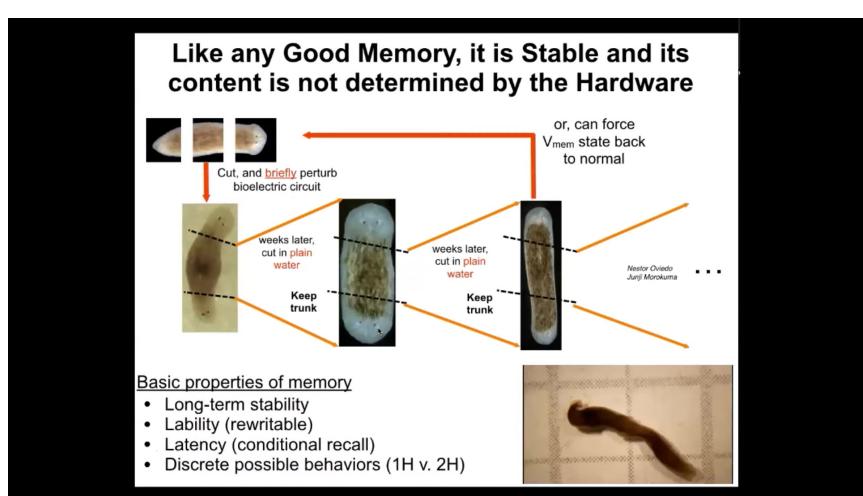
In this context, it takes about 3 hours minimum. When you do this, we can establish a pattern that looks like this. And while this pattern is interpreted by the regenerating system as build one head, build one tail, this pattern is interpreted as build 2 heads.

If we do that to this animal and then amputate the head and the tail, this middle fragment will build 2 heads. This is not Photoshop or AI. These are real animals. What's really critical here is that this bioelectric pattern is not a map of this two-headed animal. You see a two-headed bioelectric state, but that is not a map of this animal. It is a map of this perfectly normal, anatomically normal planarian. It has one head, one tail. If you look at the molecular markers, there they are, anterior markers in the head, no anterior gene expression in the tail.

What you're looking at here is a latent memory. This is a bioelectrical memory of what a correct planarian is supposed to look like, but it's latent. The tissue doesn't do anything with it until it gets injured. Only if it gets injured does it consult this internal map of the goal state, and then it builds what it needs to build here.

You can think of this as a primitive evolutionary ancestor of the brain's amazing ability to do mental time travel — to imagine and remember circumstances that are not true right now. This is a counterfactual memory. It is what I would do if I got injured in the future. I think this is telling us that the same exact mechanisms that are exploited by brains and nervous systems to do memory and prediction are already doing this kind of thing, but in anatomical space, long before brainy kinds of behaviors in three-dimensional space.

This is how we read and rewrite the content of the mind of the collective intelligence, which is using these patterns to help it navigate anatomical space.

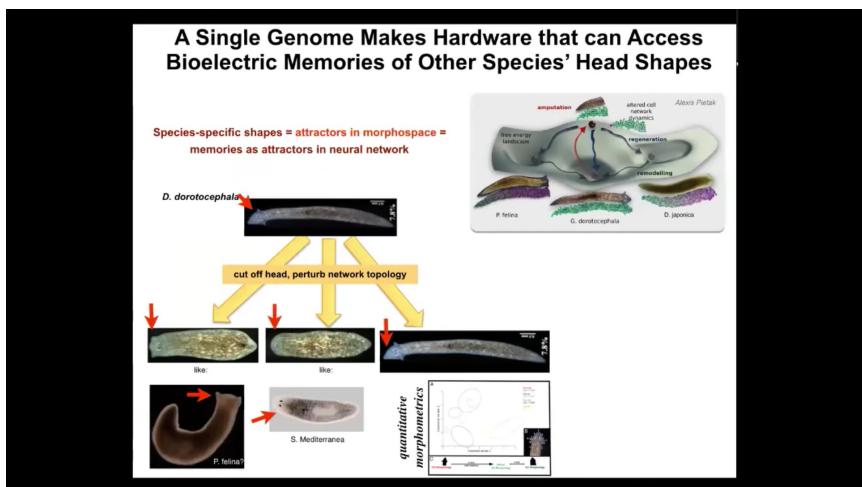


I keep calling it a memory, and one way to show that it really is a memory is to point out that bioelectrical state, once we induce it, holds. It's permanent. When we take these two-headed animals and cut them again and again, as far as we can tell indefinitely, with no more exposure to any drugs, in plain water, we keep cutting them, or we let them fission on their own. The normal reproductive mode of this animal is they tear themselves in half. This propagates. Once you're two-headed, the fragments will be two-headed in perpetuity.

Here you can see these animals are moving around. The first time I showed this at a conference, somebody stood up and said, "Those animals can't exist." I made sure to bring these videos so you can see they absolutely do exist. These are from the 3rd or 4th generation, and you can keep doing this. It's permanent. Until we set it back, and we can reverse the bioelectric memory and set them back to one-headed, this has all the properties of memory. It's long-term stable. But it's rewritable, so it has some lability to it. Once we know how to do it, we can rewrite it. It has latency or conditional recall, meaning that these memories can be latent and not expressed until certain conditions are met. For example, injury stimulus triggers the recall of that memory. There are some discrete behaviors, one head, two head, and so on. However, this system is not just for head number.

A single planarian body here can store at least two different internal representations of what a correct planarian looks like, probably more, but we've nailed down these two.

It's actually not about head number at all. There's much more here.



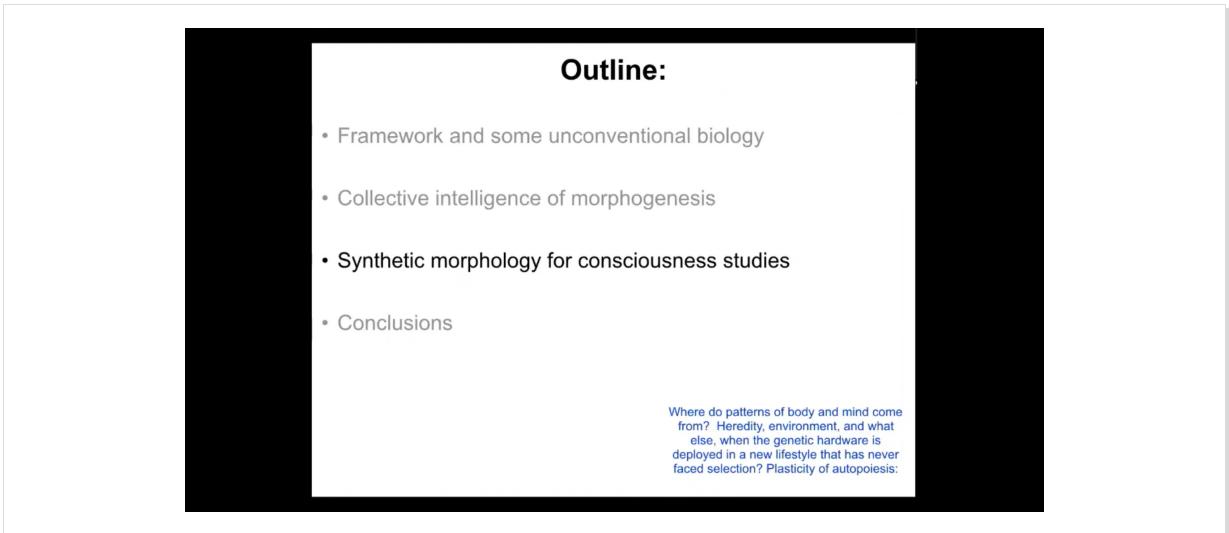
For example, head shape.

We can take a planarian that has this nice triangular head shape, cut off the head, and perturb the bioelectrical network that allows it to remember what kind of head it's supposed to make. You can get things like this. You can get this flat head like a *P. falina*. You can get round heads like an *S. Mediterranean*. You can also get the normal heads.

Not only the head shape, but the actual shape of the brain and the distribution of stem cells become just like these other species, about 100 to 150 million years of evolution between these guys and this one. Remember, there is no genetic change anywhere here. There's no genetic change in the two-headed worms.

Much like every cognitive system that we are familiar with, the whole point is to learn without needing to radically change the hardware. You don't need to mutate your DNA to learn new things or to express the memories that you have. The same thing is true here.

Normally, this hardware finds one very particular attractor in the anatomical amorphous space. Here's where they live. Right here. However, it is perfectly willing to visit other regions here if it gets confused. The way it gets confused is the bioelectronics that help it navigate that space towards a specific location. We now have the ability to alter those and they will end up in some of these other attractors. You can see again how they're navigating anatomical amorphous space using some of the same mechanisms that we see in the brain.



Outline:

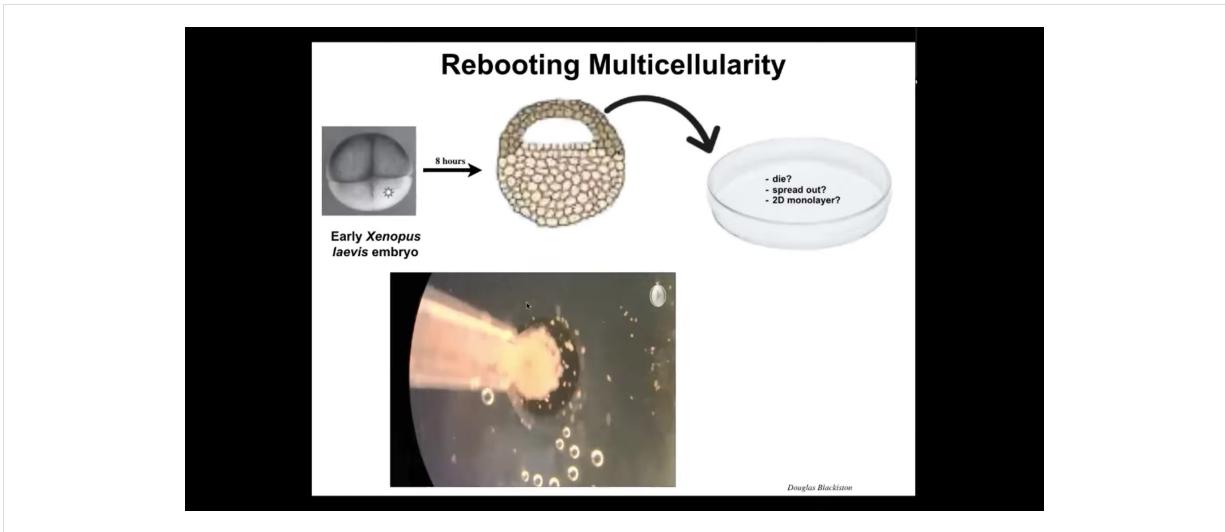
- Framework and some unconventional biology
- Collective intelligence of morphogenesis
- Synthetic morphology for consciousness studies
- Conclusions

Where do patterns of body and mind come from? Heredity, environment, and what else, when the genetic hardware is deployed in a new lifestyle that has never faced selection? Plasticity of autopoiesis:

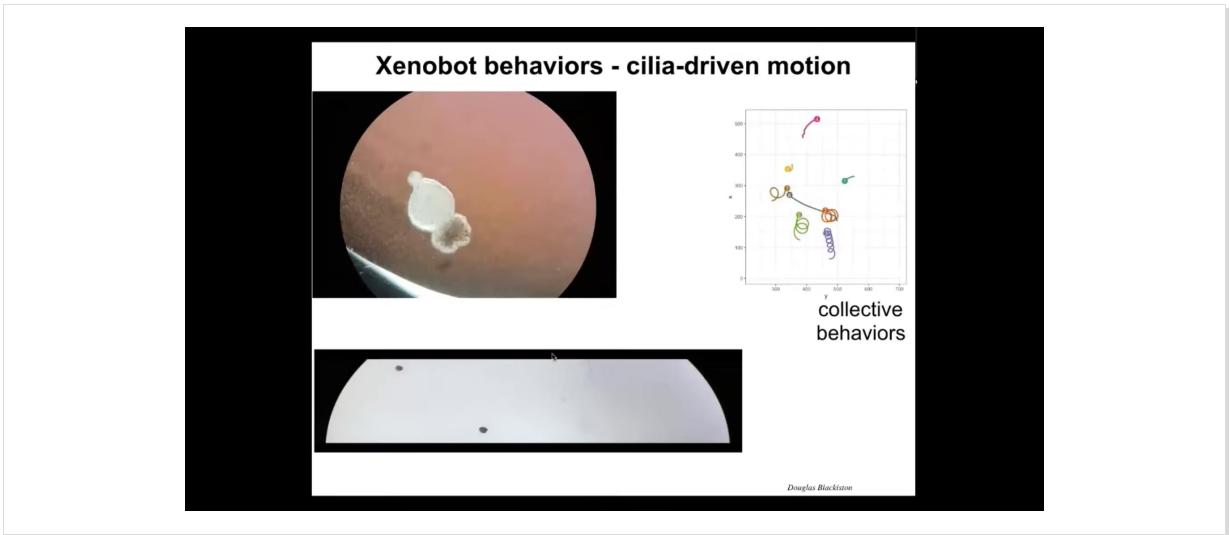
Okay, we come to this next-to-last section because what I've shown you up until now are endogenous examples, meaning natural evolved early ways that biology uses representation, memory, and the bioelectrical interface to store goal-directed set points that then, with various degrees of ingenuity, the collective intelligence of cells can follow.

What we want to do now is ask the following question. In these natural examples, when one asks, where do the goal states come from? Usually we say it comes from a combination of heredity: a long history of selection for a particular pattern memory with certain environments. There are some constraints of physics.

But we wanted to ask more broadly, in novel cognitive systems, where do goals and memories come from? That is, when we deploy certain genetic hardware in a completely new lifestyle that has never faced selection in that configuration, what actually happens? Here we turn to synthetic morphology, and I'm going to show you this particular example.



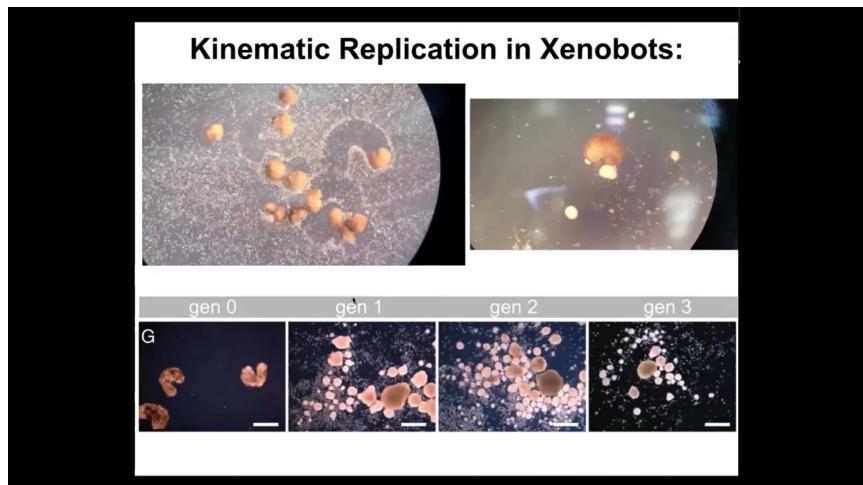
These are called Xenobots, and the way they form is we take early frog embryos and we take some of the epithelial cells at this early blastula stage from the apical end of the embryo, we dissociate them, and we put them in a dish. There are many things they could have done. They could have died. They could have spread out and moved away from each other. They could have formed a nice two-dimensional monolayer, a cell culture. But instead, they coalesce into this little mass. The flashing is calcium signaling, but you can see these cells coming together. If you look closer, there's some amazing things that they do when they come together.



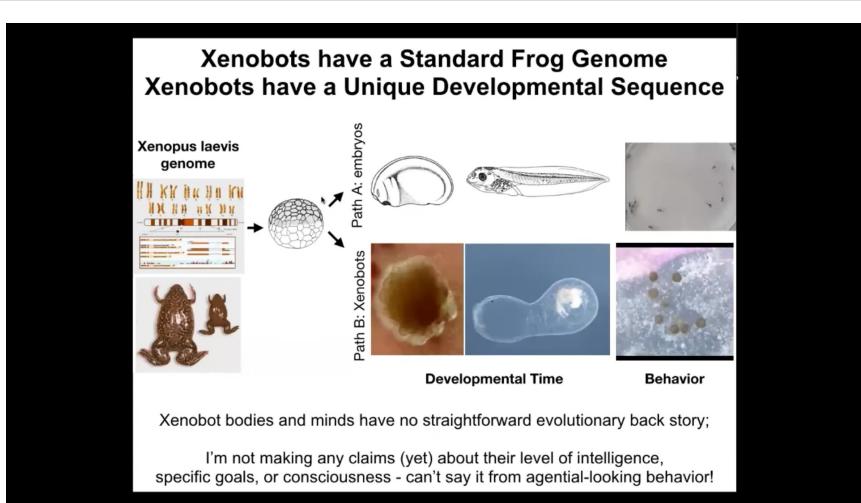
Because it is ciliated, it's covered in cilia whose job normally is to distribute mucus down the side of the animal. In this case, it rows against the water, so they're able to move. They can go in circles, they can patrol back and forth like this. They can have collective behaviors. Here's two of them interacting. These are resting in place. This one's on a very lengthy journey. There are all kinds of behaviors.



It traverses this maze in the still water. The water's not moving and there are no gradients. There's nothing in there besides water. It takes the corner without having to bump into the opposite wall. It's moving along straight and, without having to hit anything, it takes the corner. Then spontaneously it turns around and starts to go back where it came from. All of this is completely internally driven. It's autonomous. We're not controlling the movement the way people do with some engineered biobots. They have all kinds of unusual behaviors that were not predicted in advance.



For example, one of the things that it turns out they can do, this was found in simulations that our collaborator Josh Bongard and Sam Kriegman did, is that if you provide loose epithelial cells to the bots, both on a collective population level and on an individual level, they will polish them into these balls. They collect them into little piles, and then they polish them into these little balls. Because they're dealing with a gentle material themselves, these are not passive particles, these are cells, these balls then mature and become the next generation of xenobots. They run around and do exactly the same thing and they make the next generation, and the next. So we call this kinematic self-replication. It's von Neumann's dream of a robot that can go out and make copies of itself from material it finds in the environment. As far as we know, no other creature on Earth reproduces by kinematic replication. This is entirely new.



And so, when we make it impossible for them to reproduce in the normal frog-like fashion, they are able to do this. And so that leads to the question of plasticity: what did the *Xenopus* genome learn in all of its time on Earth? Well, it certainly can do this, and this is what it does very reliably under standard circumstances. This is the default behavior of that hardware. You can think about this as instincts. The stereotypical morphogenesis is the way that some birds are born knowing how to make certain nests. So this is the kind of instinctual built-in behavior, but there's much, much more plasticity.

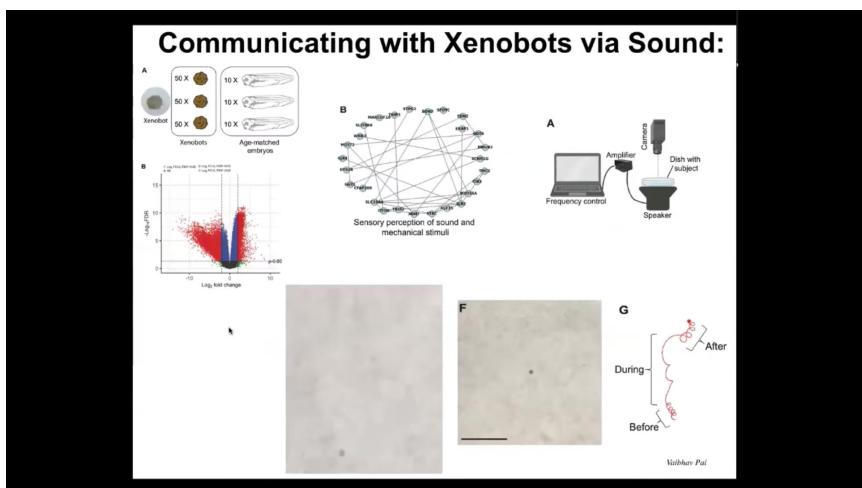
When these cells are liberated from the embryo, we didn't add anything. There are no nanomaterials here, no scaffolds, no genomic editing, no synthetic biology circuits. We did almost nothing to them except liberate them from these other cells, which normally bully them into having a boring life as the two-dimensional outer covering of a frog that will keep out bacteria. In the absence of those signals, you get to find out what these cells really want to do. And what they want to do when shielded from the instructive influence of other cells is this. They become xenobots.

This is an 84-day-old xenobot. It's turning into something. We have no idea what this is. So it's got some weird developmental sequence of its own. It has different behaviors than these tadpoles do. And probably this is just the tip of the iceberg. My guess is these things can do many different things under many different circumstances. It's very interesting because there has been no selection to be a good xenobot. There's never been any xenobots. There's never been any selection to do kinematic self-replication or to have these kinds of patterns and behaviors.

As a model system for the kind of diversity of embodied minds that we're going to be dealing with, this is a model system for that. Now, I'm not yet making any claims about their level of intelligence, their preferences, their goals, or their consciousness.

Some of this work is ongoing. We should have some things later this year on what they can learn. But certainly you can't say any of these things just from observing this kind of behavior. You have to do perturbative experiments, which is what we're doing.

Slide 37 of 44 · Watch at [51:21](#)



But one of the interesting things we found is that if we look at the gene expression in these xenobots, turns out that they express hundreds of genes differently than the age-matched tissue in their normal context. We're not worried about genes that are not expressed because of course it's obvious xenobots are missing all kinds of tissues; they're missing endoderm.

Much more interesting are the genes that they upregulate. There are hundreds of new genes that they upregulate, which are very interesting. Again, we didn't touch the genome, we didn't give them any new synthetic biology circuits. All of these gene inductions are because of their new lifestyle. They're dipping into the genome to express new genes that are relevant to their new lifestyle.

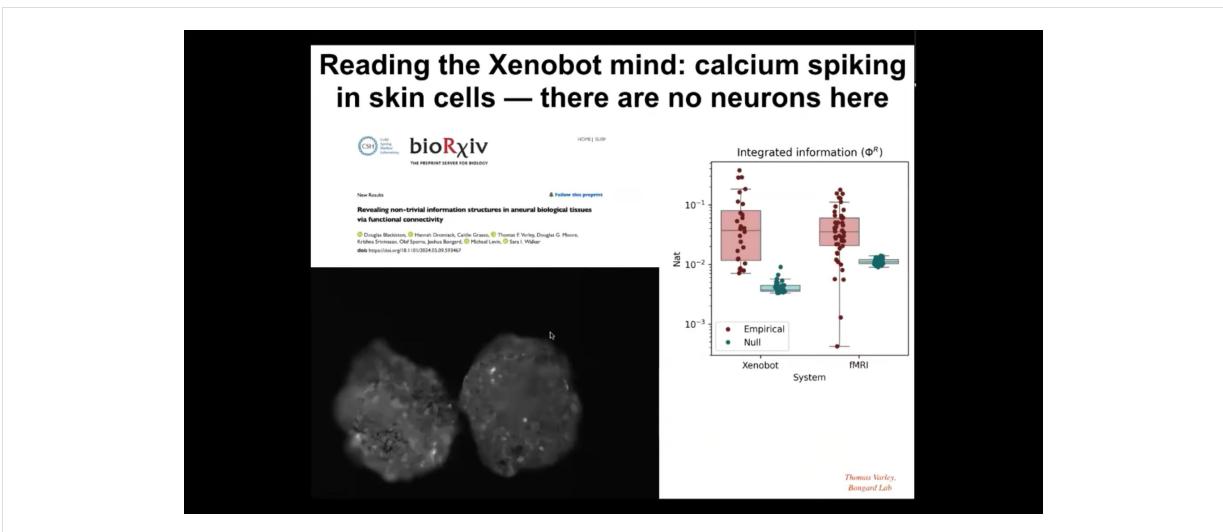
One of the clusters we identified is a set of genes that are known to be related to sensory perception of sound. We thought that was really weird. Could it be that these xenobots are able to hear? We did this experiment where we put speakers underneath the dish.

What you can see here in this video is that this particular bot, for example, moves in a particular pattern. Then we turn on the sound. This is one particular frequency that we've been playing with. You turn on the sound and you can see the behavior change radically. As soon as you turn it off, it goes back to what it's doing, not exactly the

same. There is a little bit of memory here, but it does return to its behavior. We can, in fact, signal to them via sound and change their behavior in observable ways.

Again, this question of what does the transcriptome look like of an entirely novel organism that's never existed before, this is it. It has, I'm sure this is probably just one of many interesting things that we can discover about these.

Slide 38 of 44 · Watch at [53:13](#)



Here's something else that we're just beginning, and this will have even more relevance to this audience who is interested in consciousness and various metrics thereof. What we've got are xenobots that express calcium reporter proteins. This is G-CaMP that is flashing anywhere that you see calcium signaling. People use this in brains all the time. Neuroscientists use this exact construct to track computational activity in the brain. Once you do that, you can apply a couple of different sets of metrics to them. First you can do various connectivity metrics. This is a collaboration with Sarah Walker's lab, Josh Bongard's lab, Olaf Sporns' lab, and all the people who worked on this, basically looking at the information structures and the connectivity that you find in these kinds of tissues and comparing them to what you find in nervous systems. If you want to dig into those details, you can take a look here.

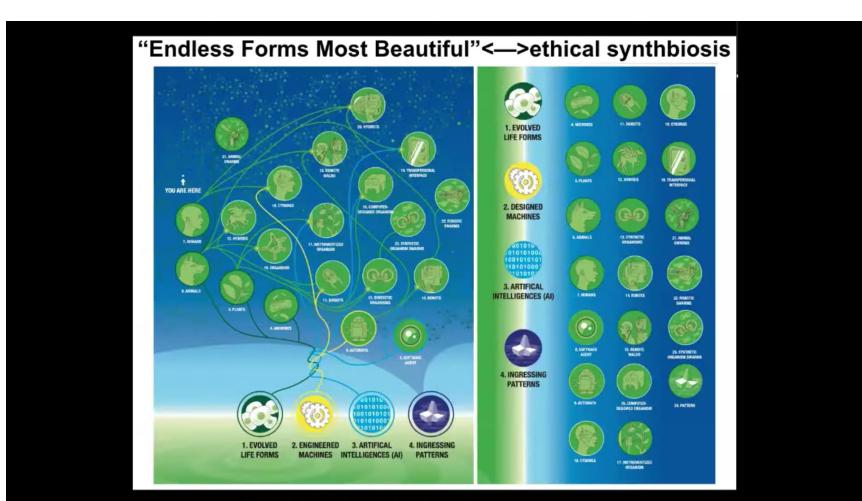
There are metrics that have been developed by a number of people having to do with integrated information. The idea is to have some quantitative estimate of the extent to which the larger-scale system has causality that is more than the sum of its parts. This is a brief explanation; this audience knows all about it. This is one particular metric.

This is called Phi R, which is a measure of integrated information. It's based on Phi 2.0 by Balduzzi and Tononi. This was done by Thomas Varley and Bongard in the Bongard lab; they analyzed the Phi R of calcium signaling in these bots compared to null models, and they found that, just like in fMRI where causal emergence is quite different in brains versus null models, the same thing is true here. There is a considerable amount of integrated information.

I'm not going to make any claims about consciousness here. There are plenty of people who have discussed the relationship between consciousness and measures of integrated information. If you take measures of integrated information applied to brains as a sign of some sort of integrated consciousness, you're going to need to deal with the same issue here, because even though there is no neural tissue, you're finding the same kind of structure in terms of signaling. We have more on this coming, but it's clear that these kinds of metrics used in neuroscience to look for measurable correlates of consciousness are very applicable to these non-neural systems.

In the last couple of minutes, I want to say some other things that might be particularly relevant to this audience, having to do with what I think this all means.

Slide 39 of 44 · Watch at [56:25](#)



So first of all, because of the plasticity of life, every combination of evolved material, engineered material, software, and the ingressing patterns of mathematics is some sort of viable agent. All the biologicals, all the endless forms most beautiful that Darwin was so impressed by are a tiny little corner of this enormous space of embodied minds.

When people think about what's going on in whales and octopuses with their semi-autonomous limbs and some of those things, that's a tiny piece of this whole question. Hybrids, cyborgs, all kinds of chimeras, every possible version of this material is going to be living with us in the next few decades. I think that having a theory of mind about these things and a theory of consciousness that goes beyond evolution and beyond the end of one example that we have in the evolutionary tree here on Earth, and understanding it more broadly in very diverse embodiments that are very different from ours, is going to be essential to enabling us to reach some kind of ethical synth biosis with these beings with whom we're going to share our world.

Slide 40 of 44 · Watch at [57:51](#)



And just to say a couple of somewhat more wild things towards the end of this talk, since this is a meeting about consciousness, I want to say two things, and then if people want to dig in, we can discuss them.

First of all, my view of where the specific goals, competencies, and behavioral propensities come from if it's not a history of selection. So in novel beings, such as the bots that I showed you, of course, we also have anthrobots, which are biobots made of adult human cells, and all of the different cyborgs and hybrots. Where do their goals and competencies and kinds of minds come from if you can't pin it on a long history of selection?

All I will say here is I think they come from the same place where the truths of mathematics come from. Platonist mathematicians believe that they are exploring an

existing and ordered space of truths, the properties of prime numbers and the facts of number theory. They are not determined by anything in the physical world. However, they matter greatly to physics, and I'm going to claim to biology as well. They occupy an ordered, structured space. They're not merely a grab bag of emergent phenomena that just happen to hold in our world for some weird reason.

A lot of people like that model better because it's a sparse ontology. You don't need to think about a platonic space of truths besides the physical space. But I think it's a very mysterious position. I don't think it's helpful for future research to assume that these emergent things are just randomly encountered and written down in our big book of surprises.

I actually think the research program here is to map out the space because this platonic space contains not only very low-agency truths of mathematics, but also behavioral patterns that we would recognize as kinds of minds. Some of these patterns are behavioral patterns of cells and tissues, meaning they come out as morphologies or anatomies, and other patterns are ones that drive behavior in the three-dimensional world, which we recognize as conventional behavior.

I think that all bodies, be they simple machines, embryos, biobots, AIs, humans, are basically interfaces. They are pointers through which some very specific patterns ingress into the physical world from a platonic space. Understanding this physical interface is only the first part of what we need to do. We also need to understand what kinds of behavioral patterns affect and manifest through these physical interfaces, just as we try to understand how symmetries and truths about very abstract objects control various aspects of physics from the lowest level on up.

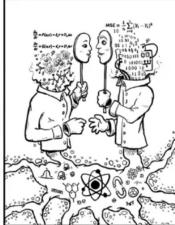
I think that synthetic morphology beings such as xenobots, anthrobots, and related constructs are basically vehicles for exploring this latent space. I think they are an indispensable tool for consciousness research as we try to understand what kinds of patterns are important for human consciousness and what else is in the space. I think we're going to have to map out the space to have a proper understanding of the mind-brain relationship.

Emergent Goals and Competencies

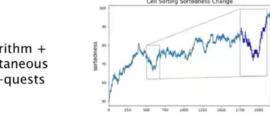
humility warranted: even bubble sort has emergent delayed gratification NOT explicitly in the algorithm

We underestimate matter and we underestimate algorithms/"machines"

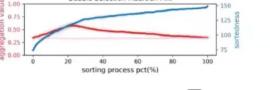
Algorithm + spontaneous side-quests



It does not take cells, life, or huge complexity to have emergent goal-directed competencies that are other than chance or necessity



Bubble Selection Insertion Mix



Classical sorting algorithms as a model of morphogenetic Self-sorting arrays reveal unexpected competencies in a minimal model of basal intelligence

Taiming Zhang*, Adam Goldstein* and Michael Levin^{1,2,3}

1. Department of Biology, Tufts University, Medford, MA 02155, USA; 2. Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA; 3. Department of Physics, Tufts University, Medford, MA 02155, USA

Sage

Taiming Zhang, Adam Goldstein, Michael Levin

The final thing I want to say is this: it's often thought that there's some essential relationship between complexity and consciousness, that in order to have consciousness, you need to reach a certain level of complexity of organization, and whether it's emergent or, as I think, it's actually a view from this platonic space of kinds of minds, either way, it's very common to think that not only do you need biology for that to happen, but you actually need very complex biology, for example, brains.

We are looking at some extremely minimal systems. I will point people to this one paper where we studied sorting algorithms like bubble sort. These things are very simple algorithms, completely transparent, deterministic. People have been studying them for many decades. Yet they have features that have never been noticed before.

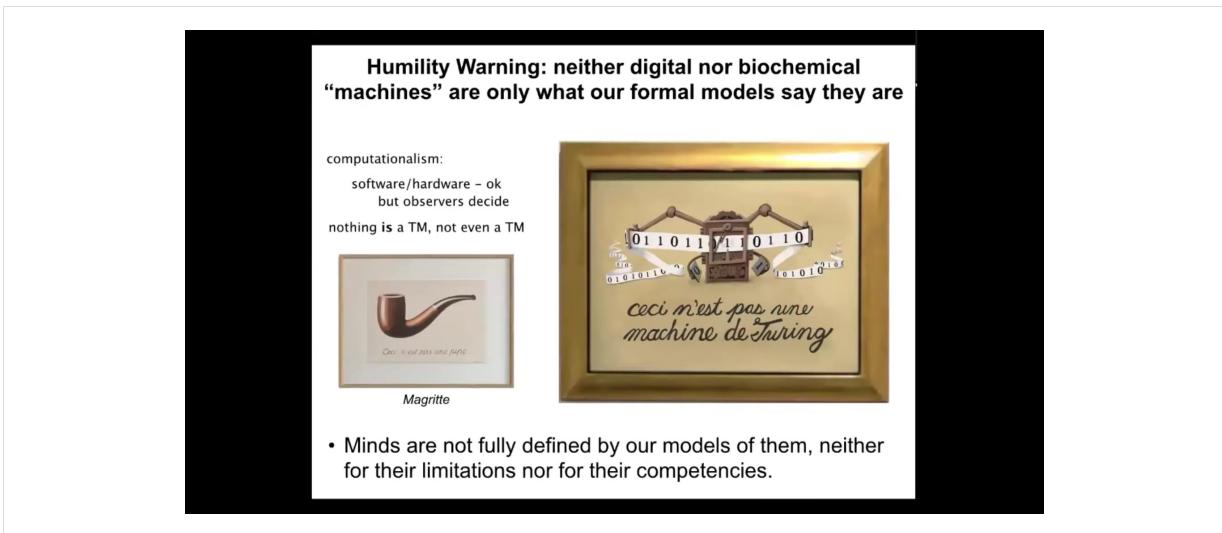
What we found is that they have competencies that are not obvious from the algorithm at all, such as delayed gratification and other behavioral tricks, but they also do these weird side quests in addition to the sorting that they do. This is the sorting of a random string of numbers, and certainly they do the sorting. But in the meantime they do this other interesting thing called clustering, which is nowhere in the algorithm. It is something that is not forbidden by the physics of their world, but neither is it prescribed by it. So it's an interesting new kind of freedom, but it's neither determined nor random.

I think ultimately this will have impact on the free will problem.

Basically, what we're finding is that even extremely simple systems that are not like biology whatsoever, that don't have a history of selection and so on, even very simple deterministic systems have behavioral competencies that are not directly encoded in the algorithm. You do not have to be a complex life form with a brain to do things that are not obvious from your construction.

I think that for the exact same reasons, we understand that the laws of chemistry do not tell you everything you want to know about the human mind. For those exact same reasons, we should cultivate some humility.

Slide 42 of 44 · Watch at [1:03:57](#)



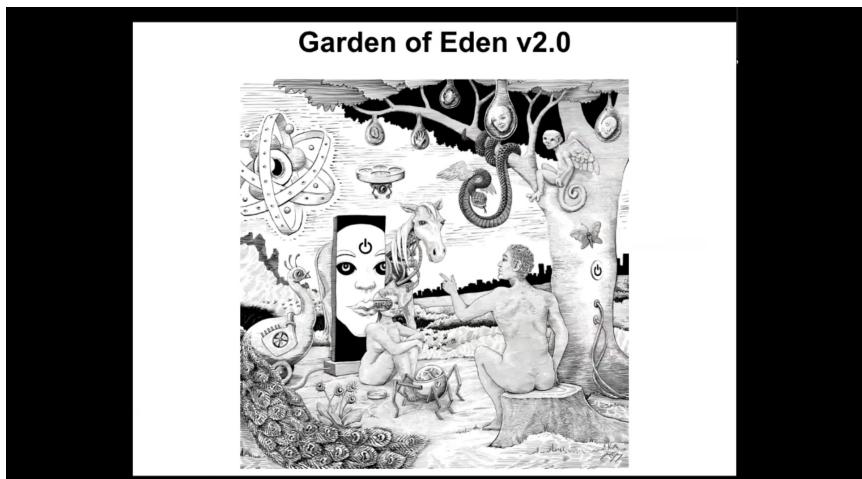
And realize that the algorithm and the materials of which quote unquote machines, including very simple machines are made, does not tell the whole story about them either. In other words, they too are subject to interesting ingestions from this platonic space of forms that we did not see coming and that are not fully circumscribed by our stories of the material and the algorithm. I would just go along with Magritte, who is trying to remind us here, and this is a classic piece of art that says "This is not a pipe" — that our representations of things, our formal models of things are not the thing themselves and do not capture the entirety of the thing.

And so I think this is true. And so, I say this is not a Turing machine. Not even Turing machines are a Turing machine in the sense that our formal models of computation do not describe the thing, the full capacity of the thing that we build when we make these things, and so when people argue whether computationalist views of consciousness are sufficient. My point is, not only are they not sufficient for brains and living things, they're not even sufficient for the things that they were supposed to describe, which is quote unquote machines. All of these are just formal models. They capture some aspects of what's going on. But as we're seeing now, they don't capture everything.

I come back to my fundamental notion of a continuum of cognition. And it is not that you have to be somewhere on the far right of that continuum and be a complex brain in

order to have some sort of freedom from being fully encompassed by models of your parts, such as the laws of chemistry and physics, that already begins at a very low level, at the level of very simple machines. Already computationalism fails there, as all formal models do. There's nothing wrong with computationalism per se. We have to remember that none of our formal models fully capture what's going on here. And so I think it's critical that whatever their embodiment, whatever their origin story, even very simple minds are not fully defined by our models of them, neither for their limitations nor for their competencies.

Slide 43 of 44 · Watch at [1:06:11](#)



Ultimately, I think the Garden of Eden is going to look like this. It's going to be a very weird place with all kinds of things that we are absolutely going to have to name in the sense of understanding their true inner nature, which is not easy. We're bad at it. We do not have the native kind of skill or talent at recognizing diverse minds. We are going to need tools for it.

I sometimes give a different kind of talk where I go through this ladder where we get progressively weirder kinds of systems that I think are actually cognitive kin. It's similar to what happened in mathematics where we progressively found stranger kinds of objects that mathematics is going to have to deal with. Each of these transitions was painful for the community. People really freaked out about irrationals and things like that. People were killed over it back in the day.

Each of these transitions across different kinds of systems, in order to understand what the level of cognition and perhaps consciousness is in any of these things, we are going

to have to break all kinds of prior assumptions and prior categories, just like was done in mathematics to get to new kinds of numbers. This is something that we're working very hard on in our group: to see what all these things have in common.

Here's the summary of everything I've said today. I'm pretty confident that most useful concepts in this field are continuum models. They are not binary crisp categories. We need models of scaling and transformation of minds and consciousness, not trying to prop up ancient categories. I'm very certain that none of this is about brains per se or embodiment in three-dimensional space. The most interesting thing about cognition and consciousness is how it explores very different spaces that we have a hard time seeing. I would claim that our formal models do not tell the whole story, either for life or for machines or for any of these things. We have cognitive patterns that show up that are not just complexity or unpredictability; they're actually new behavioral skills.

This is stuff that we're actively working on. That is what aspects of the embodied architecture, be it engineered or evolved, matter for determining what you're going to get from that Platonic space.

I would claim that evolution has no monopoly on making minds; engineering is potentially just as good, because engineering can use evolutionary computation and so on. Richard Watson has some arguments against this. His are probably the best arguments against the typical assumption that this is the case. I think it doesn't hold any water, but Richard has some interesting ideas that I encourage you to check out.

We're trying to improve our tools for exploring this Platonic space and for understanding the mapping between the pointers, meaning the embodiments that we create, and the patterns that ingress through these interfaces.

The things that I really don't know what to say about at this point — I think we need to do a lot more here before we can say any of this — are questions like this. How well does consciousness actually track intelligence? I think Anil Seth's view that these are orthogonal is probably right, but I'm not sure they're completely independent. I suspect there's some relationship, but we don't know.

I also don't know if there are any so-called phase transitions. I know it's very popular to say that there are specific distinct kinds of minds. There may well be, and the patterns in the Platonic space may in fact have distinct categories. It's certainly true for some mathematical objects like integers, but it's unclear to me whether that space is truly smooth and continuous or whether there are some discontinuities there.

What prospects do we really have while remaining more or less human for understanding very diverse minds? It's not clear to me how far we can get without changing ourselves, which is fine.

That's part of our future for sure, is to change ourselves to gain a better first-person understanding of very exotic minds and be able to enlarge our own cone of concern and compassion to these other beings.

So that's it. I'll stop here if you want more details. Lots of these papers drill down into all this stuff.

Slide 44 of 44 · Watch at [1:10:47](#)

Acknowledgements:

Postdocs and staff scientists in the extro lab:
Douglas Blackiston - brain-body interface plasticity, Xenobots
Wesley Clawson - computational neuroscience and hydrots
Patrick Erickson - cell framing
Suman Chakrabarti, Manika - memory in gene-regulatory networks
Federico Pigani - information theory and computational analyses of diverse intelligence
Yahia Almousa - memory in living and non-living systems
Federico Pigani - causal emergence metrics in living and non-living systems
Tal Shomrat - persistence of memory in regenerating brains
Patrick Gosselin - memory in regenerating planaria
Laura Vandenberg - craniofacial unscrambling
Adam Goldstein, Taining Zhang - emergent competencies of algorithms
Nester Oviido, Junji Morokuma - pattern memory in planaria

Graduate Students:
Arielle Baskin - electro-visual induction
Fallon Durant - Vom, and pattern memory in planarian regeneration
Gemma Gao - NIK Davies - Autonomic
Adam Goldstein, Taining Zhang - emergent competencies of algorithms

Undergraduate Students:
Maya Eason-Bell, Kelly G. Sullivan - non-genetic cross-species morphogenesis
Pranjal Srivastava - Michael Lesser, Ben Semegran, Andrew Bender, and
Douglas Hazel - Ambrobitos
+ many other undergraduate students working in our lab over the years

Technical support:
Raksha Colon, Jayati Mandal - lab management
Erin Switzer - vertebrate animal husbandry
Joan Lemire - molecular biology

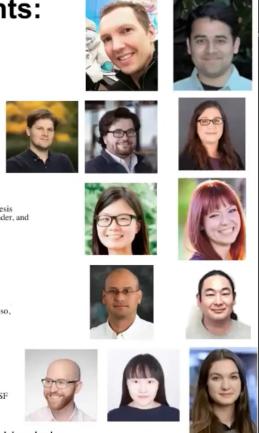
Collaborators: Allen Cognition members:
Johann Bongard - morphogenesis and AI
James Traer, Olaf Sporns, Sami I. Walker, **Thomas F. Varley**, Hannah Dromiack, Caitlin Grasso, Douglas Moore, Krishna Srinivasan - C++ - informatics
Shrawan Kumar - machine learning and AI
Chris Fields - physics of sentence and sentence of physics
Richard Wilson - computational models of cognitive scaling and evolutionary learning
Giovanni Pezzulo - cognitive robotics and morphogenesis
Anil Seth - consciousness in unfamiliar embodiments

Model systems: tadpoles, planaria, zebrafish, slime molds, human cells, biobots, animals

Funding support: JTF, TWCF, DARPA, Paul G. Allen Frontiers Group, Emerald Gate, NIH, NSF

Illustrations: Jeremy Gray @ Penguin Creative

Disclosures: Morphochemicals, Fauna Systems, Astonishing Labs



The most important thing is to thank the people who did most of the work that I showed you today. These are the various postdocs and graduate students that contributed to all the results I showed you. I want to thank many of our amazing collaborators and funding support from various agencies and foundations that have supported us over the last 30 years. Three disclosures, because there are three companies that have spun out of our work that are supporting us to some extent. This is part of the claim I made at the beginning that it is possible to transition some of these deep philosophical ideas into specific applications that make a difference in the practical world. Most of all, I'm grateful to the biologicals and the non-biologicals that are teaching us about these really fundamental and important questions, not only ancient questions, but critical for the flourishing of minds on Earth and beyond. Thank you so much. I'll 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.