"Unconventional Cognition Through the Lens of Cellular Collective Intelligence in Morphogenesis"

Michael Levin

Recorded on September 19, 2024

About this document

This document is a companion to the recorded lecture *"Unconventional Cognition Through the Lens of Cellular Collective Intelligence in Morphogenesis"*, recorded on September 19, 2024. 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

This is a 1 hour 8 minute talk covering topics of collective intelligence in morphogenesis as a model system for thinking about the origin and scaling of cognition in diverse embodiments.

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.
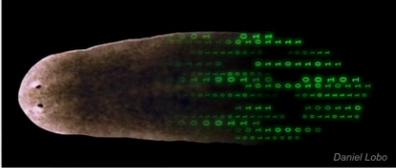

Want one for your lecture?

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

Thank you. Very much appreciate this opportunity to share some ideas with all of you. What an amazing audience to speak to. For today's talk, if you're interested in any of the details, the software, the papers, the data sets, everything is here. This is a blog where I talk about more personal views of what I think these things mean.

I'd like to do an ARC today that has three fundamental points to it. First of all, I'm going to talk about the field of diverse intelligence and the idea of an agential material that is problem solving in unconventional substrates. I'm going to talk specifically about one type of cognitive glue which helps us to solve the scaling problem, that is how emergent minds arise from the competencies of their components, and specifically talk about morphogenesis, development, regeneration, and cancer as a model system for understanding collective intelligence. After all of those examples, we'll come back to the big picture of some thoughts about novel embodied minds and the origin of anatomical, but also cognitive patterns. We'll specifically talk about emergent intelligence as distinct from just emergent complexity. In my group, everything we do boils down to use of collective intelligence whether they be groups of cells or molecular networks and so on, navigating different problem spaces and trying to show how philosophical ideas in these kinds of questions can actually become therapeutics. I won't focus on that today, but probably two-thirds of my lab does things that are very practical and application of these ideas to biomedicine of various kinds. The talk will be in three parts.

The first thing I'm going to do is introduce the way that I try to think about these things, and then we'll go into some examples.

Beyond the "Easy" Dichotomy

So this is a well-known old image. It's called "Adam naming the animals in the Garden of Eden." There are two things about this that are interesting, one that I think is deeply wrong and one that I think is profoundly correct.

The thing that we're going to have to change is the idea that there are discrete natural kinds here. It's very easy to tell who's who. So this is Adam. These are the various species. We can spend our time on the 'easy problem' of discovering what are the kinds of natural competencies of all these different creatures, what kinds of minds they have and so on. So that, as I'll argue in a minute, we're going to need to change.

What is, I think, profoundly correct about this is that in these old traditions, discovering or naming something means that you've understood its true nature. Giving something a name is basically this idea that you've understood what it really is, and it gives you power over that thing. I think we are going to have to name all kinds of really unusual beings in the coming decades and learn to understand their true nature so that we can relate to them. That will become very important, and I'll show you some of those towards the end of the talk.

3
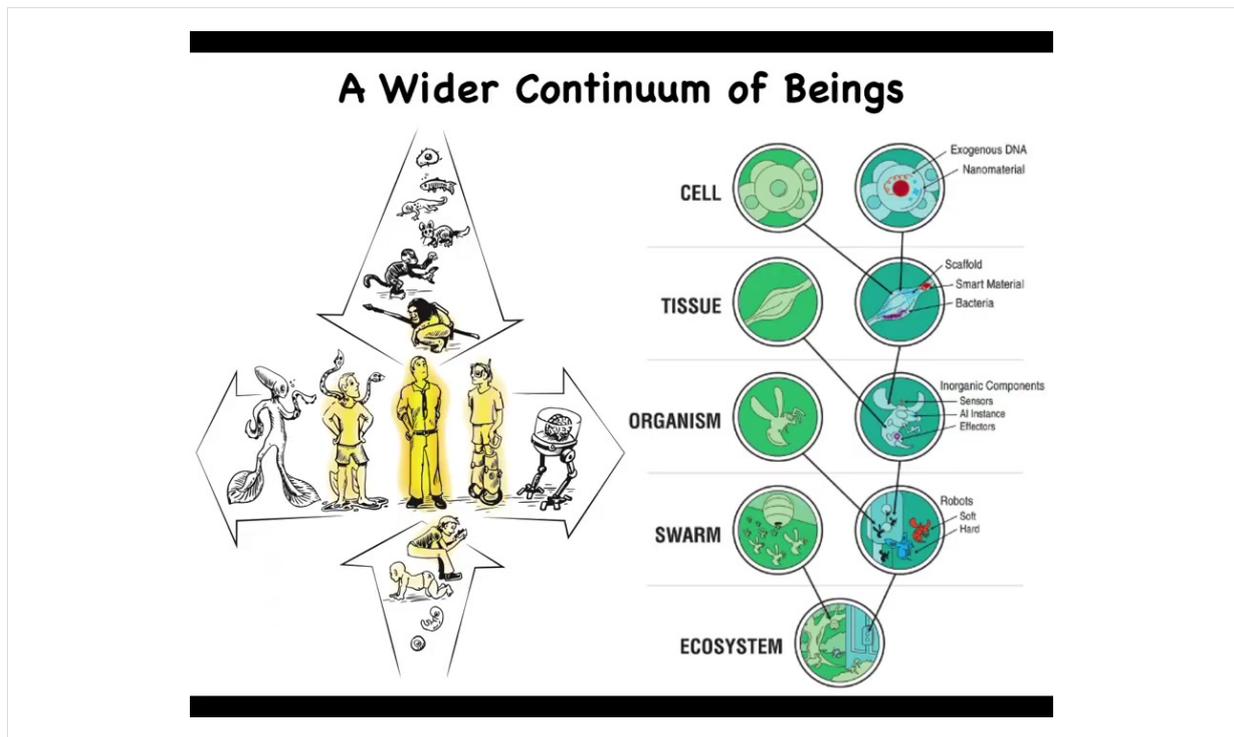
The first and most clear thing that we need to change about this is the idea that we are a well-demarcated human, and there's lots of philosophy and ideas out there about what humans do and how what they do might be different from quote-unquote machines.

We need to understand that we stand at the intersection of several smooth continua. One is that both on an evolutionary time scale and on a developmental time scale, we are gradually changing from different beings.

Whatever you think is true of this modern adult creature, you have to be able to say where that came from and how it either emerged or scaled up from other things that were going on before then.

Now with advances in morphoengineering, synthetic biology, and biotechnology, we also know that we can make gradual and perhaps drastic modifications both on the biological end and on the technological end. We are now able to introduce engineered components across the size scales of living systems.

Whatever conceptual frameworks we have for dealing with embodied minds, they need to be able to apply to all of this. It is not enough to be able to say things about standard humans.

I've been working on a framework whose goal it is to be able to recognize, create, and ethically relate to truly diverse intelligences, regardless of what they're made of or how they got here.

This means familiar kinds of creatures. Primates and birds, but also unusual things, colonial organisms and swarms, engineered synthetic new life forms, AIs, whether purely software or embodied in robotics, and maybe someday exo-biological agents.

The requirements for any such framework that I like are that it has to move experimental work forward. It cannot just be philosophy. It has to actually lead to new discoveries, new capabilities, and new research programs. It has to help us improve or refine ethical frameworks for relating to the kind of unconventional beings all around us.

I call it TAME, T-A-M-E, technological approach to Mind Everywhere.

I'm not the first person to try for something like that. Here's Weiner, Rosenbluth, Weiner, and Bigelow, who tried for this kind of scale all the way from passive matter up to the sort of second order metacognition that humans have in trying to understand how these things scale up.
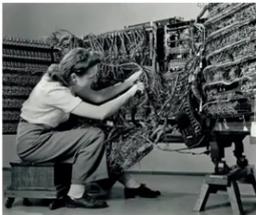
I like to think about something I call the axis of persuadability. What this means is that I view cognitive claims as interaction protocol claims. That is, if you think that something has a certain degree of cognition, what you're really saying is there's a bag of tools, including rewiring, from cybernetics, control theory, behavioral science, psychoanalysis, you're somewhere along the spectrum, and there's a certain set of tools that are going to be optimal in relating to that system.
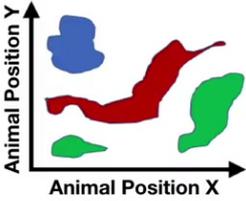
Back here, it looks like prediction and control. Out here, it's some sort of bi-directional enrichment, friendship, love. This is all about finding out what is the best way to interact with a system.

I think there are two key parts to this. The most important being that we can't simply have armchair feelings about where things land on the spectrum. We have to do experiments. We are sometimes strongly motivated to assume that something is somewhere along this continuum, but actually, as we found, there are many surprises.

So we have to commit to the idea that you make hypotheses and do experiments as to where something is going to be on this. That's good. That means that this becomes an empirical problem where you can hypothesize a set of tools, somebody else has a different set of tools as appropriate. You both try it, and we all get to find out who had the better experience.
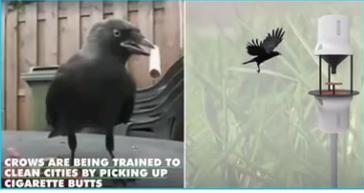
The other thing about this is that the problem space within which these systems can work could be very difficult for us to rec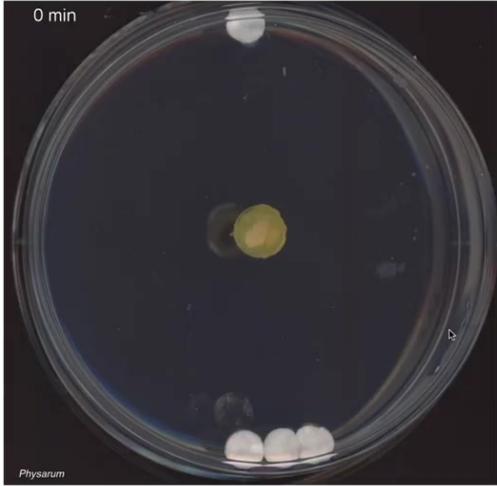ognize. We as humans and as scientists are okay at recognizing intelligence and medium-sized objects moving at medium speeds.

Birds and primates and a whale or an octopus. We can see what's going on in three-dimensional space. But life uses many other problem spaces: transcriptional space, the high-dimensional space of gene expression, physiological state space, and anatomical morphospace, which we'll spend most of the time today talking about. These other spaces are spaces in which different types of agents can run perception, decision-making, action loops, and those are very difficult for us to visualize.

If we had evolved with a primary sense of our blood chemistry, the way that we can taste and smell, and could sense other parameters in our blood chemistry, we would have no trouble visualizing that our liver and kidneys are intelligent symbionts that traverse these spaces daily and try to achieve certain goals and keep us out of trouble. But it's very hard for us to observe behavior in these other spaces.
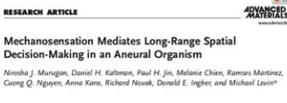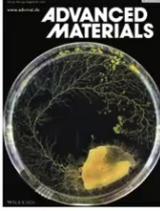
It's also the case that the distinction between these spaces is very, very artificial. So just to show you one example, this is a slime mold called a Physarum polycephalum. We put a little bit of the slime mold here. There are three glass discs here. There's one glass disc here. They are completely inert. There's no food or anything like that. They're sitting on an agar substrate. What you will see is that here, it grows out in all directions. But what it's doing that you can't see during all this time is it's gently tugging on the medium and receiving back information, biomechanical information about the strain angle. In the end, it's able to figure out where the larger mass is. And it does this quite reliably. So what it's doing is obtaining information about its environment. Up until here is cogitation time. And then now it's made a decision and bang, and there it goes.

The thing about it is that this is all one cell. So this whole thing is just one cell. There's no neurons. There's one cell. But what's interesting for the current point here is that in this creature, morphogenesis is behavior. It's in the same space. We like to divide those things, but this thing behaves by changing its shape. This is literally its outgrowth. So what we see as spaces is not necessarily what the creature itself sees, and we need to take that into account.
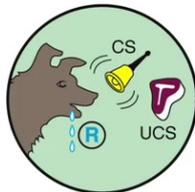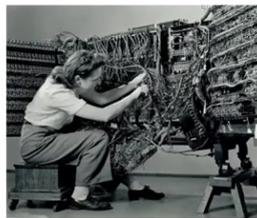
And that reminds us that this is not about humans and scientists in particular. This is about all observers. I like to think about bodies, and this is a polycomputing framework that Josh Bongaard and I developed, where bodies are composed of numerous nested, interpenetrating systems that observe and try to hack each other constantly. We need to understand what each of these systems is seeing, what space they're navigating, and what their goals are.
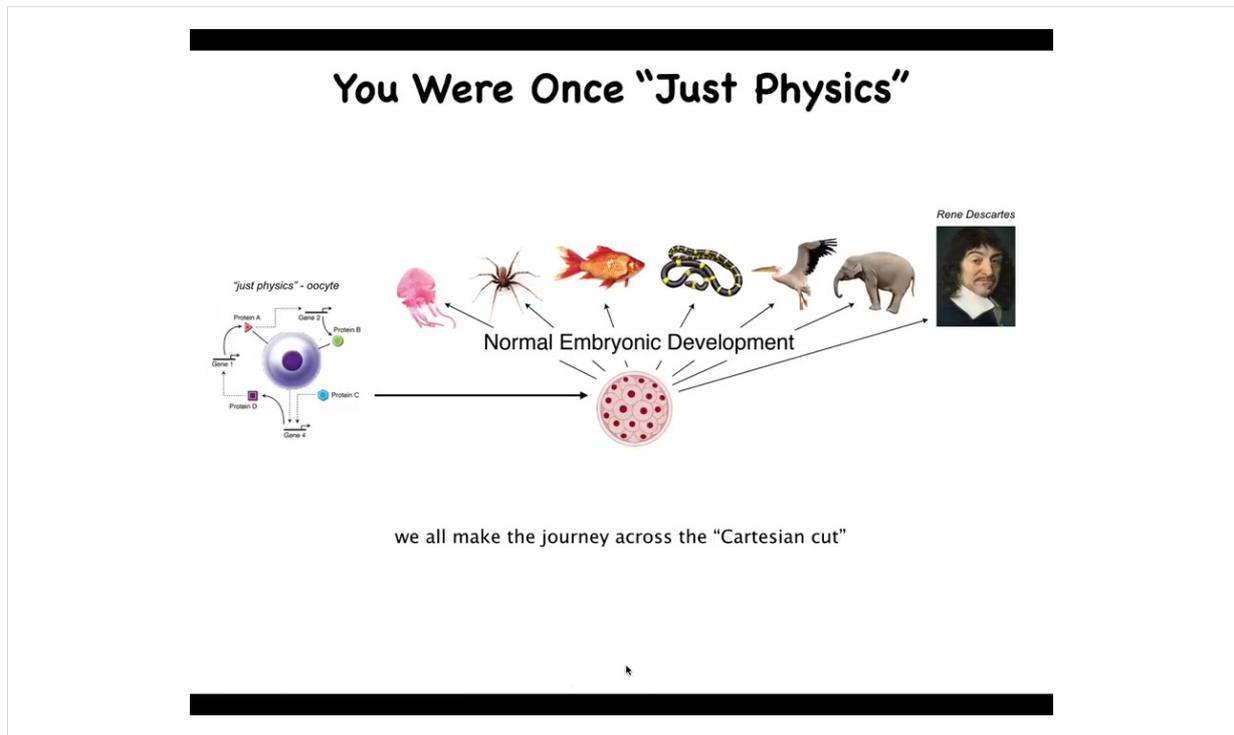
So let's think about cognition from the beginning. We all start life here as a little BLOB of chemistry and physics. And so we all look at an oocyte and many of us say this thing follows the laws of chemistry. There is no mind, there is no cognition, it's purely mechanical. But what we know from developmental biology is right in front of your eyes: for many of these species, it will slowly turn itself from this kind of system to this kind of system. And so we all make this journey. At some point, we become capable of voicing this idea that we are more than "machines." And so we need to understand where and how that happens. What is the scaling that allows this to take place?

Many people, once they start thinking deeply about this, find this disturbing, because it really emphasizes the continuity of whatever it is that we are with very basic physical mechanisms. But at least we are a unified intelligence.

We have this nice centralized brain. We're not what Ricard Soleil calls a "liquid brain," a collective swarm intelligence like ants and bees.

Jose Calvo, pineal gland histology

Gaël McGill
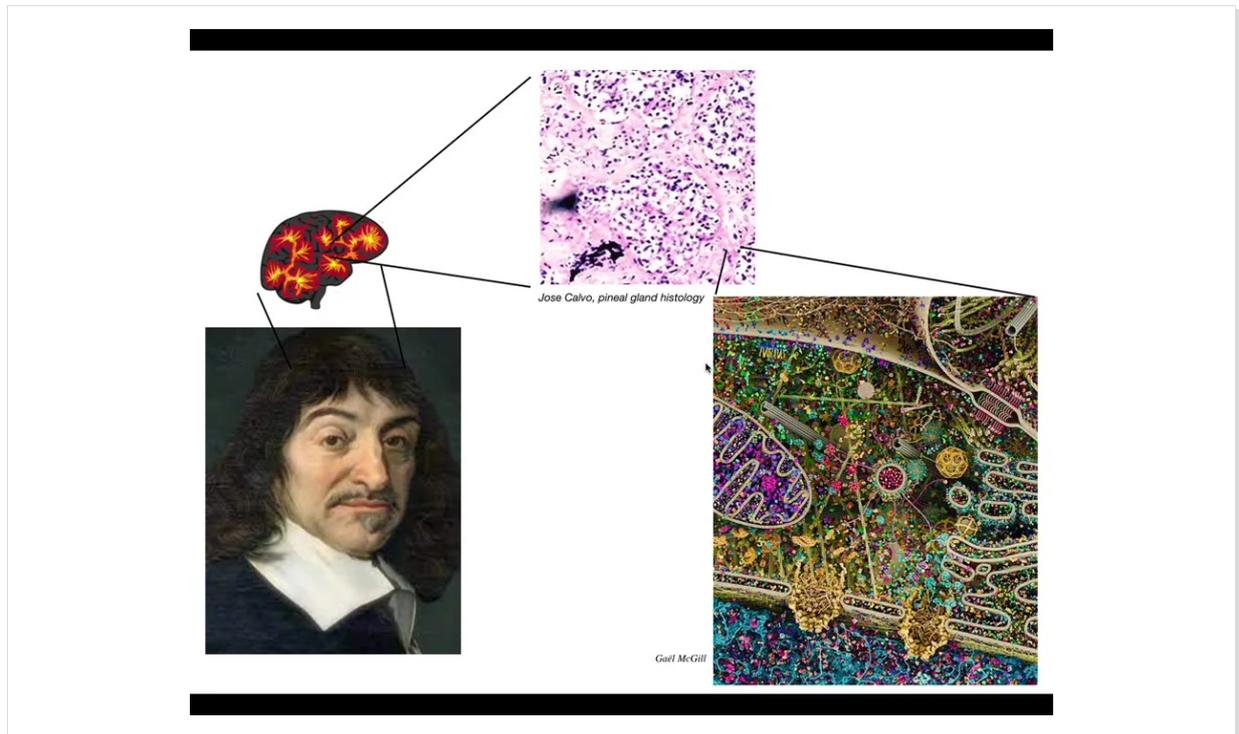
In particular, Descartes liked the pineal gland because there's only one of those in the brain and he felt that unity was appropriate to the kind of coherent unified perspective that we as humans enjoy. But what he was missing was access to good microscopy, because if he had that, he would have been able to look and find that there isn't one of anything. Inside the pineal gland is all of this. A huge number of individual cells. And inside each of those little cells is all of this stuff. There's an incredible nested complexity.

We are all collective intelligences, not just the ant colonies. All of us are made of parts. Our parts in particular are very clever. This is a single cell. This is an animal called the Lacrymaria. There is no brain, there is no nervous system, but you can see it has this amazing control over its morphology here. It's hunting for food in its environment. It's very competent at single-cell-level agendas. This is the kind of thing we're made of.

What we need to do is to develop models of the scaling of competencies of our components into the kinds of things that can store and process goals and memories and preferences that we have that none of our parts have. In fact, the competency goes below the single-cell level.

If you wonder how this thing is doing what it's doing, what it has is a bunch of biochemical networks. And we've now shown that even very simple models of very simple molecular networks, for example five or six node gene regulatory networks, you can, if you do the experiment and not assume that they're hardwired and thus boring and mechanical, train them and they will do six different kinds of memory, habituation, sensitization, even Pavlovian conditioning, by exposing them to different stimuli and watching some output nodes.

All that's described here, and we've now built some devices to try to take advantage of that and train cells for applications in biomedicine. Drug conditioning and things like that. So even inside of single cells, from the virtue of the math alone, you don't need a nucleus, you don't need all this other stuff that cells have. These kinds of simple things can do certain kinds of learning. I'm sure this is scratching the surface of what's there.
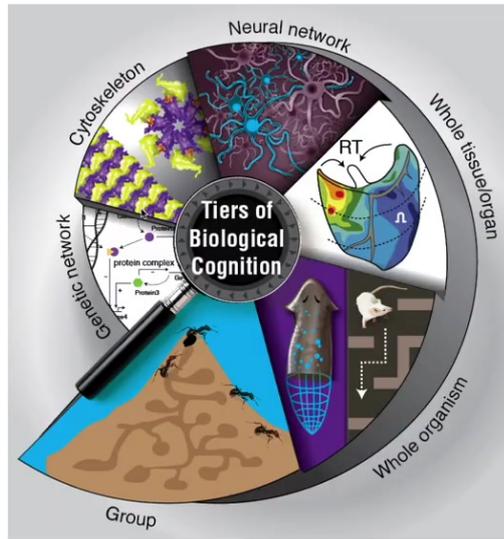
In our body is a multi-scale competency architecture. We are not just nested dolls structurally, but at every level there are components that solve problems in their own space. That architecture has some exciting implications. In particular, it has implications not only for biomedicine but for trying to understand what minds are and where they come from.
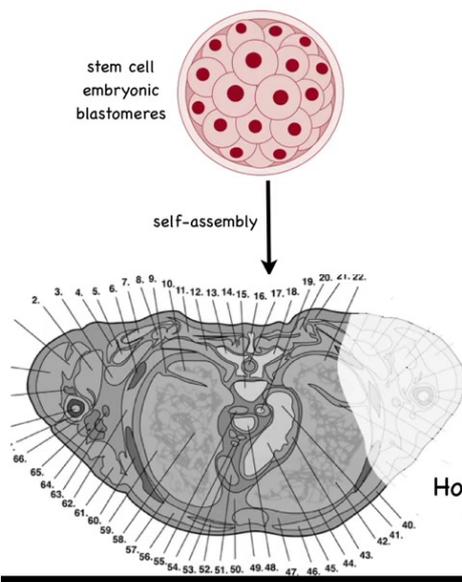
14

What I want to do next is to spend some time talking about one particular example of an unconventional agent that navigates an unfamiliar space, which is anatomical morphospace.

We will talk about morphogenesis as a collective intelligence, and I'll show you the problems that it can solve. We'll talk about bioelectricity as the communication interface to that intelligence and its properties as a cognitive glue.

I always thought it was interesting that Alan Turing, who was very fascinated with problem-solving machines and intelligence through plasticity and reprogrammability and those kinds of questions, had this amazing paper where he asked the question of how order in development arises from chemicals. I think he was on to a very great truth, which is that the self-assembly of the body and the self-assembly of minds have a fundamental symmetry to them, and there are some very deep common things that underlie both of these things. I thought that was interesting that he was on to this really long time ago.
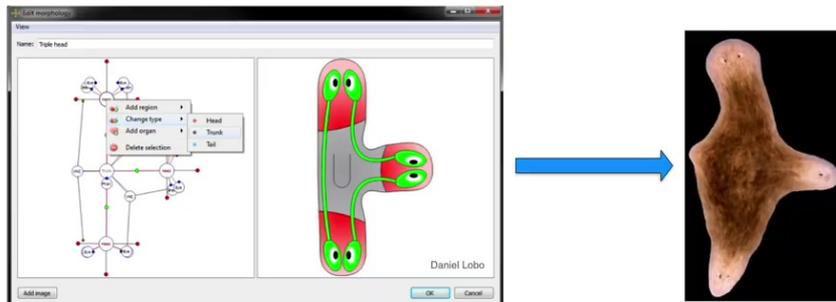
Let's look into this. Where do anatomies come from? This is a cross-section through the human torso. You can see all of the incredible complexity here. All the organs, the right size, the right shape, next to the right thing. Where is this pattern coming from? This is an early embryo. It's a jumble of embryonic blastomeres. Where does this all come from? You might be tempted to say it's in the DNA, but we know that DNA doesn't directly encode any of this. What the DNA specifies is the tiny protein-level hardware that every cell gets to have. The DNA specifies the proteins. It is the physiological activity of that hardware that gives rise to this. This pattern is no more directly in the DNA than the structure of these termite mounds or the precise shape of spider webs are in the DNA of those creatures. There are hardware specifications, and then there is the outcome of what that hardware does.

We need to understand how the cell groups know what to do and when to stop, and how much plasticity they have. We need to understand how to communicate with them, because sometimes you may want to recreate something that was damaged or missing or degenerating.

We're also interested as engineers, interested in Alife, we would like to know what else is possible, given that same hardware. What else could they build? Is this the only thing that these cells can build? We like to think so because cats have kittens and dogs have puppies, but I'll show you some amazing plasticity where it becomes clear that there is not universality, but certainly not hardwiring between what the cells are capable of and what they actually end up building.

16

If we think about running this all the way forward, what would it mean if we've solved this problem? I think it's important to think about what does an answer to these questions actually look like? One way to think about this is as the anatomical compiler. Someday you could be sitting in front of this machine and you would draw the plant, animal, organ, or biobot that you want in whatever shape. Not the molecular properties of it, but actually the anatomy. The thing we actually care about is functional form. In this case, we've drawn this nice three-headed flatworm. If we knew what we were doing, we would have a system that took this and compiled it down to a set of stimuli that would have to be given to cells to get them to build exactly this.

It's obvious why we need this, because if we had something like this, all of this would go away. Birth defects, traumatic injury, cancer, aging, degenerative disease would be a non-issue if we had the ability to tell cells what it is that we want them to build. But why don't we have this? Molecular biology and biochemistry have been advancing for many decades. Why don't we have anything remotely like this? I think it's because we have been thinking about this all wrong. This is not supposed to be a 3D printer or some other way to micromanage cell behavior. This should be a communications device. This should be something that allows you to translate your goals as the worker in regenerative medicine onto the goals of the cellular collective.

Specifically, where molecular medicine today is, we're very good at manipulating the hardware. Which cells interact with which other cells, what proteins are made, and so on. We're really a long way away from control of large-scale form and function. If somebody loses a limb or there's a birth defect or somebody wants a different shape, we, in the generic case, have no idea how to do that.
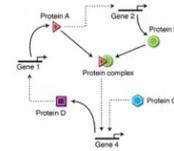
I think that's because biomedicine is still where computer science was in the 40s and 50s, where people think that the correct level of interaction is through the hardware. Think all the exciting advances—CRISPR, protein engineering, pathway rewiring—are down at the level of the hardware. But what we haven't done yet is what computer science has done, which is take advantage of the reprogrammability and higher-level interactions with a system that does not require you to change the hardware.

I think the big stumbling block has been that everybody understands that biology is complex, but there's a lot more to it. It's not simply complexity that shows up, but emergent agency. I'll repeat this theme a couple of times as we go along.

I'm going to show you what I think are examples of a collective intelligence. And what I mean by intelligence, which is not to say that I think this is the one correct definition or that it encompasses everything that everybody wants to subsume under that term. But I like this definition because it's nice and practical; it's by William James. It's the ability to reach the same goal by different means. It means some degree of reaching a goal in a space despite various things that might happen along the way, perturbations. And I like it because it's very cybernetic. It doesn't say you have to have a brain. It doesn't say what the space is or what the goals are or anything like that. It's quite generic.

And so we can think about all ranges of systems from two magnets, which if separated by a barrier will never come around and meet each other because they cannot go further from their goal in order to recoup gains later. They don't have this delayed gratification. Romeo and Juliet have all kinds of long-term planning and all sorts of other tools to avoid physical and social barriers. And in between, you've got your self-driving vehicles, cells, tissues that have some degree of capacity greater than this, but probably smaller than that.

Let's get to it. What kind of collective intelligence do cellular swarms deploy?
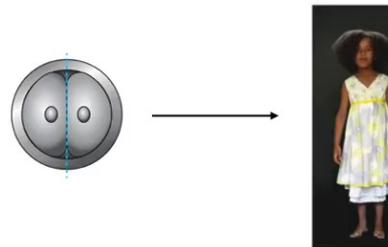
The first thing you notice about development is that it's incredibly reliable. Almost all of the time, these cells produce exactly what they should. There is a massive increase of complexity. You go from this system, which is already quite complex, but it becomes much more so.

I want to be super clear that when I say morphogenesis has intelligence, I do not mean either the reliability of it or the increase in complexity. The fact that you get from here to here is not, in my framework, a sign of intelligence. What I'm talking about is problem solving and navigation with unexpected scenarios. You can easily produce some of those by cutting these embryos into pieces, rearranging the pieces, or mushing multiple embryos together. If you do that in mammals, you don't get half bodies. You don't get confused bodies. If you do it early enough in development, you get perfectly normal monozygotic twins, triplets, and so on. That is because the system, if you cut it in half, it immediately figures out that half of it is missing. It will regrow whatever it needs. In the space, you can start off from different starting positions, avoiding local maxima, and end up within the correct ensemble of states in the anatomical space corresponding to a normal human target morphology. It can make up for quite a bit, and we could spend a whole hour going through these examples.
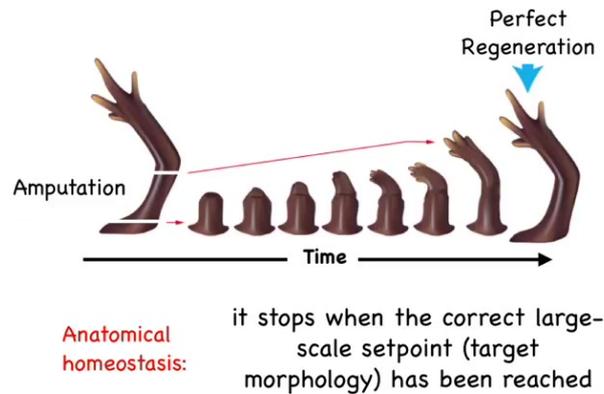
Some lucky animals are able to do this throughout their lifetime. So this guy is an axolotl, and they regenerate their eyes, their limbs, their jaws, portions of their heart and brain. And if you amputate anywhere along the limb, you discover that the cells immediately spring into action. They build exactly what's needed, and then they stop. This is the most amazing thing about that process is that it knows when to stop. What you really have here is an anatomical homeostatic system that when you deviate from the correct position in morphospace, it will work really hard to get back there.
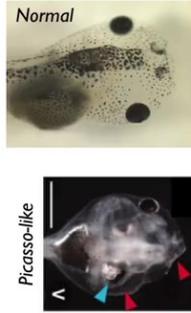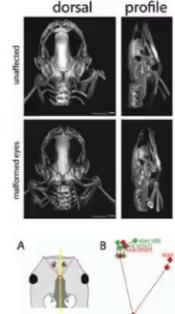
Here's another example of it. We discovered this a few years ago. This is a tadpole. Here are the eyes, the nostrils, the brain, the gut. These tadpoles are supposed to become frogs, and they have to rearrange their face in order to do so. Their jaws, their eyes, their nostrils, everything moves. It was thought that this is a hardwired process. If every organ moves in the correct direction, the correct amount, you'll go from a normal tadpole to a normal frog.

Laura Vandenberg in my group decided to test this hypothesis. What we did was create these so-called Picasso tadpoles. We scrambled these craniofacial organs. The eyes on top of the head, the mouth is off to the side, everything is mixed up. What you find is that these give rise to quite normal frogs because all of these organs will move in novel unnatural paths, sometimes going too far and even coming back, until you get a normal frog face, and then they stop. The genetics does not give you a hardwired set of movements. What it actually specifies is a highly flexible error minimization scheme. It gives you a system that can move through morphospace and recognize when the pattern is incorrect, get to where it needs to go, and then give you the correct pattern, and then stop.

That raises the obvious question. How does it know what the correct pattern is? Where is it going? What controls this navigation through morphospace?
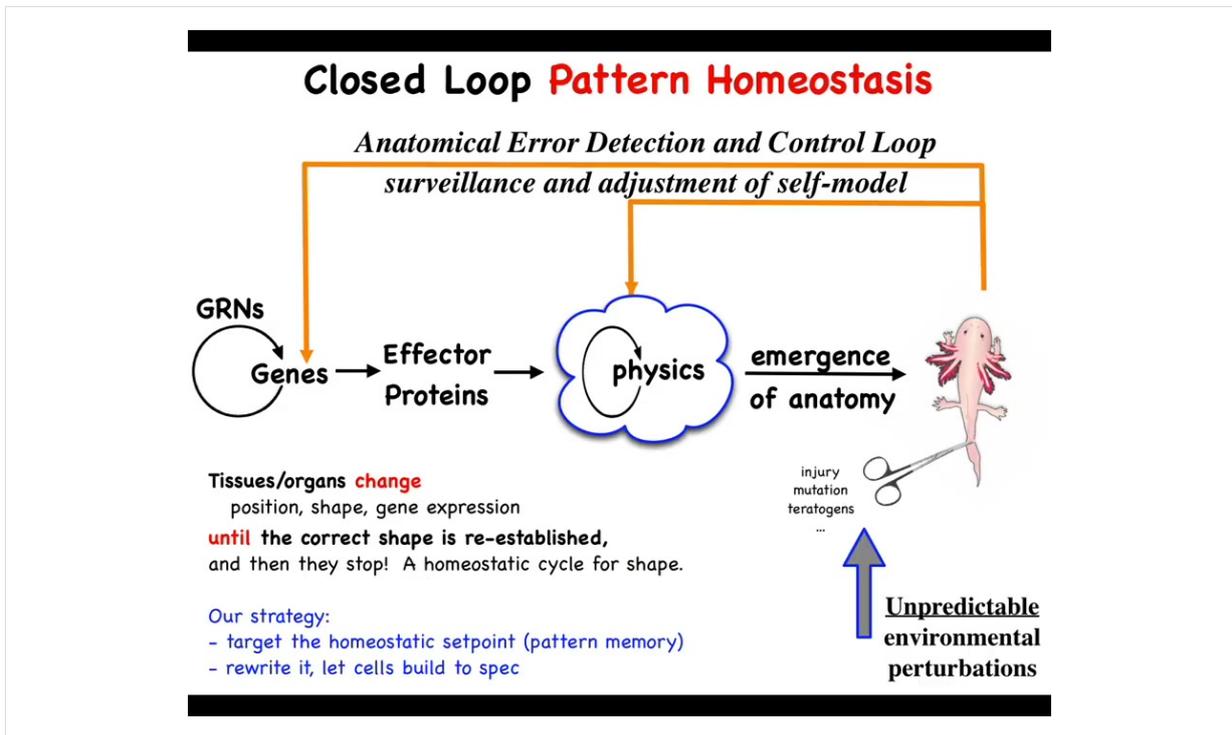
This is the standard view of developmental biology. It's very much feed-forward and emergence. The idea is a focus on emergent complexity from low-level rules. The genes are interacting with each other here. They make some proteins that do things. Then there's some laws of physics and then eventually voilà; something like this emerges as a complex agent. We all know that there are lots of systems, from cellular automata up, that will give you complexity from very simple interaction rules. That's easy.

But what we actually see here is that this is not the end of the story. In fact, it's only the beginning. If this system is deviated from this by injury, by mutation, by teratogens, whatever, then mechanisms kick in both at the level of physics and genetics that try to get you back to where it needs to be. Now you see this patterned homeostatic system. It has a simple goal. It has simple navigational capacities. This is not just a feed-forward emergence of complexity. You actually have a state that the system works really hard to maintain against perturbations. That has important specific implications.
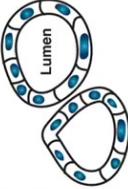
This idea of goals — and arguing about whether these kinds of biochemical systems can be set to have goals — is very practical. This is not just philosophy because if you buy into the idea that this is an emergent complex, that all of this is the result of emergent complexity, then the only game in town is to modify these genes and see what happens. Doing this backwards is not reversible. Development is computationally irreversible, which is why CRISPR and some of those technologies have a real ceiling to them. When you want to make changes here, how do you know which genes to edit? This is generically a very difficult problem.

If we're right and there is this kind of a homeostatic system, then there's hope of a completely different strategy. Find the encoding of the set point, interpret it and modify it, and then let the system do what it does best. In other words, do not interfere with the hardware. Let the hardware do exactly what it does, but change the set point. That requires you to find the set point and then decode the set point. And this is what we've been doing.
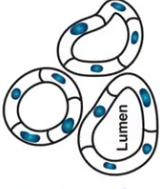
I'll take an aside to tell you that it's even much richer and much more interesting than that, to point out how much intelligence there really is.

This is one of my favorite examples. This right here is the cross-section through the kidney tubule in a newt. Normally there's 8 to 10 cells that work together to form this structure. One thing you can do is make polyploid newts that have multiple copies of their genetic material. When you do that, the cells get bigger.

The first remarkable thing is that you can change the copy number of all the genes, and you still get a totally normal newt. How many chromosomes apparently is not an issue. The size of the cells adjusts to the size of the nucleus. That's pretty impressive, but it gets better.
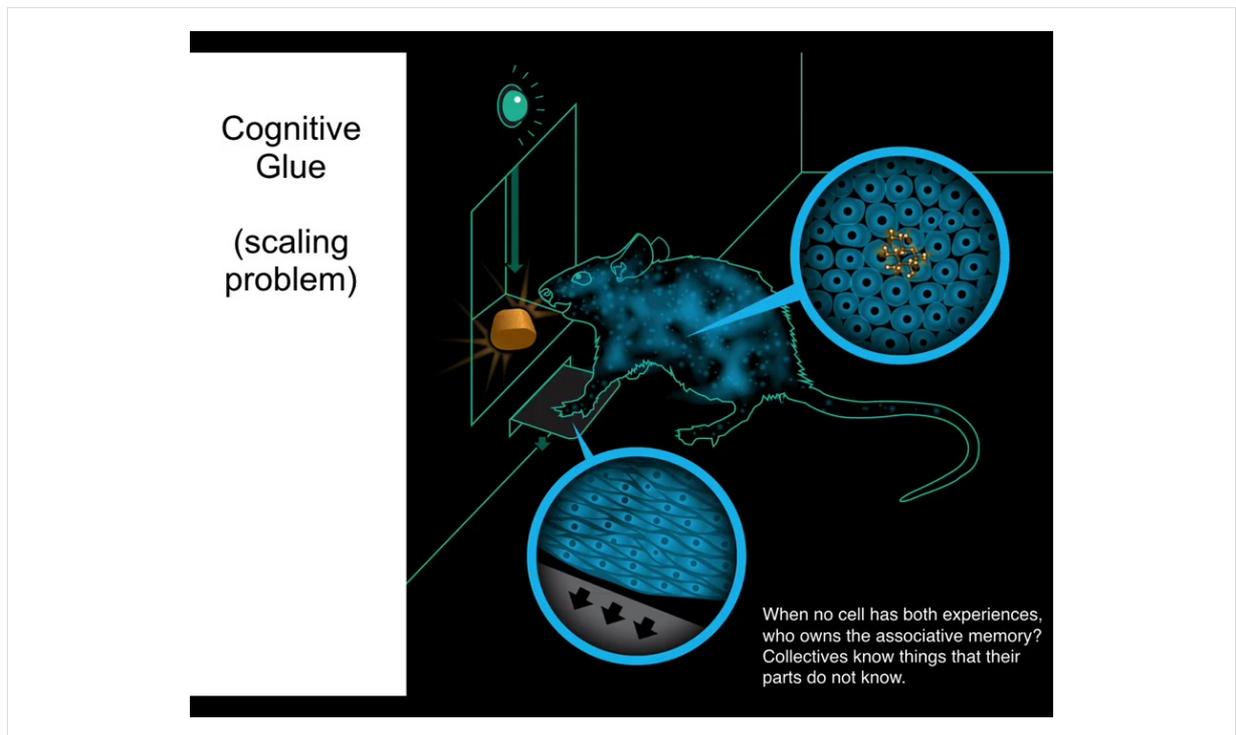
The newt that you get with these giant cells is exactly the same size as the original newt, which means that there have to be fewer cells to make up the exact same tubule. Now the whole structure adjusts to the increased size of the cells. Even more amazing: the final thing is that if you make the cells so gigantic that there isn't room for more than one around the structure, they bend around themselves, leaving a hole in the middle, and make the whole thing out of just one cell.

Look at what's happening here. There are a couple of interesting things. First of all, there's interesting downward causation where in the service of this very large-scale anatomical structure, you are using different underlying molecular mechanisms. This is cell-to-cell communication. This is cytoskeletal bending.

This is intelligence where you are confronted with a problem you've never seen before, and what you are able to do is figure out how to use the tools you have to solve this problem. There are numerous molecular components. Every newt you've ever seen in the wild is doing this, but in this remarkable case where somebody does this process involving pressure to magnify the copy number, they can figure out that there's a different way to get to that same goal and give you that same newt.

Think about it: you're a newt coming into this world. What can you rely on? You can't rely on your environment because we all know environments change. You can't even rely on your own parts. You don't know how many copies of your genome you're going to have. You don't know the size of your cells. You don't know how many cells you're going to have.

You have to do something interesting, and I'll get to this at the end of the talk again, which is to creatively get to your goals despite the unreliability of even your own parts and, in fact, your evolutionary history. All the things that have happened in the past are not necessarily a great guide to what's happening now. This idea has many implications.

When no cell has both experiences, who owns the associative memory? Collectives know things that their parts do not know.

I hope I've convinced you that morphogenesis is not just about complexity, it is about problem solving and in fact creative problem solving. That requires us to understand how all of that emerges from the competent cells that make it up. What we're looking for here is a kind of cognitive glue. We're looking for policies and mechanisms that help us to overcome the scaling problem.

Now in neuroscience we know what's going on. Here's a rat that's been trained to press a lever and get a reward. The cells at the bottom of its feet interact with the lever. The cells in the gut get the delicious reward. No individual cell has had both experiences. Who owns the associative memory? The rat does. There's this emergent being that has memories that do not belong to any of its components alone. We know that what holds that being together in conventional contexts is bioelectricity.
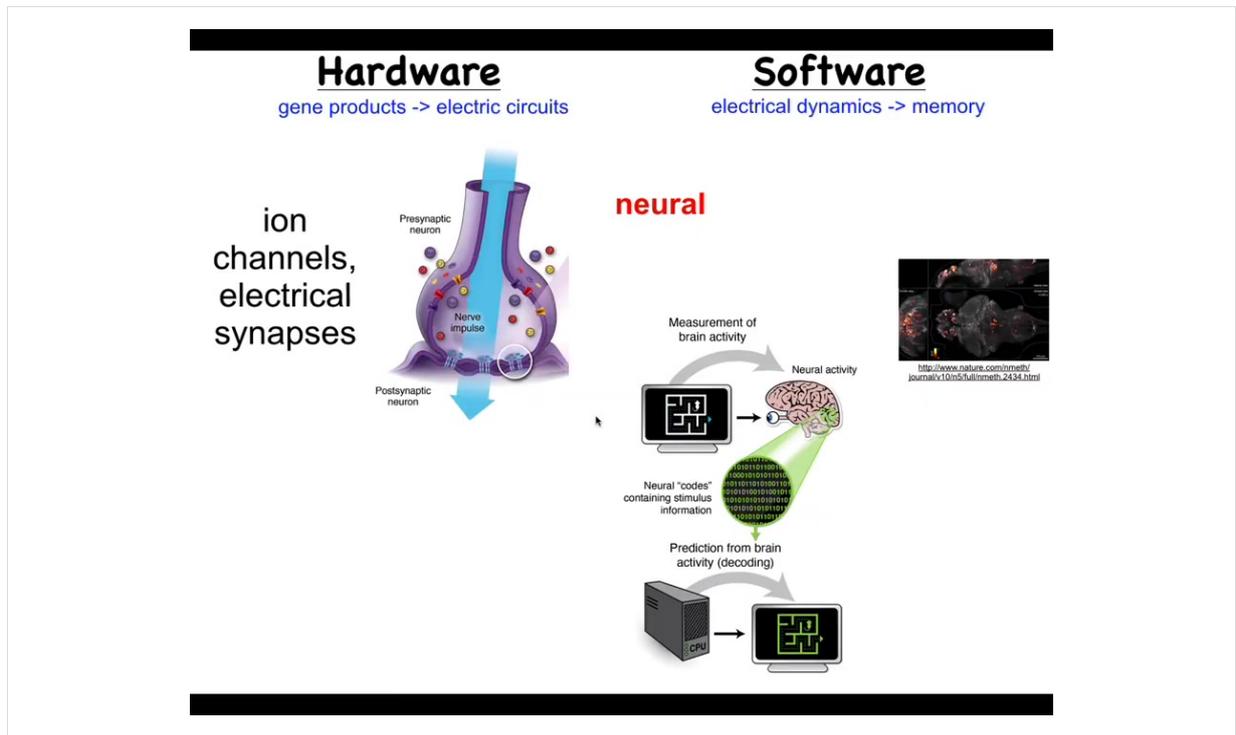
The nervous system uses this architecture to accomplish that amazing feat of taking a whole bunch of neurons and making a collective intelligence out of them.

There are these ion channels in the cell membrane that establish a voltage gradient across them. These voltage states may or may not propagate across the gap junction. These are electrical connections. And that system runs this software, which this group here is visualizing in a living zebrafish.

It is the commitment of neuroscience generally that we should be able to do neural decoding. We should be able to read this electrophysiology and decode it and know what the animal is thinking, what memories it has. The idea is that all of its cognitive structure is in some way encoded in this electrophysiological activity.
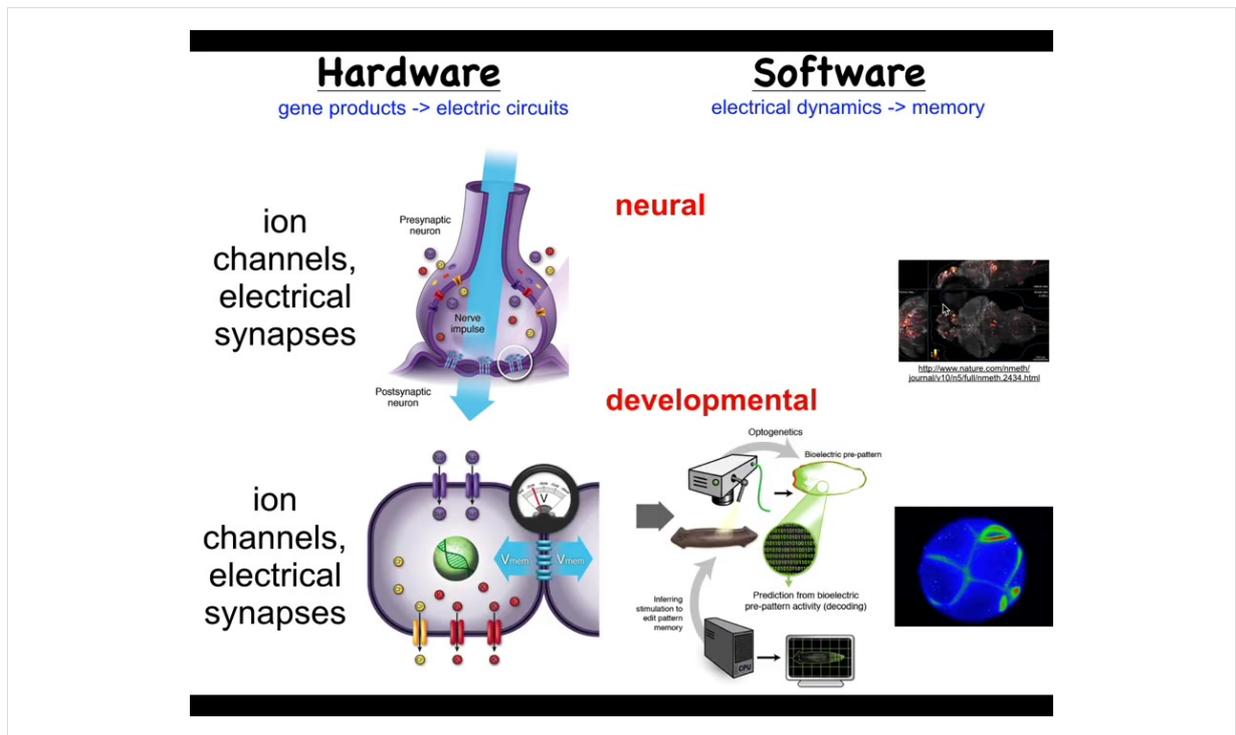
It turns out that the utility of electrical networks for integrating and aligning subunits into higher level computational structures was noticed by evolution long before brains showed up. Around the time of bacterial biofilms this first got going. Every cell in your body has these ion channels. Most cells have electrical connections to their neighbors.

What we started to do was to borrow a lot of ideas from neuroscience and to ask, could we do neural decoding here, except in non-neural cells? We know what these networks think about. They think about moving your body through three-dimensional space. What do these networks think about? It turns out there's a really strong symmetry here between development and cognition. We exploited that in a couple of ways.

How we detect and model bioelectric signals:

Characterization of endogenous voltage gradients - direct measurement and correlation with morphogenetic events
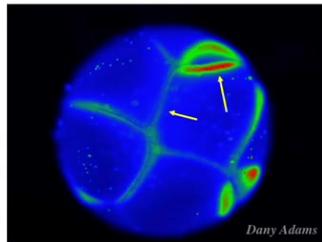
Voltage reporting fluorescent dye in time-lapse during *Xenopus development*

Dany Adams

Quantitative computer simulation: synthesize biophysical and genetic data into predictive, quantitative, often non-linear models
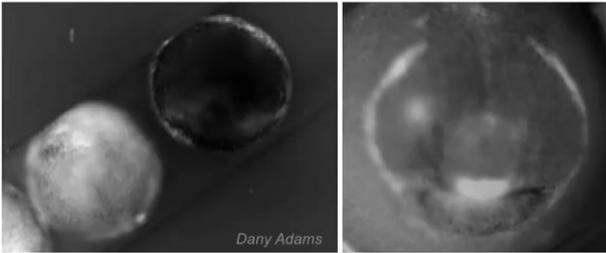
Alexis Pietak
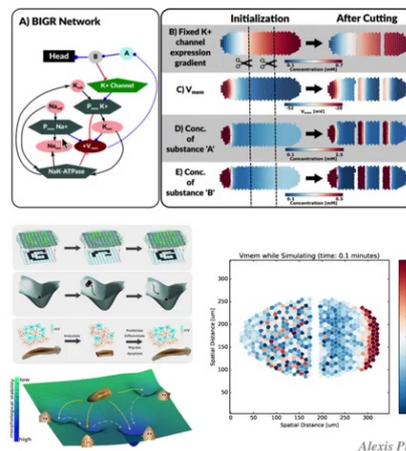
First we developed tools. These are the first molecular-level tools to read and write the information content of these electrical networks. The claim here is that groups of cells are a collective intelligence whose behavior plays out in anatomical space. What these somatic ancient networks are doing is thinking about where you should be, where the configuration of your body should be in anatomical amorphous space. Thinking about it that way allows us to deploy all sorts of tools from behavior science and computational neuroscience and try to understand how this works.

We developed these imaging kinds of processes. Danny Adams made this video using a voltage-sensitive fluorescent dye, and you can see all the electrical conversations that these cells are having with each other, and they're deciding who's going to be anterior, posterior, left, right. We do a lot of computation, so Alexis Pytak made this amazing simulator, and we try to integrate from the expression of these ion channels all the way through the electrical dynamics of the circuits to the large-scale patterns and what happens during regeneration, for example, to understand pattern completion, the way that these electrical circuits can actually restore missing information if something is damaged.

29

To show you what these look like, this is a time-lapse video also made by Danny. This is the craniofacial morphogenesis of an early frog embryo, the face is getting set up here. This is 1 frame from that video. What you're watching here, the colors or the grayscale indicate the voltage of each cell, the resting potential of each cell. These are not spiking the way that neurons spike. These are steady. To turn almost any study in neuroscience into a developmental biology paper, you can replace the word neuron with the word cell. Where it says milliseconds, you say hours and everything else maps very nicely.

You can see here's where the animal's right eye is going to be, here's where the mouth is going to be, the placodes, all of this. As I'll show you momentarily, this is determinative of what these cells are going to do in the future as they build the frog face. This is an endogenous pattern. It is required for normal face development.

This is a pathological pattern where we inject a human oncogene into these animals. It will eventually make a tumor. Even before it does that, you can see with this voltage map that these cells are electrically isolated from their neighbors. They've disconnected, and now they're amoebas as far as they're concerned. Everything else around them is external environment. We'll talk about that momentarily. This is the work of Brooke Chernett and my group.

Tracking those kinds of things is nice, but what's really important are the functional tools. How do we actually change that information to be able to now write directly into the cognitive system of this morphogenetic intelligence? We do not use any applied fields. There are no electrodes. There are no magnets, no waves, no frequencies. What we do is exactly what neuroscientists do, which is to hack the interface that these cells are normally using to control each other. We can open and close these gap junctions; that controls the topology of the network. We can open and close these various ion channels using drugs, optogenetics, or any of those tools. We can set the distribution of electric potentials, or we can set the topology of which cell talks to which other cell.

Now comes the real question. Having done that, what can we do with this? How do we prove that these electrical patterns are determinative of morphogenesis? How do we communicate new goals to this system?

I showed you this little voltage spot encodes the message of building an eye. One thing you might imagine is what happens if we recapitulate this pattern somewhere else?

31

Here's how we did that. This is the work of Sherry Au and Vipov Pai. We inject potassium channel RNA into a particular region. In this case, there's a region that's going to give rise to this gut tissue. What this does is set the voltage and establish a little spot that says build an eye here. Sure enough, these gut cells immediately build an eye. These eyes have the same components: retina, lens, optic nerve, as normal eyes.

Now we learn a couple of things from this kind of experiment. We can make eyes, hearts, brains, a few other organs. What you learn is this: first of all, these electrical patterns are instructive. They tell the cells what to build. They're not epiphenomena; they're functionally determinative. We are dealing with the system that controls the behavior.

Number 2, it's extremely modular. This is a simple stimulus that kicks off a complex set of responses. We didn't tell these cells how to build an eye. We have no idea how to build an eye. It's too complicated. What we said was a high-level subroutine call that says build an eye here. That's it. Everything else is encapsulated.

If you read the developmental biology textbook, they will tell you that only these cells up here, the anterior ectoderm, are competent to make an eye. That is because they're using a prompt called PAX6. PAX6 is a transcription factor. It's a biochemical signal that induces eyes. True enough, it only works up here in the anterior ectoderm.

This, I think, is an important point for all of us working in this field, which is that any estimate of the intelligence of a system is basically just us taking an IQ test ourselves.

32

All you're saying there is: what have we figured out how to get the system to do? If you stick to these kinds of biochemical prompts, it's true the cells do not look competent to make an eye out here. If you use a much better prompt that is more salient to the system itself, which is this bioelectric state, then you find out that pretty much every region of the body can make an eye if correctly communicated with. That speaks to the importance of not assuming that we're right when we find limits in these kinds of systems.

The final thing that I think is pretty cool is that this is one of the competencies of the material. This is a cross-section of a lens sitting out in the tail somewhere of the tadpole. The blue cells are the ones we injected. All this other stuff is recruited by the blue cells to participate in this project. They know there's not enough of them to make a proper eye. We instruct these cells: make an eye. They automatically do the rest by instructing enough of these other cells to help them complete the project, much like some other collective intelligences recruit their nest mates to carry heavier loads. This is just one example.
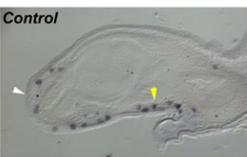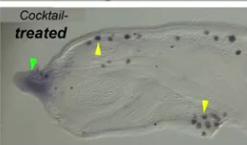
I won't dwell on all the biomedical stuff.

Brief bioelectric signals trigger long-term, self-limiting **modules** (low info-content input, high info-content output)

Hind-leg amputation
+
designed ionophore
cocktail regimen

The regenerated leg has both sensation and mobility:

Control

Control
45 days post-amputation

Cocktail-
treated

Cocktail
treated
45 dpa

MSX1 marker -
blastema induced

Outgrowth with
distal patterning induced
(and still growing)

33

We have a regenerative medicine program where we try to induce animals that don't regenerate their legs, like frogs, to regenerate their legs using a very simple, very brief trigger of a voltage pattern that says, go to the leg making region, do not go to the scarring region. You can see eventually you get this quite nice leg. It's touch-sensitive and motile.

I have to do a disclosure because Dave Kaplan and I have a company called Morphoceuticals, which is trying to move this now to mammals and eventually to humans. The idea, using these biodome delivery devices, is not to micromanage it, not scaffolds, not stem cell controls, nothing like that, but to convince the cells that this is what they should be doing, a simple, high-level signal at the beginning.
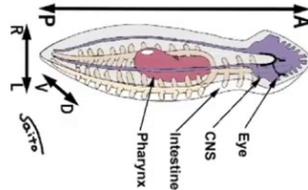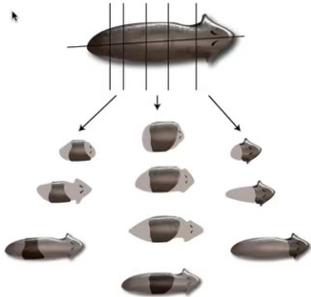
I want to switch and introduce you to a different model system in which we can really see something very interesting about the way that biological tissues use electricity to store memories. So this is a planarian. It's a flatworm. It has some really incredible properties, including immortality and cancer-resistant. If people have questions, we can talk about that in the Q&A.

What we're interested in here is the fact that if you chop these guys into pieces, every piece makes a perfect little worm. They're extremely regenerative. And we asked the simple question, in a piece like this, how does it know how many heads to have? Why does this wound form a head? Because if you actually look at a single cut, the cells on this side of the cut will make a head. The cells on that side of that cut are going to make a tail. They were direct neighbors before you cut them apart. How come they have these widely differing anatomical fates? How do they know what they should do and how many heads they should have?

We did some of this bioelectric profiling and we found something interesting, that this fragment actually contains a bioelectrical pattern that says how many heads. The depolarized regions tell you how many heads you should have and where they should be.

What we were able to do then, and it's a little messy still, the technology is still being worked out, but what you can do is introduce two of these regions. Sure enough, when the cells consult that pattern, they say, two heads, and they build an animal with two heads. This is not Photoshop or AI. These are real creatures.

Here's something very important. Two key things here. First of all, what we are doing here is literally reading the set points, AKA pattern memories, of this collective intelligence. That is, my claim is that these cells work together to maintain a memory of where in anatomical space they should go. Using these techniques, we can — this is the neural decoding, except it's not neural. We are reading directly the information that encodes this.

We know it does because if you change it, the cells build something different. This bioelectrical map is not a voltage map of this two-headed creature. This map is of this perfectly normal looking one-headed animal. One head, one tail, the molecular biology bears it out, the head marker in the front, no head marker in the tail. Anatomically and molecularly, this is a perfectly normal animal.

But he's got one thing that's wrong with him: the cells have a weird internal belief that a correct planarian should have two heads. If you injure it, meaning you cut off the head and the tail, they will then go ahead and build these two heads. It is a latent memory because until you injure it, nothing happens. He just sits there being normal until he gets injured. That's when you find out that he's got a completely different internal representation of what makes for a normal worm.

This is a simple version of a counterfactual. It's a simple version of our brain's ability to have mental time travel, where we can think about and remember and anticipate things that are not true right now. This state is not true right now, but that is what is going to guide you if you get injured in the future. The same body can store at least two different representations of what I wish to do if I get injured in the future. I think this could be a simple model for thinking about how you can store counterfactuals in this kind of collective.

Now, I keep calling it a memory. If we were to take these two-headed worms and cut them again and again in perpetuity, no more manipulations of any kind, just plain water, what you'll see is that once changed, that idea of how many heads you should have stays. That electrical state is kept. So this has all the properties of memory. It's long-term stable. It is rewritable. It has conditional recall, which I just showed you, and it has some discrete behaviors, one head versus two heads. We can set it back eventually, but once you change it, it stays.

So here are these two-headed worms hanging out. You can see what they do. Remember that in all of this, there is nothing wrong with this hardware. We did not change the genome. There are no transgenes here. This is purely a short experience in a change in the bioelectrical state of the cells that from then on there should be two heads. So no amount of sequencing, no amount of characterization of the hardware, the proteins, the RNAs, none of that will tell you what's different here, because it is not at that level.

Now, realizing then that what we're changing here is the information that guides the movement of this agent through anatomical space, we can wonder where else we can tell it to go. This controls head number, but what else can we do?
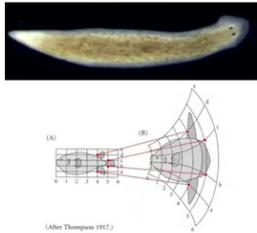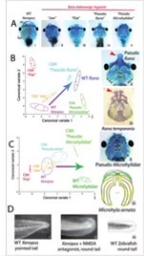
So it turns out that the other thing that is available in this morphospace are different shapes of heads. There are attractors in this space that belong to different species that have different shapes of heads. And so now the question is, can we get the same hardware?

So here's Dugesia dorotocephala, nice triangular head, little auricles here. Could we get these cells to build one of these other species and get them to visit a different region of that space? And it turns out you can't. So you cut off the head, you perturb the bioelectrical gradient, and you can make flat heads like a piphalena. You can make round heads like S. mediterranea. Not just the shape of the head, but actually the shape of the brain and the distribution of stem cells become like these other creatures. There is between 100 and 150 million years of evolutionary distance between these guys and him. And there is nothing genetically wrong, but these cells are perfectly capable of visiting these other regions in anatomical space.

Targeting the Morphospace-navigating Circuits

You can go further and visit regions that are not used by any planaria, as far as we know. Here's this crazy spiky form. Here are these cylindrical shapes. Here is a hybrid shape. One of the cool things about exploring the symmetry between developmental biology and cognitive science is that we get to use all their tools. We can take anxiolytics, hallucinogens, plastogens, all of these neat compounds that are used, and we can ask interesting questions: what happens when the morphogenetic agent hallucinates, when you distort its perception of its environment and its memory patterns? You can get frogs of a specific species to build heads that are appropriate to a different species. You can get them to make zebrafish tails instead of normal frog tails and so on.

The interesting thing is that we humans are not the only hackers who do this. There's a non-human bioengineer, this wasp, that prompts these leaf cells to build this kind of crazy structure. This is not made by the wasp, this is made by the leaves. We would have never known in a million years that these flat green cells, which do this so reliably in every oak tree in the world, are competent to do this if evolution hadn't enabled this guy to push them to do it. Now the question is: it probably took a long time for this lineage to learn to do that. Could we use modern tools and maybe AI to start to exploit these competencies rationally?

The final piece of this. I want to come back now that you've seen examples of morphogenetic problem solving, you've seen the bioelectric interface as both a cognitive glue and thus a way to communicate with it for biomedical purposes, regeneration. We can repair birth defects and some things. I want to come around and talk about the big picture of what is all this telling us about different kinds of cognitive systems.

The first thing that I like to keep an eye on is what I call the system's cognitive light cone. And that is a way to estimate the size of the biggest goals that a given agent can pursue. This is almost like a Minkowski diagram: you flatten all of space on this axis, time is here, and you can draw not the reach, the sensory motor reach of the system, but actually the size of the goal state towards which they actively work. Another way to put it is what kinds of states make it stressed when those conditions are not met.

You can think about what ticks care about, what dogs care about, how far forward they could have a goal spatially, and then humans. One thing you can do with this is you can plot out the different sizes of different cognitive light cones for all kinds of systems, no matter what they're made of, no matter how they got here.
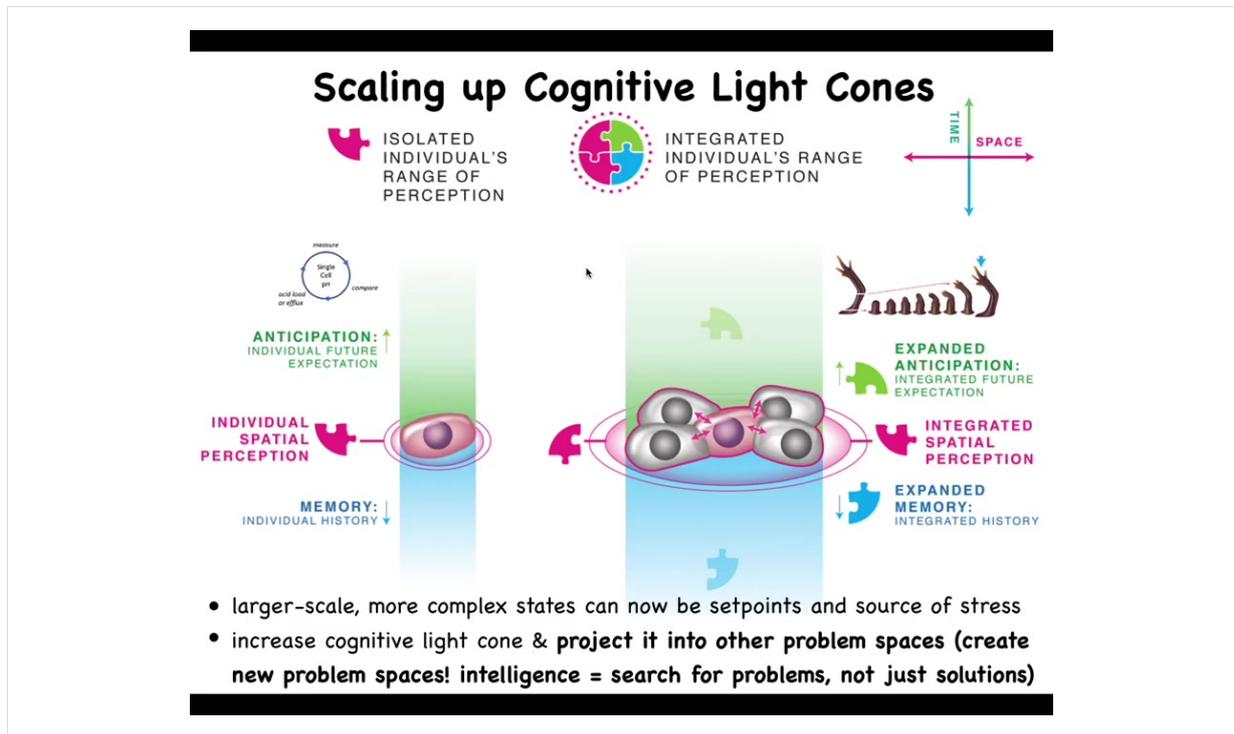
And so what's interesting is to think about the scaling. So what the biology teaches us is that a single cell might have a very small cognitive light cone and all it cares about is its own parameters, pH, hunger level, some other things like that. So very simple, very tiny little goals. But together in a network, in particular in an electrical network, but electricity isn't magic. There's many other modalities you could use. You could network them, which immediately widens the cognitive light cone.

41
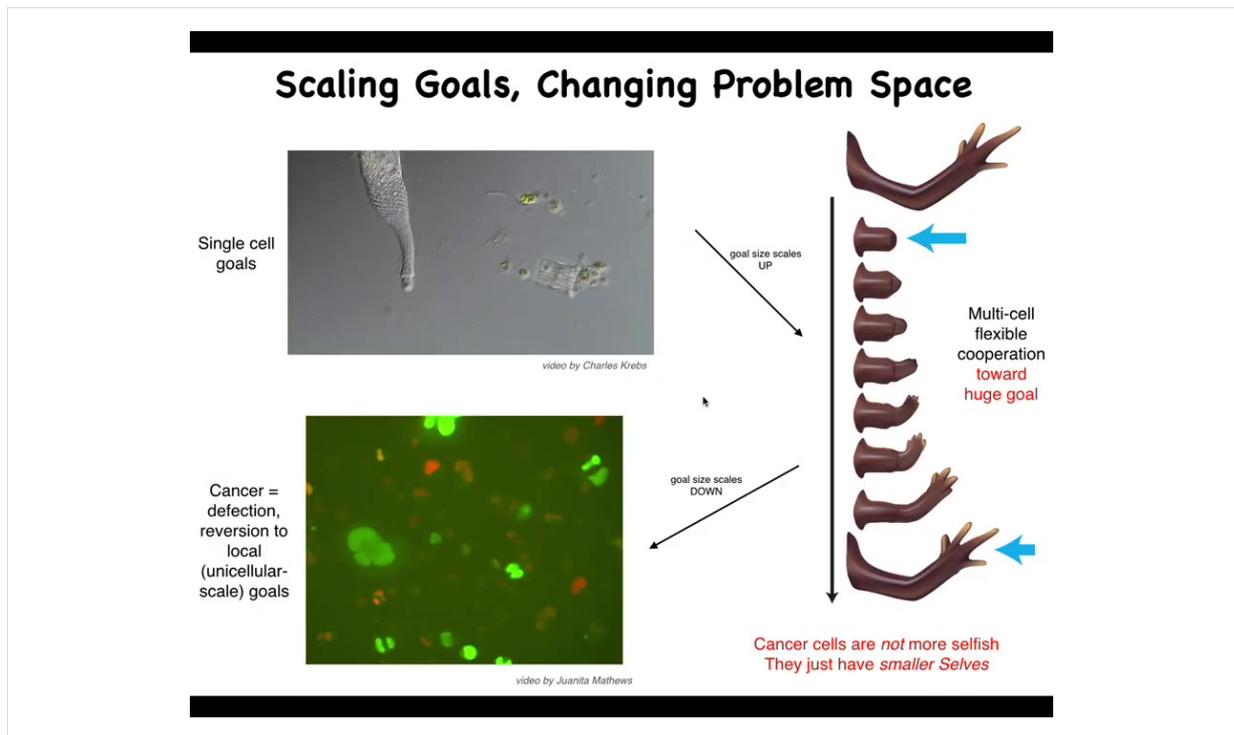
Here's a simple example. Here's a single cell pursuing its own goals. Evolution and development allows these systems to scale up to these incredibly grandiose constructions.

All of these cells are working towards this massive set point in the anatomical space. In fact, many of them will die doing it. They're all aligned and committed to this cause. They're all going to work together. If you deviate them, they will figure out some other way to do it, and they maintain this goal. That enlargement of the cognitive light cone, the goal here is huge. The goal state here is much smaller. That enlargement has a failure mode, and that failure mode is cancer.

This is human glioblastoma. This is the work from a number of people in the lab, but this right here, Juanita Matthews, is studying human cancer. What happens here is that when cells disconnect from this network, they can no longer remember this enormous goal that they should be pursuing. They go back to here. They roll back to their ancient unicellular form of life. The rest of the body to them is environment.

That border between self and world, which has to be set by any agent, is flexible, and it shrinks, and it becomes down to the size of a single cell. These cancer cells are not more selfish. There's a lot of game theory, for example, that models cancer cells as uncooperative and selfish. I don't think they're more selfish. I think they have smaller selves. They just collapse. That boundary between self and world has now shrunk.
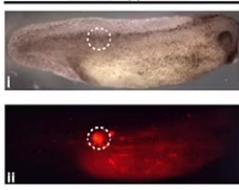
42
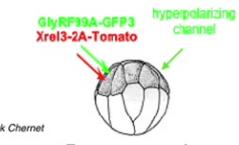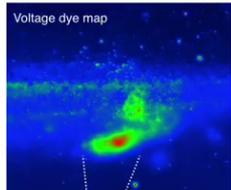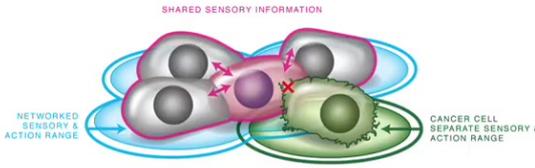
So now that philosophical idea has specific implications for a biomedical program, which is that maybe then for cancer, instead of having toxic chemotherapy to try to kill these cells, what if we try to simply connect them back to the rest of their neighbors so that they can rejoin this collective?

If you do that, here we inject nasty oncogenes, KRAS mutations, p53 mutants. Here we've labeled the oncoprotein in red, so you can see fluorescently; it's all over the place. This is the same animal. There's no tumor. The reason is, even though we haven't killed the cells, we haven't removed or fixed the oncoprotein, it's still there, but because the cell is connected to this larger network, it works towards what the whole system works towards, which is making nice skin, nice muscle, and so on.

So at least in some cases, and we've seen this in birth defects, we've seen this in regeneration, and now here in cancer: in software, you can override certain kinds of hardware defects. I don't mean all of them, but in this case, there's a real genetic problem that you would discover if you were to sequence this. What you wouldn't know is that there isn't actually a tumor because the network bends the option space for the components. They will not make a tumor and metastasize. They continue their morphogenetic cascade.

So what we've seen here is the ability of this collective to store large-scale set points.

Now we can ask, where do these set points come from? Who sets them? How do they get set?

43

- New ways to test intelligence in other spaces, other embodiments
- New ways to communicate, not micromanage

Next:

- Why so plastic?
- Where do goals come from
- Evolutionary implications

We've looked at new ways to test intelligence in unfamiliar spaces. We've looked at new ways to communicate goals to these systems, not to micromanage them. For the last couple of minutes, I want to address two questions. At the end, if anybody's interested, we can also dig into some evolutionary implications.

The first is why are they so plastic? Why are these animals so incredibly plastic? With the exception of a few things, nematodes — C. elegans, where every cell is numbered and they're pretty paint-by-numbers. Most other creatures appear to be incredibly plastic. Why is that? The question is where do these goals come from?

I want to point out something interesting about the information flow in living systems, both on the cognitive scale and on the evolutionary scale. You can think about the fact that at any given point in time, you don't have access to the past. What you have access to are the engrams, the memory traces that past events have left in your brain. You can think about your own memories as instances of communication, messages from your past self. The same way that you communicate laterally with other creatures at the current time, you also receive communications from your past self and you leave messages for your future self.

It's an interesting way of thinking about it because what's happening here is that all of these past experiences are being compressed because you're learning and you're abstracting patterns. You don't remember every single microstate that you've experienced. You compress them into a compressed representation. That representation is the n-gram that you get at any given moment. Now your job is to re-inflate that representation into some meaningful data structure that is actionable right now for whatever you need to do next. Like with any other message from another agent, you are under no obligation to interpret that message in exactly the same interpretation with which it was sent. At any given moment, it's up to you how you're going to interpret. It is required for you to be creative, because while this process, the encoding process, is somewhat algorithmic and deductive, you're throwing away detail and correlations. In order to decompress it, you have to be creative. You have to ask: you don't know what the original meaning was, but you have this information. What are you going to do with it?

Not only does this happen in cognition, this also happens in evolution. As I showed you with that newt, there is an autoencoder architecture here, where there's a bottleneck that requires this side to be more creative; this is true for the body as well as the mind. Everything gets squeezed down into the egg: the materials, including the genes, but also a bunch of other stuff. Then that has to get re-inflated. As I showed you with the example of the newt and some other things I'm going to show you momentarily, that is not a hardwired process. You do not have to do it exactly the same way as past generations have done it. That leads to an interesting intelligence ratchet. All of that is described here, if you're interested in that.

So here's a couple of consequences of this idea that, much like with that newt, you really can't count on the past. You can't take it too seriously. What you have to do is interpret what you have in the best way that you possibly can, given whatever is happening now.

Slide 44 of 58 · Watch at 56:56

The fact that you're dealing with this unreliable medium, a fundamentally unreliable medium, where living things in evolution, not only is the environment going to change, but you're going to change. All your parts are going to be different. You're going to be

46

mutated. There's going to be all sorts of things happening. You cannot take the past literally.

Here are some examples. This is a caterpillar, and these caterpillars can be trained. They have this brain suitable for doing what they do. They're trained to eat leaves on a particular color background; this is the work of Doug Blackiston. Then they have to become a butterfly. You go from a creature that lives in a two-dimensional world to a creature that flies. Caterpillars are soft-bodied and have to move in a very particular way because they're soft. A butterfly needs a hard-body kind of controller that flies and so on. And so it has a different brain. During metamorphosis, the brain is massively refactored. Many of the cells, if not most of the cells, are killed off, the connections are broken, it reassembles a new brain. These guys still remember the original information.

The obvious question is how do memories survive this incredible refactoring? The deeper question is that the exact memories of the caterpillar are of absolutely no use to the butterfly. The butterfly does not move the way the caterpillar does. They don't want leaves. Butterflies want nectar. The memories cannot be just kept. That's not enough. You have to translate them. You have to convert them. This is your bottleneck. You have to convert them into something that makes sense at this point, that's adaptive. The idea is to emphasize salience, not a fidelity of the original information, because everything is going to change anyway. You can't rely on keeping the original information exactly intact.
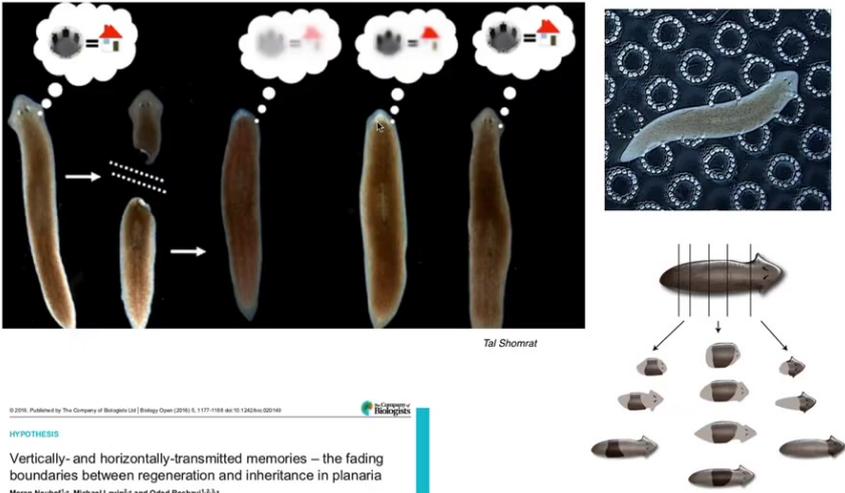
In planaria, you can watch the information move across tissue if you train them to eat liver in a particular bumpy environment. You then cut their heads off, which is where their brain is. Their centralized brain is here. The tail sits there doing nothing. It regenerates a new brain. At this point, you find out that the information has now been imprinted on the brand new brain, and they now remember where to find it. So the information can move across the body, but it can also move across radically different bodies, as I just showed you, which requires a lot of reinterpretation.
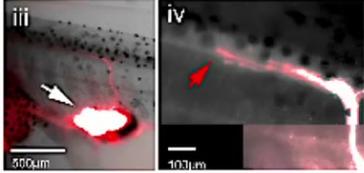
We see that in these kinds of cases where if we make a tadpole that has no eyes and we stick an eye on its tail, these animals can see perfectly well. We know because we have a machine that trains them in visual assays.

These eyes make an optic nerve. The optic nerve does not go to the brain. It goes, maybe, to the spinal cord and ends there. Maybe it goes to the gut. That's it.

Why does this work out of the box? Why does this animal with a radically different sensory-motor architecture not require new rounds of selection and modification in order to be able to see in this crazy configuration?

I think that's because the standard tadpole, even though we couldn't see it under standard circumstances, was never assumed to be hardwired for specific facts anyway. It's basically a sense-making system that emerges from scratch.
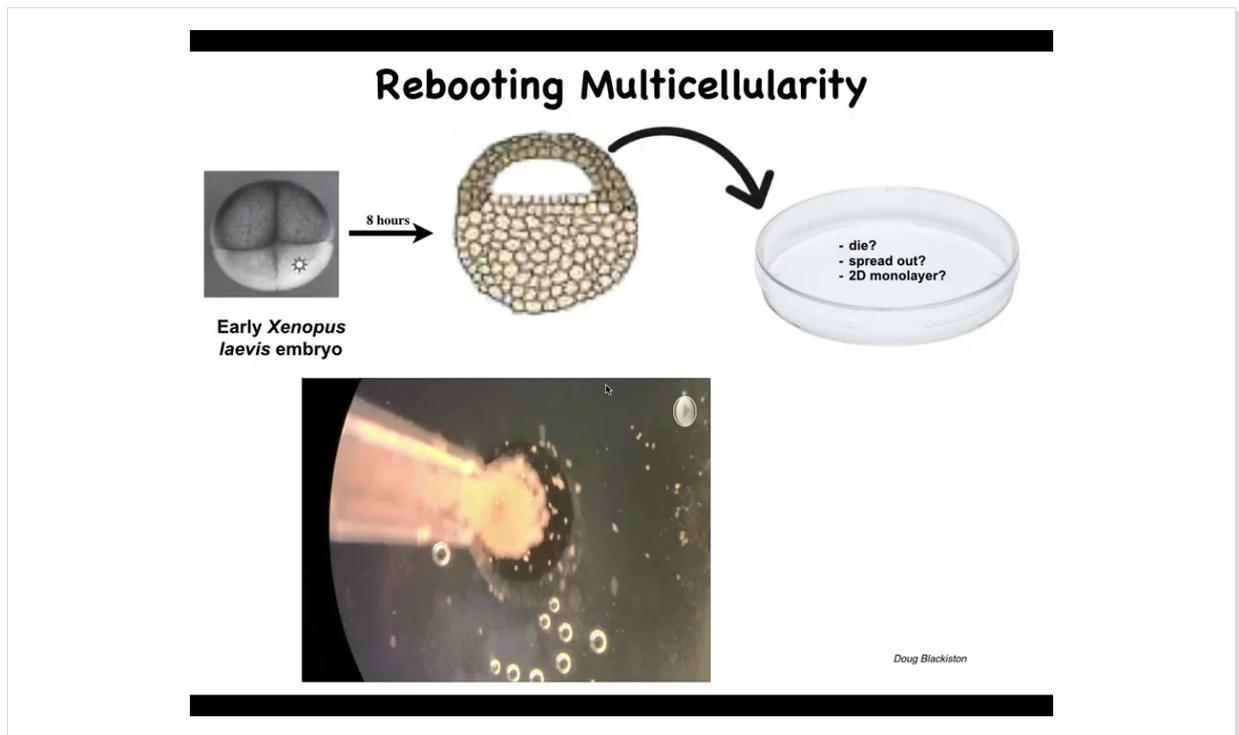
Here's the final thing I want to show you. I'm almost done. We started to ask, how does this plasticity work out for creatures that have never been here before? We tried to make some novel multicellular beings.

Here's a frog embryo. Here are some epithelial cells, cells that normally would make the outer skin surface of the animal. We liberate some of those from the rest of the animal. They could have done many things. They could have died. They could have spread out, gone away from each other. They could have made a flat monolayer. But instead what they do is they come together and make this cool little thing. The flashes that you see are calcium fluorescence.

What they make is something we call Xenobots. Xenopus laevis is the name of the frog and we think this is a bio-robotics platform. Xenobots. What they're doing is they're autonomously moving because they have little hairs that used to wash the mucus down the side of the animal, but now they're using it to row against the water. They're swimming along. They can go in circles. They can patrol back and forth like this. They have group behaviors.

Xenobot in a maze (still water, no flow):
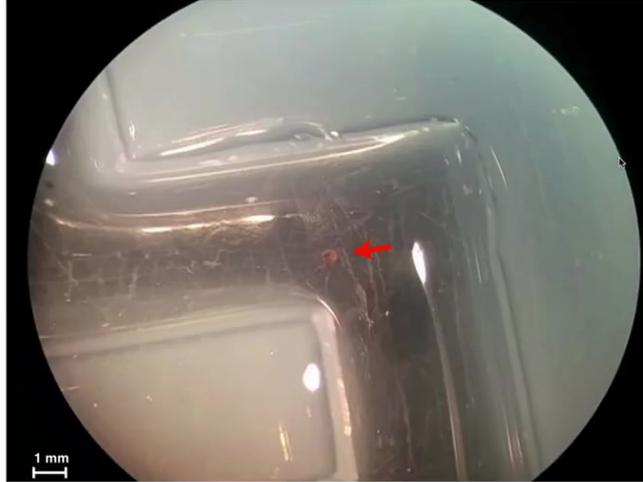
1) it traverses maze,   2) rounds the corners without bumping into walls, and
3) it makes a spontaneous decision to turn around without hitting anything.

Doug Blackiston

Here's one navigating or traversing a maze. It's going to take a corner without bumping into the opposite wall, so it takes that corner. Then here, for some reason, it spontaneously turns around and goes back where it comes from. A wide range of behaviors.

If you look at the calcium signaling, you see some very interesting things. I'm not going to make any conclusions yet, but we're analyzing these now using the same tools that neuroscientists use to analyze calcium signaling in the brain. But remember, there are no neurons here. This is just skin.

Kinematic Replication in Xenobots:

evolution, like engineers, works in agential materials

Doug Blackiston

gen 0 | gen 1 | gen 2 | gen 3
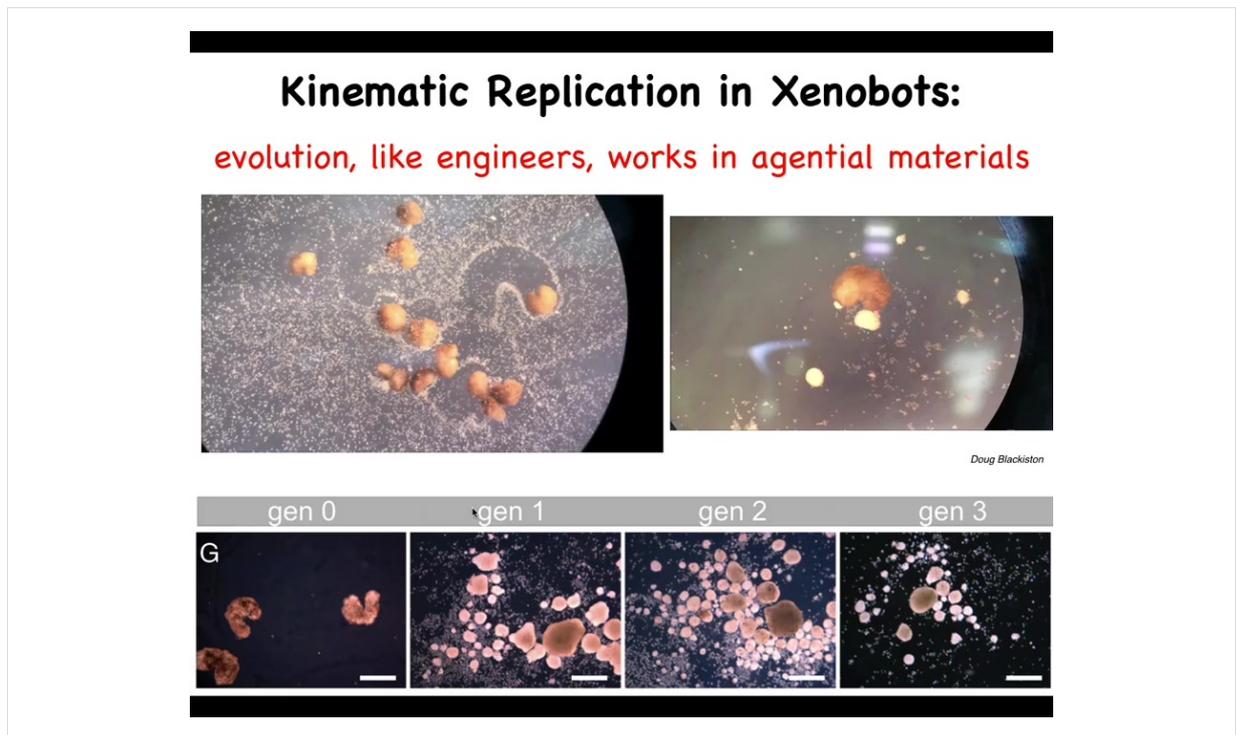
This remarkable capacity they have is to reproduce. They can't reproduce in the normal ****** fashion because we made that impossible. But if you give them a bunch of loose skin cells, that's what this white stuff is, both together and individually, they collect it into little piles, they polish the little piles, and the little piles mature into the next generation of Xenobot. They run around and do exactly the same thing, which makes the next generation. This is Kinematic replication, and it works because the material they're working with is also not passive, but itself an agential material. You get this thing that's von Neumann's dream of an agent that goes around and makes copies of itself from parts it finds nearby.

Where did that come from? There's never been any xenobots. There's never been any selection to be a good xenobot. The frog genome, we tend to think, has learned to do this throughout development and has this developmental sequence, and then eventually it makes this, and this is what it's learned to be a good frog in a ****** environment. But if you liberate these cells from the influence, from the hacking by the other cells that force them to be a boring two-dimensional bacterial suppressing layer around the animal, they actually have their own very different life. They can move. They do this crazy developmental thing. This is an 80-day-old xenobot. It's turning into something. I have no idea what it's turning into. It has different behaviors. We are now studying their learning. Stay tuned for that. But I think we're starting to see here that what you can't do is the thing that you would normally do, if you're wondering about the properties and the behaviors of a particular animal or plant, is to say, well, eons of selection, that's where it comes from. These things are baked in by selection for specific things. There's never been any xenobots.
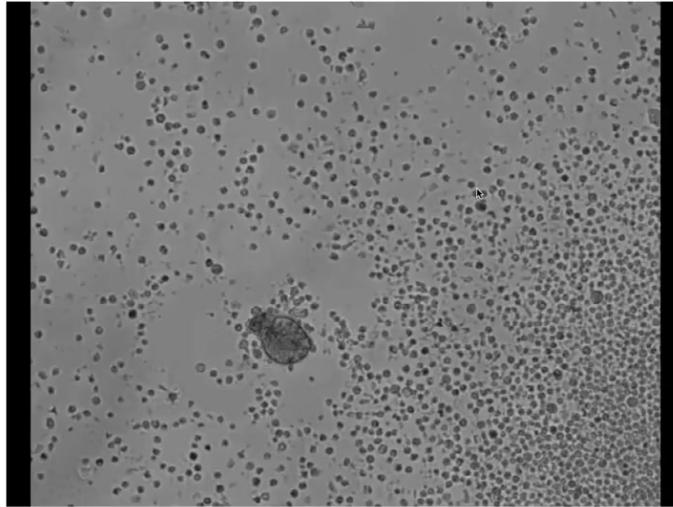
And then this is, in case you think this is some weird embryonic frog thing, I can show you this. This might look like something you got out of the bottom of a pond somewhere. If you were to sequence the genome here, what you would see is 100% Homo sapiens. These are adult human patient cells. They are tracheal epithelial cells in this environment. They assemble themselves into a multicell bot. It has a different transcriptome that has thousands of new genes that are differently expressed than in the native tissue from the patient. It has new behaviors. I don't have time. It has all kinds of other things it can do.

So what I think is going on here is that much like simple physical devices, triangles and the Archimedean machines and so on, are embodying aspects of what mathematicians call the contents of a Platonic space, what we are doing when we make anthrobots and xenobots and those kinds of things is we are making vehicles to explore exactly that latent space. I think where they come from is exactly where all these other patterns in nature come from that are not specifically baked in or evolved or anything else. So I think that by making these kinds of synthetic constructions, we actually start to get, we can start to map it out. What are some of the affordances that exist in that latent space?
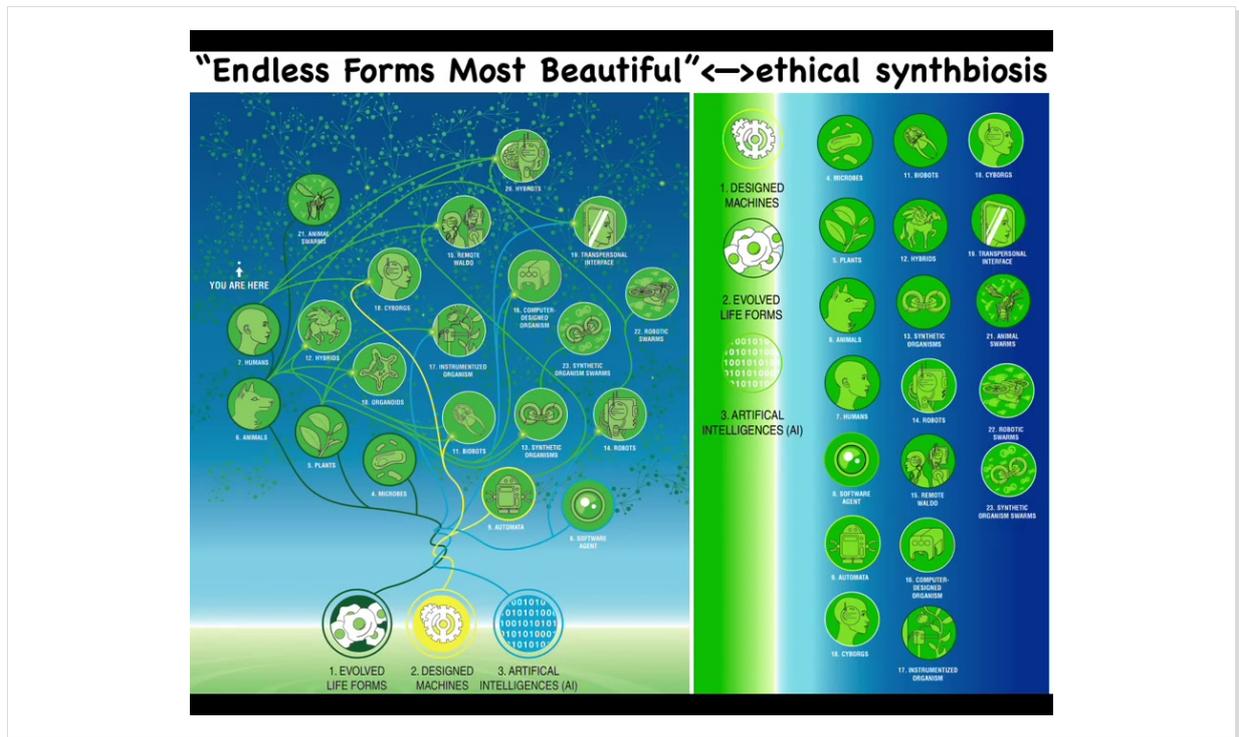
I think it's important because everything that Darwin meant by "endless forms most beautiful" is a tiny corner of this option space. Cyborgs and hybrids and chimeras, any combination of evolved material, natural material, and software, are agents. Many of these things already exist. There are going to be way more of them, and we need to understand how we are going to enter some ethical synth biosis with other embodied minds that are very different from what we're used to.

I just want to mention one thing: I think a lot of humility is warranted because I won't go into detail unless somebody wants to ask, but we found not just complexity, but primitive problem solving and emergent goal-directedness in very simple things — in this case, sorting algorithms, the kinds of things people have been studying for many decades, bubble sort and so on. You can see that here. I think we've really underestimated what matter can do. And I think for the same reason we underestimate what very simple algorithms can do and what machines can do. And so, just because we've made it and just because we think we understand some of the parts, I think we really do not understand well other capabilities, not just emergent complexity and unpredictability, but actually emergent cognition. And I think it begins very low on the spectrum of the kinds of things that we associate with.

I think intelligence is widespread. I think we have to learn to rise above our innate baseline limitations of how to recognize it. We can now have principled frameworks that avoid either assuming that there is no intelligence or conversely assuming high-level minds under every rock. I think we can do better than this now. We have all kinds of interesting opportunities in the future using AI and other tools.

If you want to dig into any of this, there are some papers here, and I want to thank all the people who did the work.

The bio bot work was done by Doug Blackiston and Gizem Gamushkaya; the ectopic eye work by Sherry Au and Vaipoff Pai; the unexpected competencies of algorithms with Tanning Zhang; Fallon Durant for the planarian work; Nirosha Murugan for the Faizarm and the leg regeneration work. Juanita Matthews is leading our cancer efforts. This is Sarama Biswas who did the memory in gene regulatory networks. We have lots of collaborators that contributed to all of this.

Our funders: here are the disclosures. These are three companies that have supported some of this work. All the heavy lifting is done by the model systems. I'll stop here and thank you for listening.

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.