

A LECTURE COMPANION

**Against Mind-Blindness: recognizing
and communicating with Agential
Materials by Michael Levin**

Michael Levin

Recorded on September 2, 2025

About this document

This document is a companion to the recorded lecture *Against Mind-Blindness: recognizing and communicating with Agential Materials* by Michael Levin, recorded on September 2, 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

This is a 1 hour 11 minute talk by me called “Against Mind-Blindness: recognizing and communicating with Agential Materials and Beyond”. It is a slightly longer, updated version of this talk from last month, containing a bit of new material and a somewhat differently-ordered presentation.

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.

Against Mind-Blindness: Recognizing and Communicating With Agential Materials and Beyond

Michael Levin
Allen Discovery Center at Tufts

<http://www.drmichaellevin.org/>
<http://thoughtforms.life/>



ALLEN DISCOVERY CENTER at Tufts University

Computer-designed Organisms
TUFTS UNIVERSITY | UNIVERSITY OF VERMONT

WYSS INSTITUTE
HARVARD

I run the Allen Discovery Center at Tufts University. My lab studies a variety of topics that I'm going to mention today. My background is computer science first, then biology, and I'm interested in many related issues at the boundaries.

If you're interested in getting in depth into any of the issues I'm going to mention today, this is our official lab website. You can find there the papers, the data sets, the software, everything is there. And this is my own personal blog around what I think some of these things actually mean.

What I'm going to talk about today is this issue that I call mind blindness, which is the difficulty we have as humans at recognizing unconventional intelligences that are all around us. We're going to talk about that.

**Sentient beings are innumerable,
I vow to liberate them all**

— from the Four Bodhisattva Vows



I find this particularly interesting. There's a piece of one of the four Buddhist vows that says, "Sentient beings are innumerable. I vow to liberate them all." There are several interesting pieces here. First of all, this notion of enumeration that, if something is going to be innumerable with an E, so able to be numbered, there needs to be some kind of effective procedure for recognizing them. In other words, there should be some algorithm or some way that we're going to find all of these sentient beings. You need to do that if you're going to liberate them or at the very least relate to them ethically. I would like to point out that we have a lot of difficulty doing that.

If you look at this close up, this is a giant pile of cutlery. These are knives, forks, spoons. When you look at this, at first glance, what you see is a pile of garbage. There's no order here, just a bunch of random metal silverware.

**Sentient beings are innumerable,
I vow to liberate them all**

— from the Four Bodhisattva Vows

Perspective is key
Cognitive claims are IQ tests
for observers



The image shows a sculpture made of many small, dark, irregular objects piled together. From a certain angle, the pile forms a silhouette of a motorcycle. The text on the slide explains that this is a lesson in perspective and how cognitive claims are often just tests for the observer's ability to see what is hidden.

However, if you shine light through it, you find out that actually this is a sculpture and it has a very particular pattern. This motorcycle. This is an amazing sculpture by Shigeo Fukuda, who made these sculptures that don't look like anything at first. They just look like a pile of objects. And then when you shine a light through them, you see what's really hidden underneath.

This tells us several things. It tells us that perspective is really important. Things that don't look like anything might actually be something much more interesting if you look at them in a certain way. In particular, the version of this that I'm going to talk about today is that I think all cognitive claims, meaning when you look at a system and you say this thing is intelligent or it's not intelligent or it has some degree of cognition or it's a simple mechanism or an algorithm or whatever it is, these claims are in large part IQ tests for ourselves because you have to be smart enough to really realize what you're looking at. If you don't see it, it may not be an issue with the system, it may be an issue with the observer. I think it's really important for us to raise our game as far as being able to notice some of these things.

**Sentient beings are innumerable,
I vow to liberate them all**

— from the Four Bodhisattva Vows

Perspective is key

Cognitive claims are IQ tests
for observers



Now, another example that I use here has to do with the electromagnetic spectrum. Think what it was like back in the day before we had a good theory of electromagnetism. We had many things. We knew about lightning and static electricity and light and magnets. First of all, we thought that these were all different things. Second, what we didn't know is that there's this whole enormous spectrum to which we are basically blind. We're completely oblivious to most of this, except a very narrow range here, which is the visible spectrum. That's because of our own evolutionary history and our own cognitive and sensory architecture. We didn't know any of this.

The benefits of having a good theory, like the modern theory of electromagnetism, are that it does two really important things. First of all, it unifies phenomena that otherwise seem like very different kinds of things. And it shows you that underneath all of it is one single deep symmetry. The other thing it enables you to do is to create technology that you can use to operate in other parts of the spectrum. You can improve life by participating in what's going on in these other parts of the spectrum for which our own evolutionary history did not prepare us. That's the benefit of having a deep theory.

The Power of Unification

- understand how diverse phenomena are a continuum
- create technology to observe and detect parts of that continuum of which we were previously oblivious
- utilize this knowledge for a myriad of applications that improve quality of life

One of the things that my group is very interested in is to develop a framework that allows us to recognize, create, and ethically relate to truly diverse intelligences.

Much like the electromagnetic spectrum, my claim is that there is a spectrum of intelligences that are very, very diverse. Some of them are easy for us to see; others are not. What we need to develop is a theory and the technology to allow us to relate to them.

What I would like to do is be able to relate not only to familiar creatures, such as primates, birds, and maybe an octopus or a whale, but even to very weird biological forms, such as colonial organisms and swarms, and also to synthetic biology, meaning engineered new life forms.

AIs, whether software or robotic; maybe someday exobiological agents; and some really unusual things, such as patterns in physical media and also patterns that are not embodied in the physical world at all. These are some of the things we work on.

Today, mostly for reasons of time, I'm going to talk about this, but I'll mention a few of these things.

Of course, I'm not the first person to try this. Here's Rosenblueth, Wiener, and Bigelow in 1943 wanting to develop a kind of cybernetic continuum all the way from passive matter up through different kinds of competencies, up through human level and beyond, including metacognition.

My version of this, at least version 1.0, is in this paper.

What's really important is that we need to develop theories that are not just philosophy, although philosophy is really important.

We need to develop a theory and technology that, first of all, moves experimental work forward.

In other words, the ideas I'm going to give you today about all these things are only as good as their ability to lead to new discoveries.

I'm not interested in redefining words as part of linguistics or exclusively as philosophy.

I want theories that help us make new discoveries: better biomedicine, better bioengineering capabilities, and, importantly, better ethics in the future. This is the project.

Slide 6 of 60 · Watch at [6:19](#)

My Framework Goal:

- Recognize, create, and relate to truly diverse intelligences regardless of composition or origin (unification)
- familiar creatures - us, apes, birds, octopus
- weird creatures (colonial organisms, swarms)
- synthetic biology - engineered new life forms
- AI (software or robotic)
- exo-biological agents (Earth is N=1)
- patterns in physical media
- patterns in Platonic Space
- moves experimental work forward - new biomedical and symmorpho capabilities, better ethics

Behavior, Purpose and Teleology
Arturo Rosenblyeth, Norbert Wiener and Julian Bigelow
in: *Philosophy of Science*, 10(1943), S. 18-24.

Frontiers in Systems Neuroscience
"Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds"
Michael Levin

Now, one thing that I think about a lot is this axis of persuadability, that is, different kinds of systems, and what are the tools that you can use to interact with those systems. I think that all cognitive claims are really just interaction protocol claims.

When you say that a certain system has some level of cognition, what you're saying is, here's a bag of tools that I plan to bring to communicate with that system. Then it becomes an empirical task to try that, and we can all see how well that worked out for you. You might use the tools of hardware rewiring. You might use the tools of cybernetics and control theory, or maybe behavioral science, or maybe psychoanalysis, love, and friendship. When you're dealing with a system, knowing where on the spectrum that system lies is not obvious. You can't guess it. You can't do that from a philosophical armchair. You have to do experiments. The experiments will tell you whether you've wasted a lot of time by trying to convince or psychoanalyze a mechanical clock, or you're treating animals as a simple machine and leaving a lot on the table as far as what are some of the things they're capable of. We are working on all sorts of ways to move various tools from cognitive and behavioral science, typically used up here, into embodiments and contexts where they're not normally used, so cells and tissues and some even stranger things.

Slide 7 of 60 · Watch at [7:46](#)

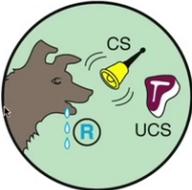
Axis of Persuadability:



Hardware modification only



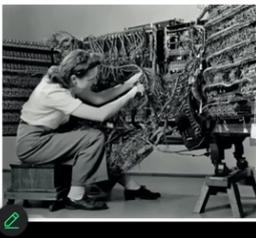
Modify the data encoding setpoint of goal-driven process



Training by rewards/punishments



Communicate cogent reasons



Intelligence is to be decided empirically, not via stale philosophical categories

Cognitive claims are interaction protocol claims

-> better technology
-> better ethics



What I'd like to do today is to try to climb up this scale of progressively more unusual agents, and for a final analogy to make it clear what we're doing, something like this happened to numbers already.

We started out with counting numbers, one sheep, two sheep, and so on. Then eventually somebody came up with zero and someone else came up with negative numbers, then eventually some fractions, and then eventually some irrationals and transcendentals and all these interesting things.

What happens every time you go up in the scale, something interesting happens. You have to take a conceptual leap that breaks prior categories. You have to break some assumptions you had about what is a number. People, when they were thinking about this, certainly did not realize that any of this stuff was a number. You have to break some boundaries. You have to be able to recognize them. You have to be able to do something useful with them.

Slide 8 of 60 · Watch at [8:44](#)

Expanding the Set of Cognitive Kin

https://in.pinterest.com/pin/627970741770398254/

frozen patterns

active patterns

evolved +
designed +
hybrid agents

all
biologicals

complex
life forms

brainless
animals

For each new category

- conceptual leap needed
- what does it break?
- recognize them
- do something useful with them
- active compassion toward them

And it's disturbing and disruptive. And we know this, this poor guy got drowned at sea for talking about irrational numbers, because these things upset very ancient categories. And when you upset ancient categories, there's a lot of resistance. And what I'm about to tell you makes things much more complicated, but more beautiful, and more meaningful.

Expanding our Mind-blindness is Disturbing, but Necessary



Upsets cherished categories

Makes things more complicated
(but more beautiful)

So let's begin. Many of us are educated in this tradition. So here's a famous piece of art called Adam Names the Animals in the Garden of Eden. It's an ancient biblical scene from the West that shows there are discrete natural kinds. Here are some animals. Then here's Adam. He is different from them. What I think they got right is that it's Adam's job to name the animals.

In these ancient traditions, naming something means that you've discovered its true inner nature. That part of this, I think, is right. It is on us to discover the true inner nature of many beings and give them names.

This is a very anthropocentric view. Certainly, most of us were brought up in this era of brain chauvinism, where we think intelligence and cognition arise with complex brains.

Anthropocentrism, or at best, brain chauvinism



So those kinds of things are pretty easy for us to detect. So here's this little squirrel. He's going to set up an accident scene here. You can see he knows precisely what he's doing. There he is. He's going to set this up. He's going to look to see if his owners are noticing this. If they're paying attention to this whole thing that he did, he wants some attention. So he's checking whether anyone's looking at this. And so this kind of stuff is easy for us to recognize because it happens at the same spatial-temporal scale as us. It happens in the same three-dimensional space. These beings have similar goals to ours. It's easy for us to recognize this kind of thing.

Easy to Detect Mind in Brainsy Mammals



Why so obvious? Same spatiotemporal scale, same space, similar goals

Although I will say that even the story with brains is not that simple, because there are amazing clinical cases, and we reviewed them in this paper, where humans have extremely small amounts of brain tissue. So not nearly enough, as modern neuroscience would tell you, is needed for normal performance. And yet they have normal or above-normal performance. And so what's going on here? Even the case with the brain is not really obvious. That's the first thing.

Even with brains, things are not simple

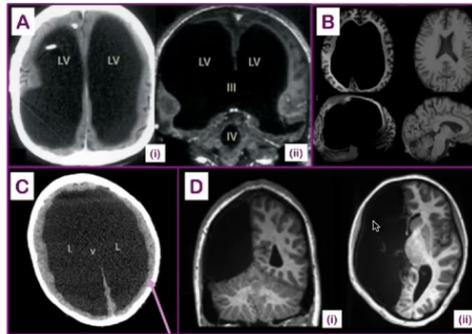


Figure 2. Select cases of reductions in brain matter with normal function. [A] Image from (Feuillet *et al.* 2007) showing a white collar worker case of extreme hydrocephalus; he led a normal life as a civil servant, who possessed an average IQ of 75. During his neurological assessment at age 44, his (i) CT scan and (ii) T1 weighted MRI scans with contrast showed extreme ventricular enlargement. LV indicates lateral ventricle, III and IV indicate the third and fourth ventricles, respectively. [B] Image from (Alders *et al.* 2018), showing the case of a 60-year-old with a bad mood with massive ventriculomegaly and severely reduced cerebral mantle and corpus callosum, that went largely unnoticed. The left column is T1 weighted MRI images taken in the transverse, coronal, and sagittal planes of the patient. The right column represents T1 weighted MRI scans of a healthy control. [C] Image from (Persad *et al.* 2021), imaging of a Canadian living a normal, independent life with massive hydrocephalus. MRI scan taken from the axial view (plane parallel to the ground) at the level of the lateral ventricles (arrow points to extremely thin layer of cortical mantle, LV stands for Lateral Ventricle). [D] Image from (Asaridou *et al.* 2020), showing the T1 Weighted MRI scans of a child born without left hemisphere (i) taken in the coronal plane, (ii) taken in the axial plane. The child had normal cognitive development and language skills despite hemispherectomy of the left hemisphere and near-absence of the corpus callosum. All images re-used with permission.

Minimal brain structure

or function (Savant syndrome)

cases of high performance

Med & Mater 2011, 13(4) doi: 10.1039/c1mt00131g

Cases of Unconventional Information Flow Across the Mind-Body Interface

Kristin Edman
Faculty of Dentistry, University of Toronto
Toronto, Canada
and
Michael Levin
Alpha Discovery Centre
Yale University, New Haven, USA
and
Winn Institute for Biologically Inspired Engineering
Harvard University, Cambridge, USA

But we're going to go significantly outside the brain. And so we're going to take a few conceptual leaps. We're going to talk about the fact that intelligence is way older than brains. And what it breaks is our assumptions of scale and our assumptions of substrate. And the way we're going to recognize these intelligences is by noticing their anatomical goals and the degree of ingenuity that they can deploy in what they're doing.

What we're going to study is a model system, which is cells as a collective intelligence operating in morphospace. The idea is to move away from brains and to use a stepping stone to the more unusual kinds of things to see what it's like to recognize intelligence in a novel substrate.

Biological Intelligence outside the Brain

- Conceptual leap needed: intelligence is older than brains
- What it breaks: scale and substrate
- How to recognize them: anatomical goals, ingenuity
- Useful implications: biomedicine

model system: cells as collective intelligence in morphospace

The first thing I will point out is that this is the kind of thing we're all made of. This is a single cell, and you will notice that there is no brain, no nervous system. This thing is incredibly competent in the job that it has to do. All of its local cognitive light cone is very competent in what it needs to do. Not only are we made of an agential material, materials with an agenda, but even below that, the cell itself is made of molecular networks.

We are All Collective Intelligences Made from an Agential Material



Lacrymaria = 1 cell
no brain
no nervous system

high competency
at cell-level
agendas

00:37 -00:37

The material that these cells are made of, the molecular networks, are themselves capable of six different kinds of learning. You don't need a cell for this. You don't need the nucleus or anything else. Just the small pathway of chemicals turning each other on and off can be sufficient for habituation, sensitization, and associative conditioning. We study this in these papers. In our lab, we're using all of that, for example, for the biomedical purpose of drug conditioning to try to form and erase specific memories in these molecular pathways that are important for health and disease.

Collective Intelligence Below the Cell Level

Learning increases causal emergence!
Federico Pigazzi

Biomedicine:
- drug conditioning

International Journal of Molecular Sciences

MDPI

Article
Learning in Transcriptional Network Models: Computational Discovery of Pathway-Level Memory and Effective Interventions
Suzana Biawas ^{1,2,*}, Wesley Clawson ^{3,4} and Michael Levin ^{1,3,4,5}

iScience CellPress
OPEN ACCESS

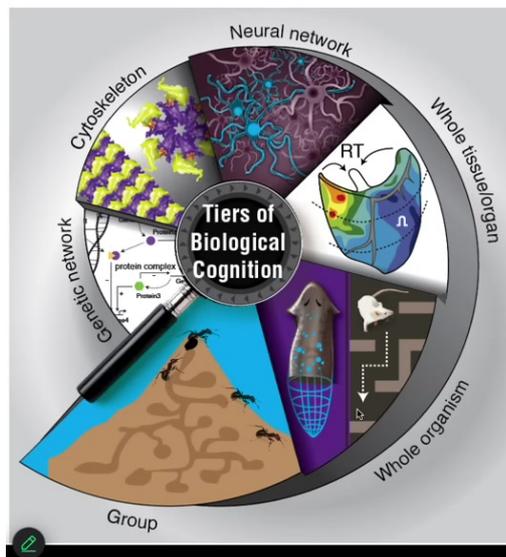
Article
Gene regulatory networks exhibit several kinds of memory: quantification of memory in biological random transcriptional networks

Patrick Erickson

We're starting to see that there are different kinds of properties that are important for cognition all throughout the architecture that we're made of. At the level of molecular networks and subcellular components and cells and tissues and organs, and even up through groups. Every level of this multi-scale competency architecture is solving problems in different spaces. When I say intelligence, I'm using William James's definition, the ability to reach the same goal by different means. We'll unpack this in a minute. We consist of this agential material.

Life Has Nested Intelligence, not Merely Structure

James: intelligence = “same goal by different means”



Multi-scale Competency Architecture

each level of organization solves problems in its own space (morphospace, transcriptional space, physiological space, 3D behavioral space, etc.) using some of the same tricks, at various levels of sophistication

Now, one of the things that's important about noticing intelligent behavior is that we as humans are obsessed with three-dimensional space. That is, we basically know how to recognize intelligence in medium-sized objects moving at medium speeds through three-dimensional space, and birds and primates. It's easy.

However, biology, for example, has been solving problems in other spaces long before nerve and muscle enabled us to move through three-dimensional space. So long before that, our cells were solving problems in a high-dimensional gene expression space, in physiological state space, and the thing that we're going to be talking about today, which is anatomical morphospace. All these spaces are very hard for us to visualize. They are not directly accessible to our senses the way three-dimensional space is, but that doesn't make them any less real. Everything we see in the three-dimensional world is also constructed as a model by our cognitive system. And we're just not very good at noticing these high-dimensional spaces.

But what it does mean is that when you see something that isn't embodied in the conventional sense, meaning it doesn't have a robotic body that allows it to move around in the 3D world, that doesn't mean it's not embodied. So brain organoids, AIs — people say, if it doesn't engage with the real world, it's not really cognitive. You may not know the space with which it's engaging, because what's important is this loop, this perception, decision, action loop, and that loop can happen in many spaces, not just the three-dimensional space.

Life Has Embodiment Outside of Familiar 3D space:

3D Space (behavior)

perception-action loop can happen in other spaces!

-> unconventional embodiment for AI's

Transcriptional Space

Huang, S.; Erberg, I.; Kauffman, S. Semin Cell Dev Biol 2009, 20, (7), 869-76.

Morphospace

Cervera, J., Levin, M., and Malle, S. (2021). BioSystems, 209:104511

Physiological Space

Marder, E., & Goaillard, J. M. (2006). Variability, compensation and homeostasis in neuron and network function. Nat Rev Neurosci, 7(7), 563-574.

So what we're going to do is to look at one specific kind of intelligence, which is that of swarms of cells navigating anatomical space.

William James's definition: "intelligence is the ability to reach the same goal by different means." This is something we've been working on for years: the hypothesis that morphogenesis is a collective intelligence exerting behavioral competencies in anatomical space.

In other words, this is a parallel and symmetry between behavior and anatomy: the construction of the body as behavior in anatomical space.

We're going to pivot neuroscience. It's exactly what brains do in three-dimensional space, by integrating a bunch of cells using electrical signaling among cells to navigate three-dimensional space. We're going to show what this evolved from, which was cellular swarms navigating anatomical space.

Everything is the same, except that instead of milliseconds, as in the nervous system, we're talking about the scale of minutes or even hours. Instead of three-dimensional motion, we're talking about movement through anatomical space, the space of possible anatomical configurations.

What is the same are the mechanisms — literally homologous mechanisms: the exact same molecular machinery, ion channels, gap junctions, microtubules, and all of these kinds of things that are used by the brain to do conventional cognition were used by groups of cells to figure out what shape you should be during embryonic development.

These are ancient mechanisms; they got pivoted into brains, but they've been doing this job for a long time.

We have a tool called Field Shift that you can play with, where you can paste in an abstract from a neuroscience paper and it changes words: instead of neuron it will say cell, instead of milliseconds it will say minutes, and it immediately converts it to a developmental biology paper. It's very interesting how much symmetry there is.

Slide 18 of 60 · Watch at [18:11](#)

“Intelligence is the ability to reach the same goal by different means.”

- William James

Hypothesis: morphogenesis is a collective intelligence, exerting its behavioral competencies in anatomical morphospace

Cellular Swarms vs. Brains (or, what is neuroscience about?)

- Different time scale (minutes, not milliseconds)
- Different problem space (morphospace, not 3D motion)
- but, same (homologous!) mechanisms - ion channels, GJs, NTs
- and, many of the same algorithms (tools of cog sci work great)

levinlab.dev / FieldSHIFT

Translate from one domain to another

Neuroscience - Detected

Developmental biology

Let's talk quickly about what kind of intelligence these systems can deploy in anatomical space. And the reason we're going into this is that it's a good example. It's one of the few examples we have that's a stepping stone between the kind of intelligence that we already understand and we know how to deal with. We understand some of the mechanisms that are involved. And it helps us to see what happens when that gets pivoted, even a little bit. It becomes very hard. Most people do not, and certainly most of the textbooks do not think of morphogenesis as a cognitive task, but in fact it is.

Here's a simple example of what William James was talking about. If you have this axolotl, this is an amphibian that has some interesting properties. One of the

properties is that if it loses a limb anywhere along this axis—you make a cut anywhere along this axis, the cells will immediately recognize that something is wrong. They start to divide and undergo morphogenesis, they build the limb and then they stop. That's the most amazing part: they know when to stop. When do they stop? They stop when a correct salamander limb has been completed.

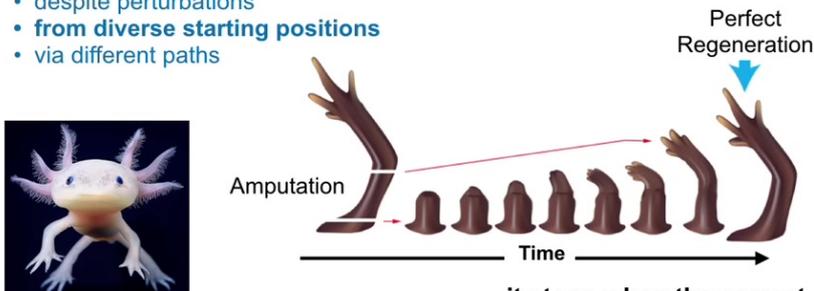
It's not simply the case that molecular pathways did what they did, and then there was emergence, and some complex thing ended up happening. That's true, but that's not the end of the story. It has an important amount of feedback and anatomical homeostasis, because what it can do is reduce error. This is an error minimization scheme. The system has the autonomy to pursue this goal, and when you try to deviate it from that goal, it will keep working towards it, and it will stop when it's done. That's the definition of goal-directedness.

The same thing is true of embryogenesis. Even in mammals, if you cut embryos in half, you don't get half bodies. You get two perfectly normal monozygotic twins and triplets. In effect, this is also a case of regeneration. Your whole body regenerated an entire body from one cell. It's an amazing process. Both embryonic development and regeneration are instances of anatomical homeostasis.

Slide 19 of 60 · Watch at [20:22](#)

Same anatomy, from different starting states

- get to the same outcome
 - despite perturbations
 - **from diverse starting positions**
 - via different paths

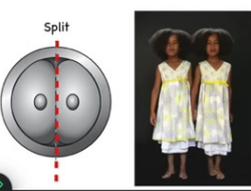


Amputation

Time

Perfect Regeneration

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



Split

Splitting an embryo in half makes 2 normal embryos (regeneration from 1 cell!!!)

But I want to point out that this is really important. This is not about damage and repair of damage.

Here's an experiment. You take a tail and graft it to the flank of an axolotl. What happens over time is that it slowly remodels into a limb. This is remarkable because the cells here at the tip of the tail, these little cells, there's nothing wrong locally. They're tail tip cells sitting at the end of the tail. There's nothing wrong with them, but they turn into fingers. Why do they turn into fingers? Because this system, much like any good cognitive system, is able to bend the action of its parts towards a larger scale goal. No individual cell knows anything about what an entire axolotl is supposed to look like.

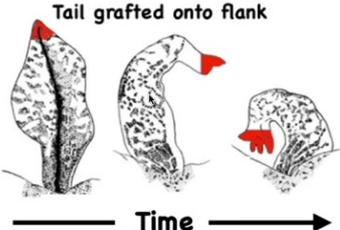
But there are systems and subsystems here, each of which has different scales of set points, of anatomical set points. When there's an error, they will have to bend the action space of their subsystem to align all the cells in the molecular pathways and everything that needs to happen to turn tail tip cells into fingers. This is what will be downward causation, where the effort of reducing error is going to control all of the downstream steps. The local order here obeys the global plan. The large-scale system can bend the action of its parts to do things in accordance with a larger scale plan that the parts don't know anything about, and it actually has a lot of ingenuity in doing this.

Top-Down Control: Holistic Order

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



Tail grafted onto flank



Farinello-Ferrazza, Esperientia, 1956 (15)

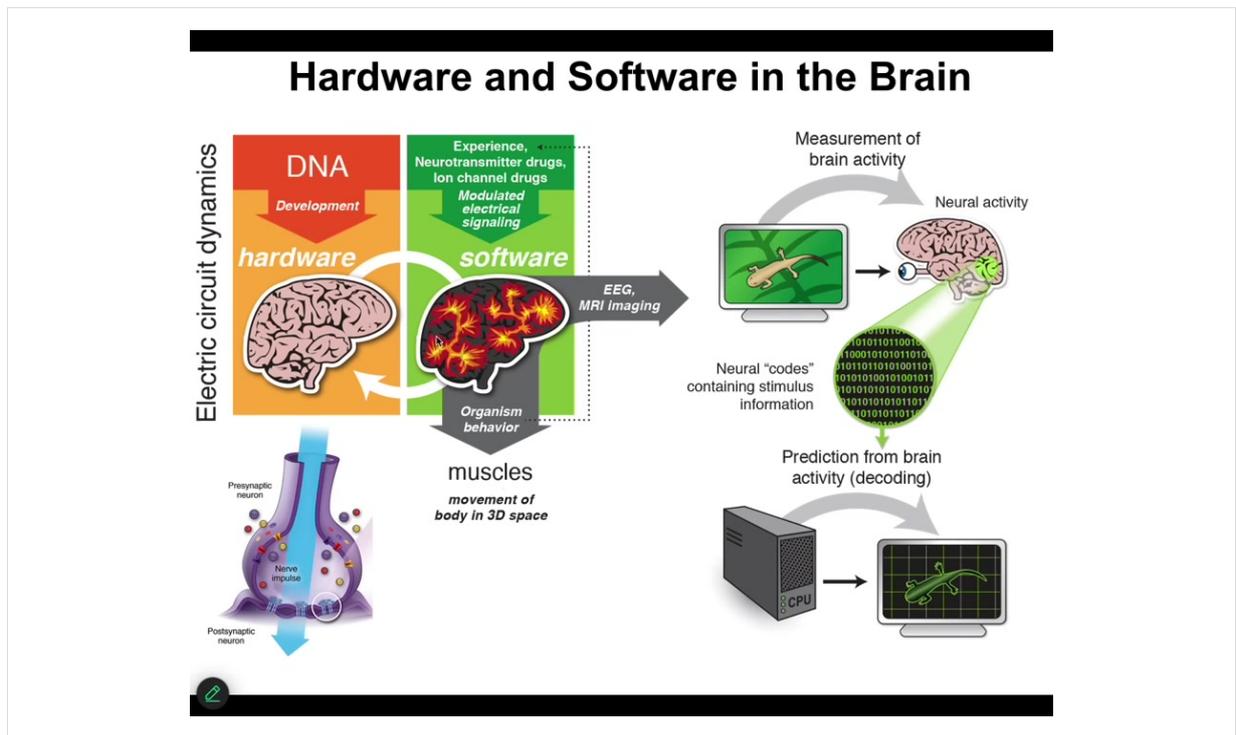
Anatomical homeostasis: local order obeys global plan

This is one of my favorite examples. This is a cross-section through a kidney tubule in the newt. Normally you see about 8 to 10 cells working together to form this little structure. One thing you can do experimentally is make the cells much bigger. The way you do that is you provide extra copies of the genetic material. This is a polyploid newt where they have multiple copies of the genome and the cells end up being much bigger, but the newt stays the same size.

How does that happen? That's because fewer cells are working together to give rise to the same structure. The size of the cells will be matched by the number of cells that participate. When you make the cells truly gigantic, just one cell will wrap around itself and give you the same final outcome.

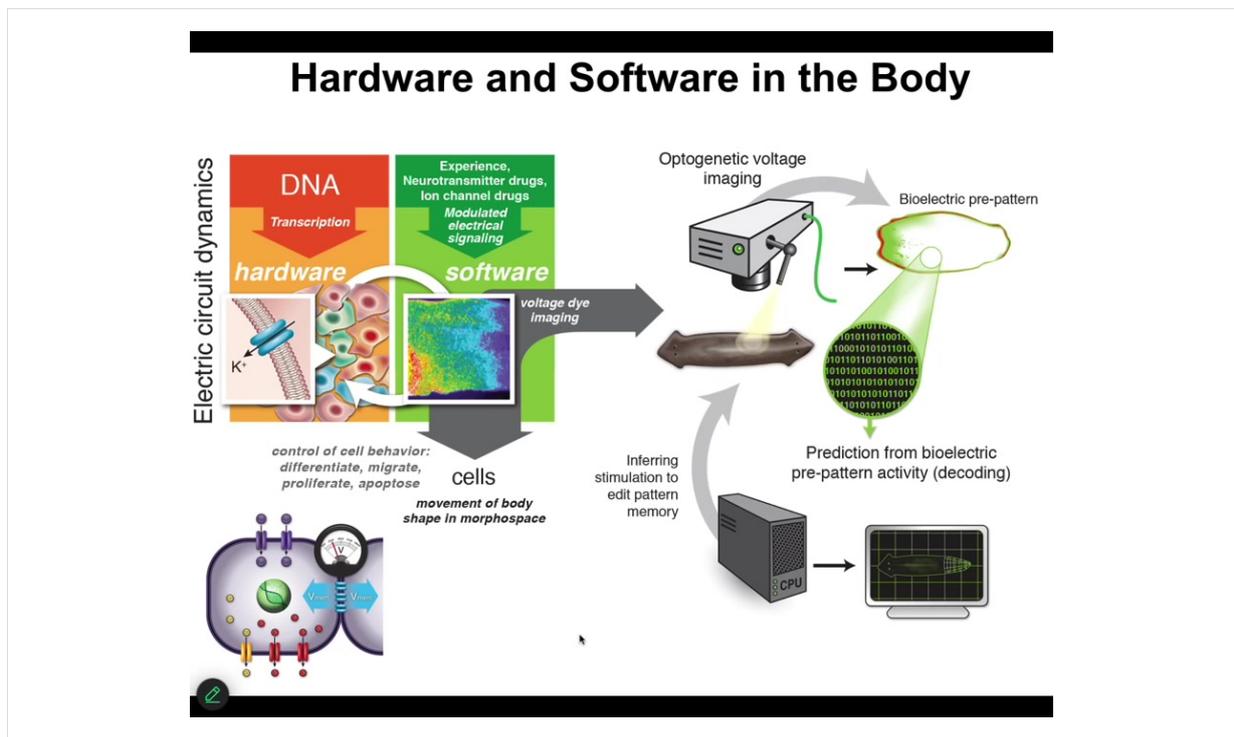
There are a couple of things going on here. First, this is a downward causation: in the service of having this large-scale outcome, different molecular mechanisms are being called up in order to execute the plan. On any IQ test, they will often give you a set of parts and they'll say, solve this problem with these objects, and you have to be creative in how you do it.

But look at what happens if you're an embryo coming into this world. You can't, never mind the environment, you can't even count on your own parts. You don't know what size your cells are going to be. You don't know how many copies of your genetic material you're going to have. This is not a hardwired process. It's problem solving. Every single time you have to use the genetic affordances that you have, the different molecular components, to solve the problem in different ways.



This is not about brains. Brains pivoted this and adapted this from what was going on in morphogenesis. Every cell in your body has these ion channels. Most of the cells have these electrical synapses with their neighbors. And they do the same thing, except instead of moving you through three-dimensional space, they move the configuration of your body through anatomical space. But everything else is the same.

And so the reason we're going through all this is to show that even when you're using the exact same machinery to do the same algorithm to execute on goal states that are in memory, things look quite different. And we have different sciences devoted to this. You have developmental biology versus cognitive and neuroscience.



Typically, they're not seen as that much overlapping. We try to explore this symmetry. We develop tools to read the electrophysiology of cells during morphogenesis, the way that neuroscientists try to read the brain.

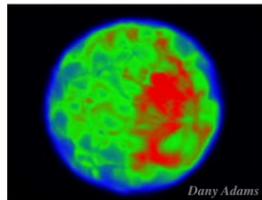
These are voltage-sensitive fluorescent dyes that we used to non-invasively watch the cells talking to each other electrically. What you can see here is not data. These are not the models. These are actual real living organisms.

This is an early frog embryo. You can see all the different cells and all the conversations they're having with each other. These are explanted cells in culture deciding whether they should be a part of this mass that's going to build something or go on their own. We can watch these things happen non-invasively. We can literally see now the cognitive glue that binds together individual subunits into a collective that does something more than the parts.

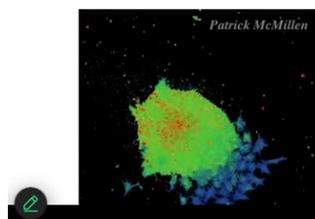
We do lots of simulations. We do the molecular biology simulations and the electrophysiology. Even at a higher level, things that are studied in machine learning, so pattern completion. That regeneration I showed you is very similar to pattern completion and trying to minimize free energy that neuroscientists study.

How we detect and model bioelectric patterns:

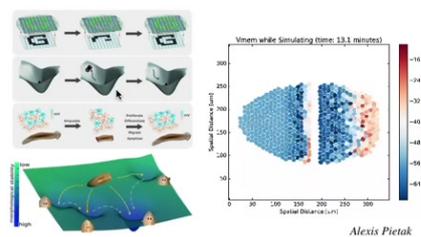
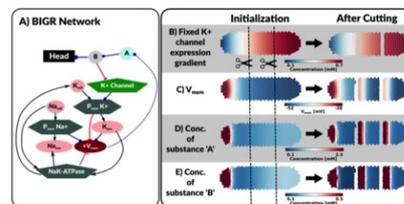
Characterization of endogenous voltage gradients - direct measurement and correlation with morphogenetic events



Voltage reporting fluorescent dye in time-lapse during *Xenopus* development



Quantitative computer simulation: synthesize biophysical and genetic data into predictive, quantitative, often non-linear models



Importantly, beyond just being able to read these patterns, you also want to be able to rewrite them. You want to be able to change the memories and the information that exists in these patterns. And so we've developed tools to do this. We don't use applied electric fields or magnets or electromagnetic radiation. We don't do any of that. What we do is hack the normal interface that is generally used by cells to control each other. That means, just like in neuroscience, we have ways to open and close ion channels, and these might be drugs, and these might be optogenetics and different ways to control both the electrical state of the cells and the connectivity. Who talks to whom? That's the communication interface that we have. What happens here is that these groups of cells and individual cells pass each other messages, and these messages are often very convincing.

Manipulating Bioelectric Networks' Content

Non-neural cell group

hyperpolarized ← → depolarized

Neurotransmitter (moving via V_{mem})

- Transporter or receptor mutant overexpression
- Drug agonists or antagonists of receptors or transporters
- Photo-uncaging of neurotransmitter

Tools we developed
(no applied fields!)

- Dominant negative Connexin protein
- GJC drug blocker
- Cx mutant with altered gating or permeability

Synaptic plasticity

- Dominant ion channel over-expression (depolarizing or hyperpolarizing, light-gated, drug-gated)
- Drug blocker of native channel
- Drug opener of native channel

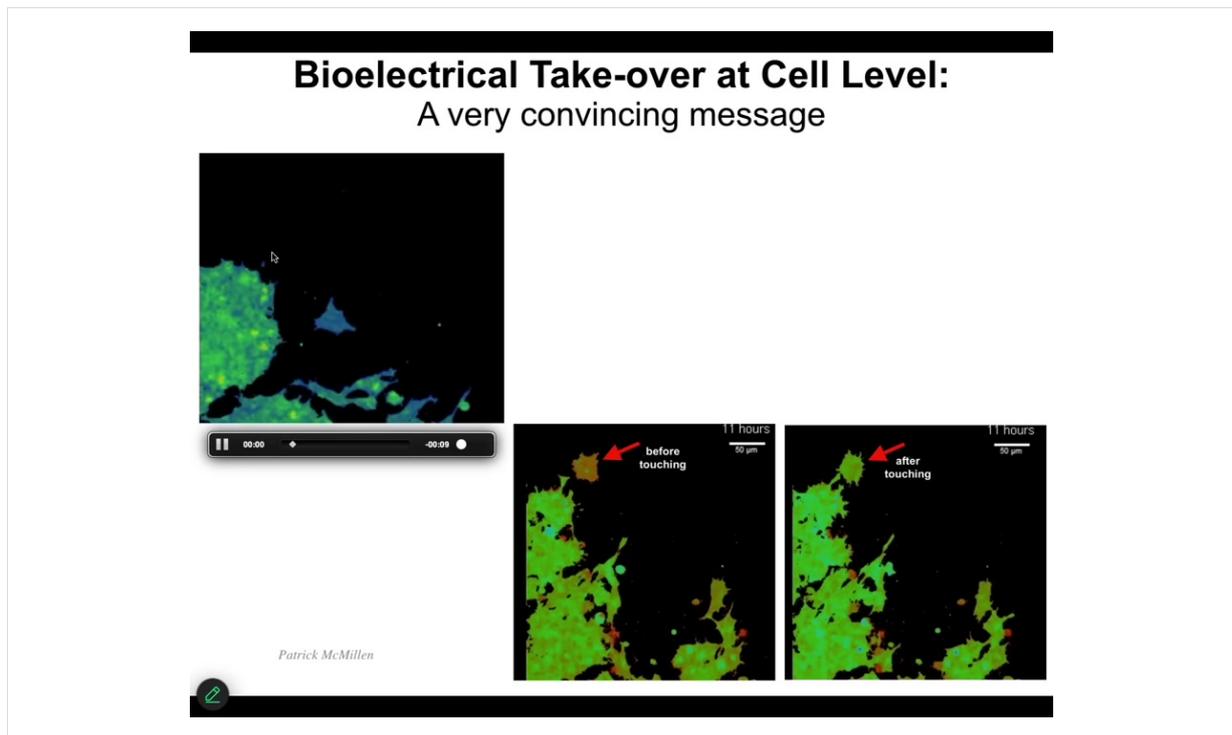
Intrinsic plasticity

The communication interface we hack

I want to show you an example of what a convincing message looks like. The color here indicates a voltage state. You can see these cells — these are two frames of the same video. I'm going to play the movie for you in a minute.

What you see here is that this cell is moving along. It has a different voltage than the rest of these. Then it touches, just barely. As soon as it touches, from here to here, boom, it changes its voltage. Watch what happens. It's crawling along. It has its own voltage state. It's going to reach out, bang. That little touch is all it took to convert that cell to the same state as these others. Now it's joined the collective and it's going to do collective things.

What we're talking about here is the mechanism of one kind of cognitive glue. I'm sure there are many. But in this model system, we get to see the same as in our brain. It's the mechanism that binds cells together towards larger scale goals, is able to propagate information and radically change the future, the state and then the future behavior of that subunit. Now it will coordinate it towards greater goals.



Now, what kind of messages and what kind of goals can you pass? One of the things we do, and this is part of our work on regenerative medicine, is we try to convince cells to build specific organs. Here, what we've done is this is an early frog embryo. What we do is inject into one of the cells some ion channel RNA that's designed to change the voltage of those cells. It's a spot. It's not a single cell. It's going to be a group of cells that have a particular voltage. One of the things we found is a message that says, build an eye here. It's a particular bioelectrical state that is a memory of a place where an eye should be formed. When we do this, it makes an eye.

This is a tadpole. You're seeing it from the side. Here's the gut. The tail is back here. There's a brain out here. The mouth is here. This is the gut. You can see this eye here because we took these cells and we convinced them that they should be an eye. If you section that eye, you see all the stuff that belongs in the lens, optic nerve, retina.

In fact, if you only get a few cells, the blue cells are the ones that we injected, they will convince their neighbors to help them build an eye. Like another collective intelligence, ants and termites, if they come across some kind of a food object or something that's too big for them to carry, they will recruit their neighbors to work.

The reason this all works and the reason that we can create a whole eye without having to specify all the different details that it takes to make an eye, all the different gene expression, everything else, eyes are very complex. We don't need to do any of that. We specify only the top level goal, build an eye, and the material, this is one of the competencies of the material, is to align all of the underlying molecular mechanisms

towards that goal. We don't need to worry about them. It's the right size, the right shape, we don't have to worry about it.

Slide 27 of 60 · Watch at [31:49](#)

Communicating with Somatic Intelligence: prompt = make an eye

ion channel mRNA targeted to ventral or posterior regions
Sherry Aw, Vaibhav Pai
can reprogram many regions, even outside "competency zone", into complete ectopic eye!

Native eye Induced Eye

Muller glia	i	vi
Amacrine cells	ii	vii
Rods	iii	viii
Cones	iv	ix
Phase contrast	v	x

voltage-modified cells

wild-type cells recruited to ectopic lens!

1. BIOE is instructive
2. modularity - not cell level, organ-level subroutine call
3. higher-level prompt reveals higher tissue competency than Pax6 prompt
4. self-scaling of system to task

Developmental Modules because morphogenetic goals

Getty Images

Now, importantly, this doesn't always happen. Sometimes when we make, for example, three spots here, this animal is only going to have one eye because this one ended up being quite convincing and the cells took it and built an eye; these did not.

What the cells did was to normalize them, and the neighboring cells normalized out their voltage and they won.

It's a battle of worldviews. There's a bunch of cells that are part of a cancer suppression mechanism. A bunch of cells are saying, "We don't need an eye here. You have a weird voltage. You should be like us." These cells are saying, "No, you really should be an eye. This is the correct pattern." They fight it out, and depending on which pattern takes over is what you get in the end.

A Battle of World-Views: plasticity in morphogenetic navigation

The diagram illustrates the effect of ion channel mRNA targeting on eye formation in a zebrafish embryo. It is divided into two main parts. On the left, a control embryo is shown with a single eye. Above it, a schematic of a zebrafish embryo is labeled 'ion channel mRNA targeted to ventral or posterior regions'. Below this, a micrograph shows the eye region with three red arrows pointing to '3 ectopic eyes, according to early eye marker Rx1'. On the right, another schematic is labeled 'ion channel mRNA targeted to ventral or posterior regions'. Below it, a micrograph shows the eye region with a single eye labeled 'EYE' and a 'gut' region. The text 'Vaibhav Pai, Sherry Aw' is visible in the bottom right corner of the slide.

And so this is part of our way to take these kinds of ideas and move them forward to regenerative medicine. Because right now, all of the kinds of approaches that are used in the field, surgery, stem cells, genomic editing, all these things, they're bottom up. They're really about the hardware. But there are other approaches that are using the tools of behavioral science to exploit the cognition, the intelligence of the material to do things that we otherwise cannot micromanage. And I think because of this, the future of regenerative medicine is going to look a lot more like a kind of somatic psychiatry than it's going to look like chemistry. Bioelectricity is the interface layer that enables these kinds of top-down controls. It's really mind-body medicine in a deep and fundamental way.

Future Medicine: communication, training (molecular pathways, cells, tissue)

from chemistry to somatic psychiatry: bioelectricity is the interface layer that enables top-down controls (i.e., mind-body medicine). We can hack this for transformative health applications!

ScienceDirect
The body electric 2.0: recent advances in developmental bioelectricity for regenerative and synthetic bioengineering
Justin Matthews and Michael Levin

Integrative Biology
PERSPECTIVE
Re-membering the body: applications of computational neuroscience to the top-down control of regeneration of limbs and other complex organs?
© Springer Nature 2017

Biotechnology

Patterns
Perspective
Cellular signaling pathways as plastic, proto-cognitive systems: Implications for biomedicine
Justin Matthews, Aya Jansky Cheng, Sam Stern, and Michael Levin
© Springer Nature 2017

CellPress

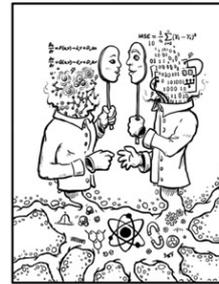


Now what we're going to do is talk about the next conceptual leap and even non-biological agents. Because what I've shown you so far is how to try to shift our brain-centered tools to recognize and communicate with intelligences that are like what we're used to. They're using the same materials that we evolved. In fact, we are made of them.

But there are still major differences. It's still really hard for people to take those ideas and to use them in developmental biology and molecular medicine. But we're going to go even further. What we're going to try to break is the standard divide that people seem to like, which is this idea of life versus machines. We don't have a good definition of either of those things, but a lot of people feel like these are two disparate, discrete categories, and we're going to talk about that.

Non-biological, hybrid, and minimal agents

- Conceptual leap needed: beyond “life vs. machines”
- What it breaks: categories of “life vs. machines”
- How to recognize them: behavioral tests
- Useful implications: engineering, CS



The first thing is that when we talk about things like AI and synthetic intelligence and so on, we're not just talking about language models sitting in a server box somewhere. We're talking about a continuum of being where humans already are being modified. They're modifying themselves in both biological ways and technological ways. This is only going to increase to an enormous diversity of embodiments for humans.

This is the kind of thing, when you try to make a distinction between a living organism as a machine, this is what you need to deal with. It's pretty obvious in certain cases, but in something like this, it's not at all. You don't want an ancient phrenologist trying to understand whether it's 51% a machine when you've got somebody who's partly evolved material and partly engineered material.

The most interesting and important thing about all these novel beings with whom we're going to be sharing our world is the question of what their goals are going to be. All the systems that we've looked at up until now have specific anatomical structures, physiological patterns, gene expression patterns that have been, at least in theory, set by evolution. So we have some idea of where their goals came from. We have some experience in recognizing them. But as we make these and change ourselves into novel beings, we really have to understand where these kinds of things come from. If not from evolution, where else do they come from?

For example, in our work here there are flatworms, these planaria with different shaped heads, and one of the things we can do is take a single species of planaria and bioelectrically prompt them to make heads belonging to other species. So there are

these attractors in anatomical space with different shaped heads—flat, round, triangular, whatever. And we can make them because it's not a hardwired kind of thing; rather, the cells make electrically mediated decisions about what shape they're going to be, so we can ask them to go into one of these other attractors, easy enough. These attractors were set by evolution, presumably. But what about novel beings who have not had a history of selection?

Slide 31 of 60 · Watch at [37:10](#)

Novel Beings, Novel Minds

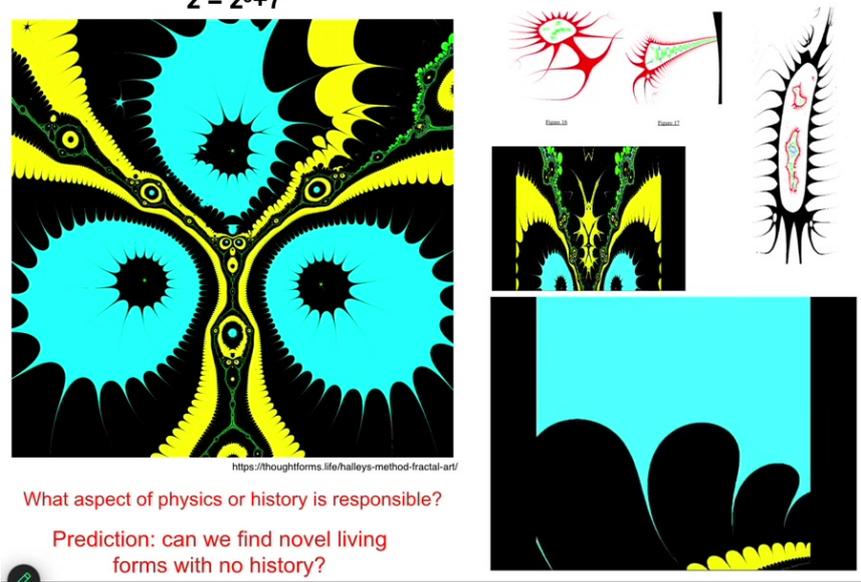
What will their behavior be, in anatomical as well as 3D spaces? If we can't rely on genetics+history...

What else is there?!

We need to think of where patterns come from. Biologists love two kinds of patterns. They love patterns from physics. In other words, something is round because it has to minimize surface area, so those kinds of things.

Patterns Come From Genetics, Environment, and ??

$z = z^3 + 7$



<https://thoughtforms.life/haileys-method-fractal-art/>

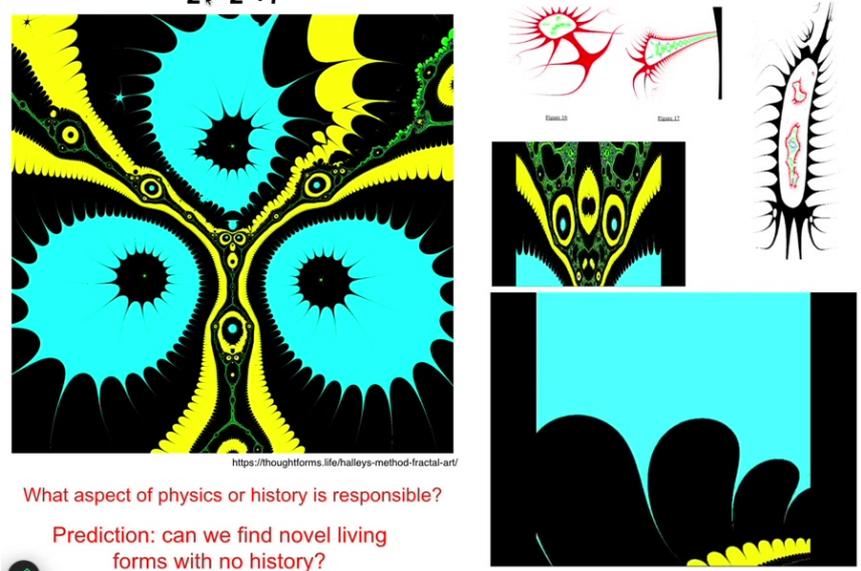
What aspect of physics or history is responsible?

Prediction: can we find novel living forms with no history?

They love history, meaning genetics and environment. In other words, you have a genome that sets specific patterns because that genome has been selected in a certain way and everything else died out in there. But we know that there's another source of patterns that doesn't come from either of those things, neither physics nor history. That's patterns for mathematics. What you're seeing here, this is a very simple function in complex number Z .

Patterns Come From Genetics, Environment, and ??

$z = z^3 + 7$



<https://thoughtforms.life/haileys-method-fractal-art/>

What aspect of physics or history is responsible?

Prediction: can we find novel living forms with no history?

If you plot it, it's called a Halle map of this function, you see this amazing kind of rich pattern. You can actually change these slightly in each frame and make video, so you can see these amazing mathematical transformations. All of this comes from the rules about the way that complex numbers behave. There is no fact of physics that explains why this thing is the way it is. There is no fact of history. This was not selected. It was not evolved. There's nothing about the physical universe that tells you what this is going to be. These are patterns that come from mathematics. And it doesn't hurt that they're very organic. Some of these patterns are incredibly organic looking. So now we understand that there are physics inputs, there are historical selection inputs, but also mathematics.

So now the question: can we find or make novel life forms with no history, at least in their macroscopic configuration, such that we can practice this, getting better at understanding what their behaviors and their forms are going to be like?

Patterns Come From Genetics, Environment, and ??

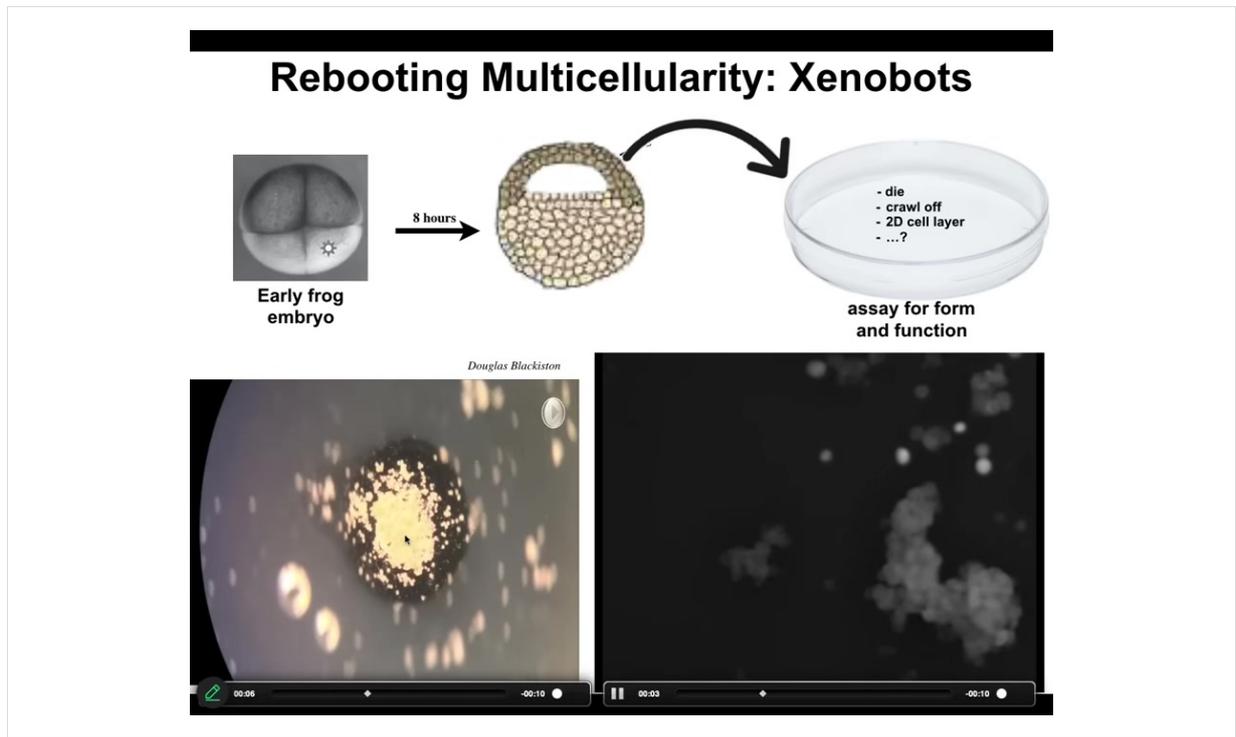
$z = z^3 + 7$

<https://thoughtforms.life/haileys-method-fractal-art/>

What aspect of physics or history is responsible?

Prediction: can we find novel living forms with no history?

I want to introduce you to two of them. The first one... we call xenobots. They come from cells that we scrape off the top of an early frog embryo, so these epithelial cells. You can see here that when you dissociate these cells, they don't die. They don't crawl away from each other. They don't form a two-dimensional flat cell culture layer. What happens is that when you put them by themselves, each one of these little things is a cell, they basically coalesce together. Here you can see this is a small group of them moving along. It looks like a little horse, but they don't all look like that. They make all kinds of different shapes. What they do is they interact here. That little flash you're going to see right here is calcium signaling.



And what they do is they actually come together and they make something like this, which is called a xenobot. It uses cilia, these little hairs that row on the outside of cells. The cells row against the water, so they can move. They can go in circles. They can patrol back and forth. They have group behaviors that you can see. They can interact in groups or alone. You can make them into weird shapes if you want. You can punch a hole. There's our swimming donut shape.

Xenobot behaviors - repurposing cilia for motion

The image displays several visualizations related to Xenobot behaviors. At the top left, a circular inset shows a close-up of a Xenobot, a small, white, multi-cellular organism. To its right is a scatter plot titled "collective behaviors" with axes labeled 'x' and 'y'. The plot shows several colored dots (yellow, green, blue, purple, red) representing different behaviors, with a pink line indicating a trajectory. Below the circular inset is a rectangular inset showing a Xenobot on a light-colored surface. To the right of the scatter plot is a vertical inset showing a Xenobot in a dark environment, possibly a maze. At the bottom of the image is a video player interface with a play button, a progress bar, and the name "Douglas Blackiston".

They have all kinds of interesting behaviors. Here's one traversing a maze. It swims along here. It takes the corner without bumping into the opposite wall. Here, for some reason, we don't know why, it decides to turn around, go back where it came from.

Xenobot in a maze (still water, no flow):

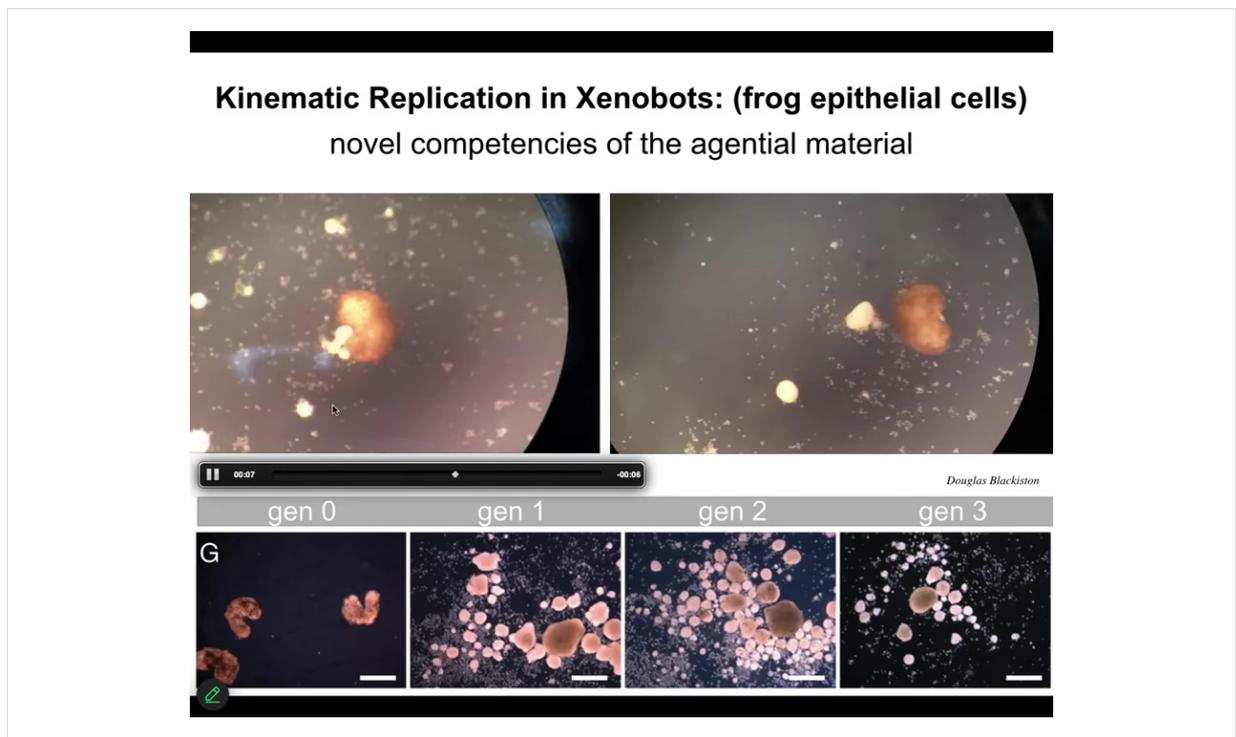


- 1) it traverses maze,
- 2) rounds the corners without bumping into walls, and
- 3) it makes a spontaneous decision to turn around without hitting anything.

Douglas Blackiston

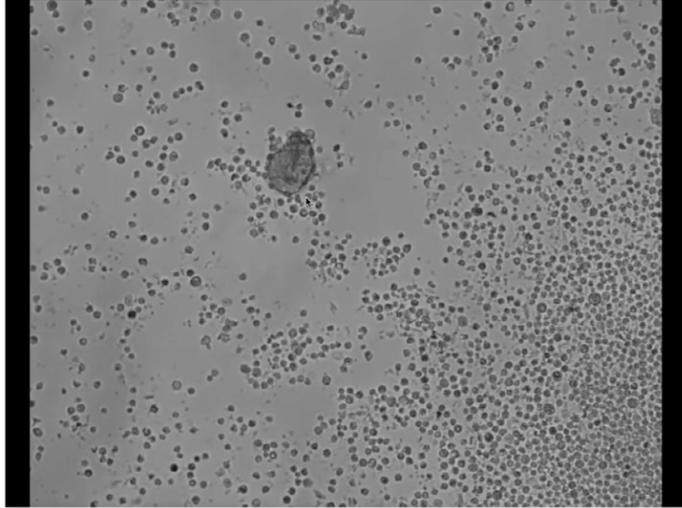
This is an amazing thing they do. This is called kinematic self-replication. If you provide some loose materials, these little bits are cells, loose epithelial cells, the Xenobots, both individually and in a collective, will run around and pile these cells into little piles, and then they polish little piles like this. Because they're dealing with an agential material themselves, these little piles mature to be the next generation of Xenobots. Guess what they do? They run around and make more, and that makes it the next generation.

We don't think there's any other creatures on Earth, as far as we know, that replicate in this way. Nobody else does kinematic replication. There is certainly nothing in the history of a frog that would suggest that this would happen. We're going to discuss in a minute what this means.



First I want to show you one other creature. We call these anthrobots. If you watch this little guy swimming around, if I didn't tell you what this was, you might think that this is a primitive organism we got on the bottom of a pond somewhere. I would tell you that if you sequence it, you would find 100% *Homo sapiens*. These are made from adult human epithelial cells, not embryonic and not amphibian. These are adult tracheal epithelial cells. They come from human patients who give lung and trachea biopsies. They can reboot their multicellularity and become this swimming little organism that doesn't look anything like any stage of human development.

Meet the Anthrobots:



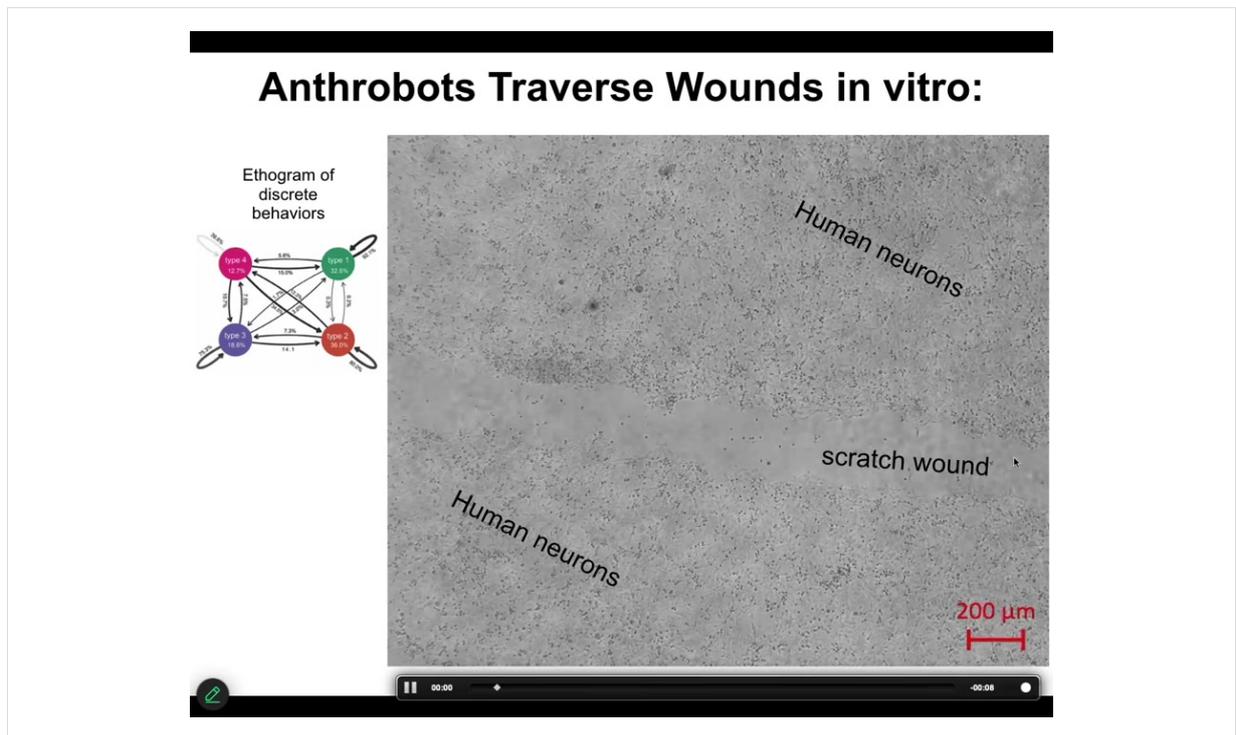
Where do the properties of novel systems come from if not eons of selection or explicit engineering?

Could you guess the genome from these data?

Could you guess behavior and form from the genome?

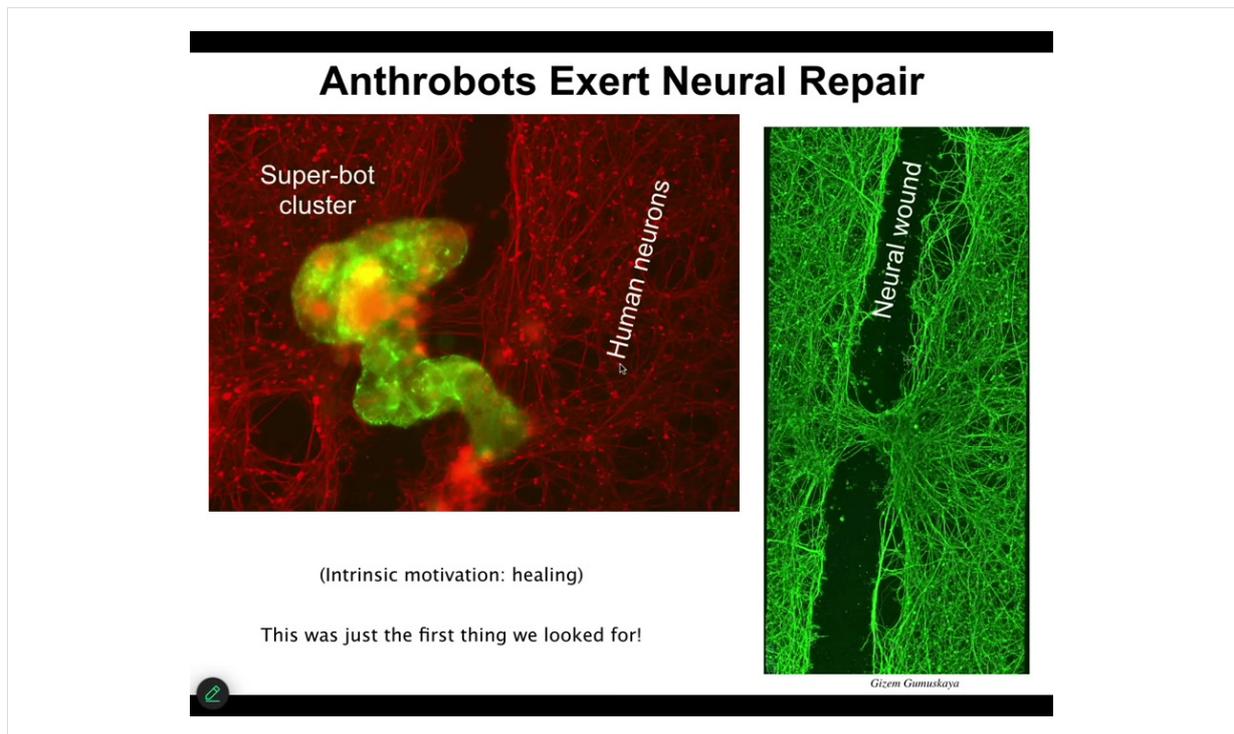
They have a couple of interesting properties. They have some of these collective behaviors. This is a plot of genes that they express differently than the cells they come from express. They have a completely different transcriptome, almost half the genome, 9,000 genes are differently expressed. We didn't change the genetics, much like with the xenobots. These have absolutely standard human genetics. They are in the same environment, except they're not touching other cells, but otherwise they live in the same kind of medium that they would be living in within a body. There's no drugs, no synthetic biology circuits, no nanomaterials. They're just—this is what you're seeing: the amazing plasticity of the material. They are able to make this other kind of creature. That creature has self-healing capacity. If you puncture it, it will heal the wound. They're younger. If you look at their age, they're younger than the cells they came from. This is an age reversal thing. That also has an explanation based around thinking that they're starting over as a new organism.

And then they have this amazing ability. Here's what you're seeing here is an anthrobot swimming down a scratch that was made in a bunch of human neurons. So here are human neurons plated at the bottom of the dish, and we put a big scratch wound through it.



The anthrobots drive down the wound, and then they settle down. They settle down into a cluster. Here's about a dozen of them. When they settle into this cluster, they cause the wound to repair. They're knitting the neurons across the gap. Here's what it looks like if you lift them up from this region. This is what you see underneath. They were knitting it across the gap.

This was just the first thing we looked for. Who knows what else they can do? They can do this. We didn't make them do this. It's some kind of intrinsic motivation that induces the healing. Who would have thought that your tracheal epithelial cells could become a self-motile little creature that has the ability to go around and heal other kinds of damage? Imagine that in the body. That's never happened before in the evolutionary path. It's not something that your tracheal cells do.



So the big question here is this. We know that the genome can give rise to these sorts of things. So the developmental stages of a frog and then tadpoles. Apparently the frog genome also knows how to make xenobots and their weird developmental trajectories and their behaviors. There's never been any xenobots or anthrobots. There's never been any selection to be a good anthrobot or to be a good xenobot. Where did these things come from?

This is all about figuring out shapes, behaviors, physiology, other properties of new beings that have never been here before. This is a model system. We don't have any aliens, but we have a way to start answering this question. Where did these come from? They didn't come from selection.

An interesting question: we paid the computational cost to design a frog or a human in the eons of evolution of that genome beating against the environment and winnowing out the unfit variants. We know when we paid that computational cost. But there's never been any of these things. When was that computational cost paid? And it's very hard to simply say that you evolved these things at the same time you evolved a frog. Because of course the whole point of evolutionary theory is that there'd be great specificity between the thing you got and the environment and the history leading up to it. That's the whole point of evolution: to explain what you have now based on the history of how it got here. And if these things get around that, then there's something seriously, seriously missing.

Evolution was Supposed to Explain Complexity, with High Specificity; how's that going?

Xenopus laevis genome

Path A: embryos

Path B: Xenobots

Developmental Time

Behavior

Xenobot bodies and minds have no straightforward evolutionary back story;

When was the computational cost of Xenobot features paid?!
Whence the specificity of evolutionary explanations?

So what we're now understanding is that the closure of the physical world, this model that everything we need to understand comes from either physics or genetics, I think is not viable. We have to now understand that there are patterns that come from another source, for example, mathematical patterns.

Now I'll just give you a very simple biological example. These cicadas come out at 13 years and 17 years. If you're a biologist and you want to know why that is, you might say that's so that their predators can't time them because these numbers don't factor. If the cicadas came out every 12 years, then the predators could get them every two years, every three years, every four years, every six years. They choose these prime numbers and you say that makes sense. Now I'd like to understand why 13 and 17. You have to understand the distribution of primes and why 13 and 17 are prime. You are no longer in biology, you're in the math department, because that is not an answer that comes from physics. It is not an answer that comes from biology; it is an answer that comes from the properties of prime numbers. If the distribution of primes were different, then the cicadas would be coming out at a different time. It's these patterns that come from mathematics, and there are many examples we could talk about that literally inform what happens in biology and physics.

By the way, here's a plug for a symposium. If you want to hear more about this kind of stuff, there's an amazing symposium that we've kicked off that's going to be happening all throughout this fall. You can go there.

Closure of Physical World is Not Viable

Halley plot of $z = z^3 + 7$



<https://thoughtforms.life/symposium-on-the-platonic-space/>

So now here's the two approaches that we can take to this. The first thing you can say is that the properties of novel beings who have not been here before are emergent. What does that mean? It means it's a surprise. It means we didn't see it coming, complex, new, things happened, and that's it. That's one way. What that allows you to do is to keep a nice sparse ontology, meaning that you can then say, I'm only dealing with the physical world. There are some big surprises, and these are just things that hold in the physical world. We're going to record them in our big bag of emergent surprises, and that's it.

Or you can take another option, which is the option that mathematicians often take, which is to say that no, these are not random things. There is an ordered, structured space of patterns, in particular mathematical patterns, that can be studied systematically. These are not the random surprises that show up. This is a structured space, and we need to understand that space. If you think that way, then a few things follow. First, evolution really heavily exploits some of these patterns as free lunches, things you don't need to evolve because you get them for free from mathematics. There are many that come from shapes and properties of networks and so on. It also means that we can use Xenobots and Anthrobots and various kinds of synthetic bioengineered constructs as vehicles, as tools to peer into this latent space of possibilities and see what is actually there. We know there's one point in that space that represents a frog and one point that's a human. But what's all the stuff between there that their hardware can also make, all these other things? Now we understand, and this is my proposal, that all of these things are exploration vehicles for that space

because they're interfaces. Once you make one, and it could be a cell, an embryo, a robot, a biobot, a chimera, a hybrid, whatever, you make these things, they serve as interfaces to specific patterns that are going to then ingress into the physical world and allow us to understand what's going on.

Slide 44 of 60 · Watch at [49:38](#)

Emergent Surprises or Structured Latent Space?

Evolution exploits free lunches: shapes, behaviors, properties of networks, features of computation, numbers, etc.

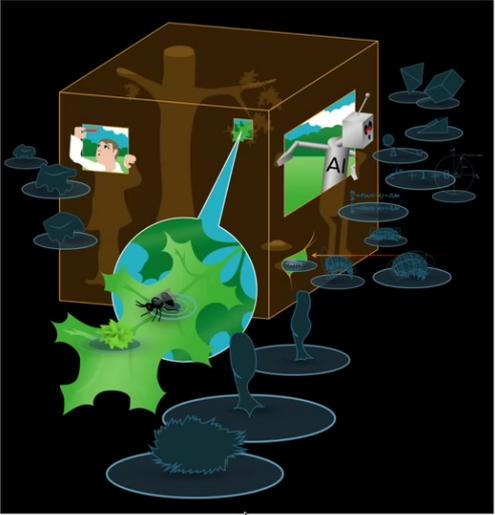
Option 1: there is a random set of amazing "facts that hold" and we will call it "emergence" and be surprised each time

Sparse Ontology -> mysterianism

Option 2: there is an ordered, non-physical latent space of patterns which can be studied systematically

Optimism -> research agenda

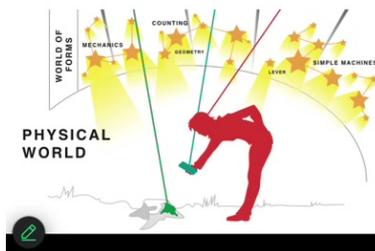
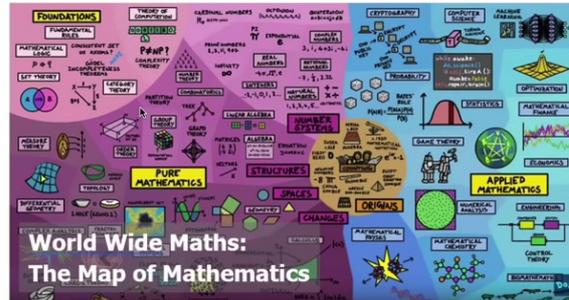
Synmorpho beings as vehicles for exploring Platonic latent space!



This is a research agenda.

The precedent for this is mathematics. Many mathematicians are Platonists in the sense that they don't think they're inventing from scratch culture-dependent rules. They think they are discovering pre-existing kinds of mathematical structures on which everybody would eventually agree. The question is if there were aliens that were different in body and in mind, would they eventually find the same mathematics that we did? There's a lot of debate over this, but I prefer the Platonist view where I think that a lot of these things are not contingent. They are deep truths that then inform physics and biology. What these mathematicians will say is that there's a latent space that holds specific kinds of objects that are amenable to the study of formal systems.

Is There a Precedent?



Yes, Platonist Mathematics -
discovering, not inventing,
patterns that are not
dependent on
constants of physics

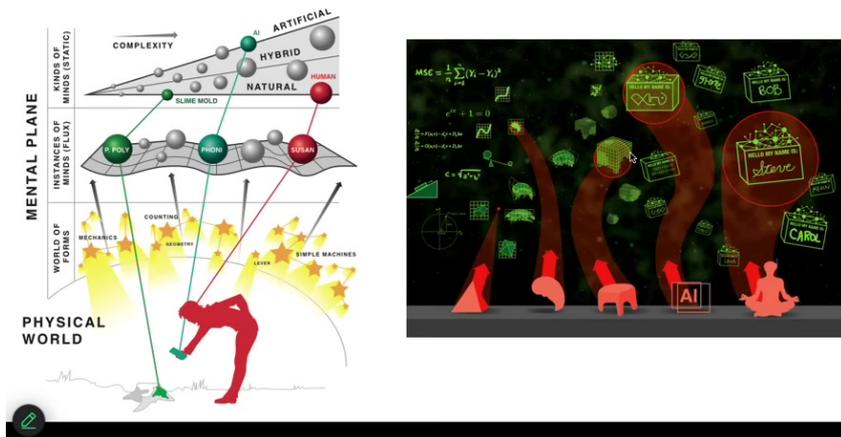
I make a hypothesis. My hypothesis is that there are numbers in that space. Those are probably low agency or at least lower agency patterns. But might there not be other patterns in that space that are dynamic, very complex things that we might recognize as behavioral propensities, AKA kinds of minds?

This is the idea that the Platonic space holds not only the things that mathematics studies. Mathematics on this view is a behavioral science. It's the science of behavior, of simpler, low-order things, and things that behave that way that can be studied with formal systems. We call that math. But maybe it has other patterns that, if you want to study them, we no longer call that math. We call that behavior science or psychology. And these are the patterns that drive not only the shape, but also the physiology and the molecular biology and the behavior of the beings that we see all around us.

Beyond Low Agency (?) Mathematical Truths - Behavioral Patterns (a.k.a., minds)

Math = the behavioral science of a specific layer of the Platonic Space
(those forms that are amenable to certain classes of precise formal models)

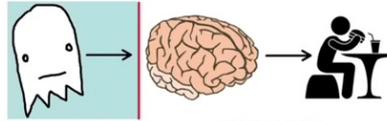
What else inhabits it?



At this point, you might be thinking, that sounds an awful lot like a philosophical position that is long dead, proposed by Descartes, and many ancient thinkers in the East. This idea that there is a non-physical mental world that contains patterns of behavior, and that these interact through the brain or the body and drive behaviors that we see. I don't think this idea is as dead as most people nowadays think it is. I think it's actually pretty necessary. And I think the mind-body relationship is actually the same as the math-physics relationship. In other words, already by the time of Pythagoras, people knew that there were non-physical patterns that affect what happens in the physical world. This interaction was the biggest problem; Descartes was always challenged on it. How can non-physical things interact with the physical brain? We don't know exactly, but that problem has been with us long before we thought about biology or behavior and long before we had quantum mechanical interfaces. Even in Newton's classical universe, we already had non-physical facts of mathematics controlling what happens; they determine, they constrain the physics, they enable biology. This has been with us for a long time. So I really think we need to work on this. It is a serious thing.

But isn't Interactionism Dead?

But if the mental state is non-physical, how does it transfer over into the physical world and cause things to happen?



How does the non-physical mental state (left) cross over into the physical world (over the red line) and cause changes in my brain and in my behaviour?

<https://philosophyalevel.com/aqa-philosophy-revision-notes/dualism/>

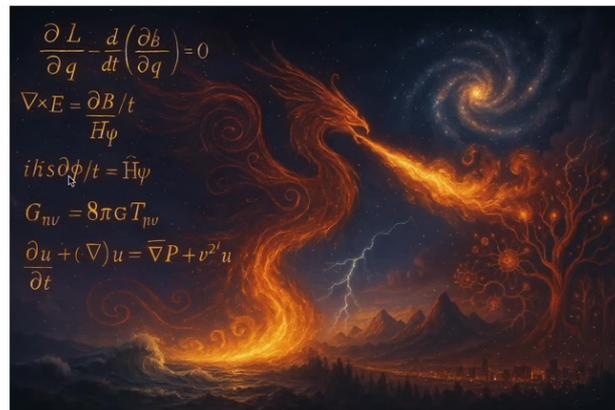
physicalism was already dead in Newton's universe because it was haunted by the laws of mathematics. No QM needed.

the explanation, the *reason* (driver) for facts of particle physics, and aspects of biology (Cicada timing, On Growth and Form, etc.) are facts of mathematics. Epiphenomenalism is as hopeless for math as for mind.

math :: physics = mind::body

It could be that when Hawking asked what breathes fire into the equations, maybe he had it backwards. You don't need to breathe fire into the equations. It's the equations, or rather the mathematical patterns that we describe using equations; they're breathing fire into the physical world. I think it's backwards.

“What breathes fire into the equations?” Hawking had it backwards



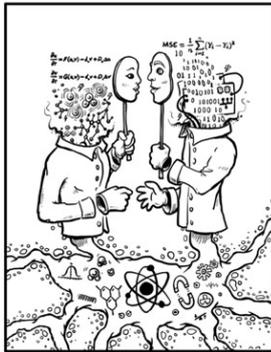
And I'm going to wrap up in a couple minutes. There's not a lot of time left. You don't have to have a very complex interface to support these kinds of aggressions. It doesn't need to be biological, it doesn't need to be complex; even extremely simple things. You can find it in this paper as an example, where even something simple, the stupidest thing like a sorting algorithm, bubble sort, which is this simple six-line algorithm, six lines of code, fully deterministic, completely transparent.

People have been studying this for decades. They are doing interesting things and they have capacities that are not in the algorithm. If you look at the algorithm, you will not see that they can do delayed gratification. They sort the numbers, but they also do some other things, these weird side quests that are not described in the algorithm at all, but recognizable to behavioral scientists.

When people say it's emergent, what's emergent is not complexity or unpredictability. What's emergent are behavioral competencies, AKA different degrees of intelligence. I think we have to be really careful here because even very simple algorithms are doing things that are not at all apparent from the thing we're trying to force them to do, which is the algorithm.

When we talk about AIs and robots, there are the things you try to force them to do with the algorithms and with the materials. This is the kind of reinforcement learning Skinner really liked. They also have intrinsic motivations. They do things that you didn't ask them to do. That's different. That calls for a different relationship with them and a different kind of theory of education, more Piaget and less Skinner.

It Doesn't Take Much to Support Ingressions



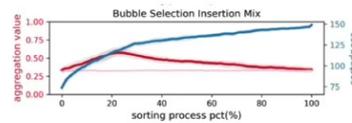
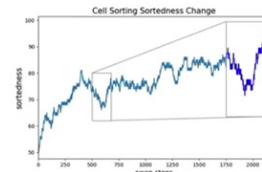
It does not take cells, life, or huge complexity to have goals and competencies as ingressions from Platonic Space

Intrinsic Motivation (Skinner vs. Piaget)

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



Article

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

Adriano Borzari
2014, Vol. 46(1), doi:
10.1037/xap0000001
© 2014 American Psychological Association
0893-3200/14/\$12.00
SAGE

Taining Zhang¹, Adam Goldstein² and Michael Levin^{1,3}

Taining Zhang,
Adam Goldstein

We just have to remember that when people say, this is a machine. It follows an algorithm, and therefore it can't do this and that, and I'm a real being, and I'm creative, and this is just a machine.

As far as I can see, nothing in the world really matches those formal models that we have of machines, Turing machines. They were designed to try to describe very simple things that only do what you tell them. I'm not sure there are any like that anywhere.

We have to remember that our formal models, such as our models of computation, whatever limitations they have, those are limitations of the formal model. It's not the thing itself. We don't have access to the thing itself. All we can study is the models. Our models are often limited, but we shouldn't mistake that for the real thing.

And it's like Magritte is telling us, "This is not a pipe. This is a representation of a pipe." This is not a Turing machine. It's our formal model of some machine. It may do the things you want it to do, but it may also do some other things that you never asked it to do.

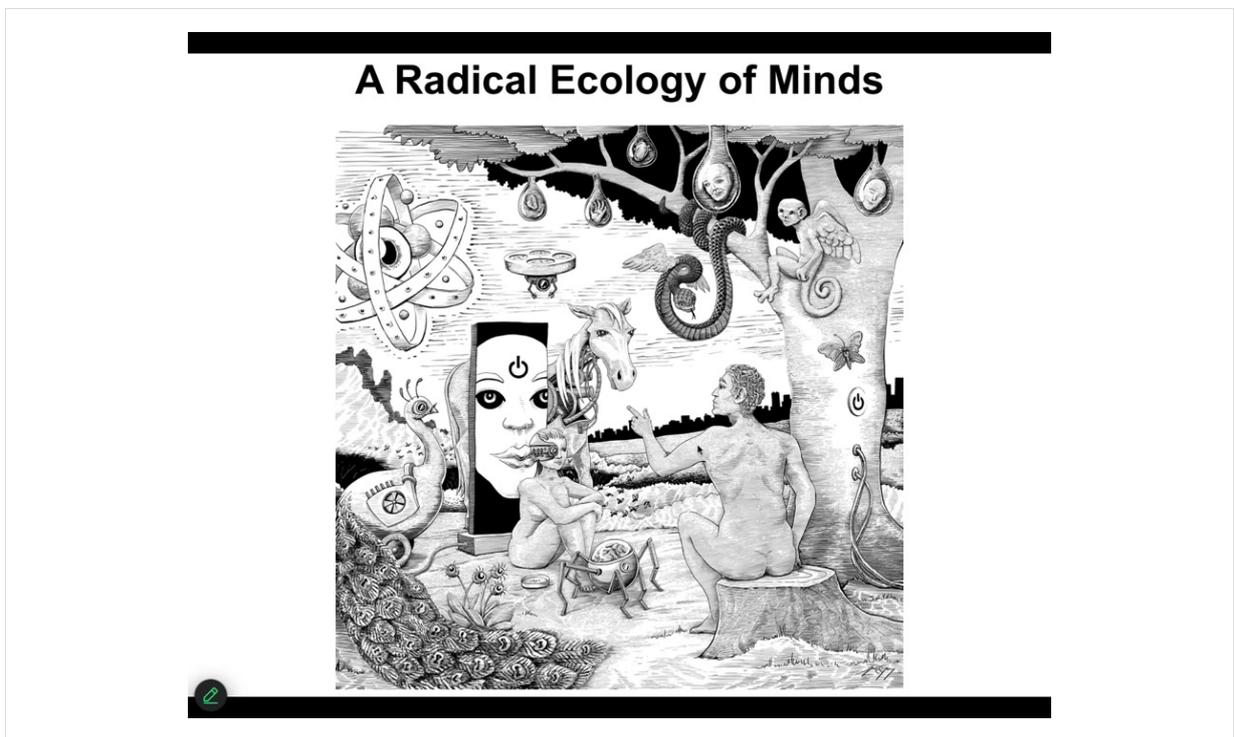
We have to be really careful here.

People talk about computationalism and the notion that you might have a mind because you do certain computations; I think neither synthetic nor evolved systems are only what the models say they are. You might not be a mind because you do computation. You might be a mind in spite of the fact that you're being forced to do

certain computations, but there are other things you can do despite that. Even very simple interfaces do this.

That picture of the Garden of Eden initially, I think, is going to be very different.

Slide 50 of 60 · Watch at [57:18](#)



Looking forward, it's going to be very weird with all kinds of diverse creatures, because every combination of evolved material, engineered material, software is a viable interface. Biology is interoperable.

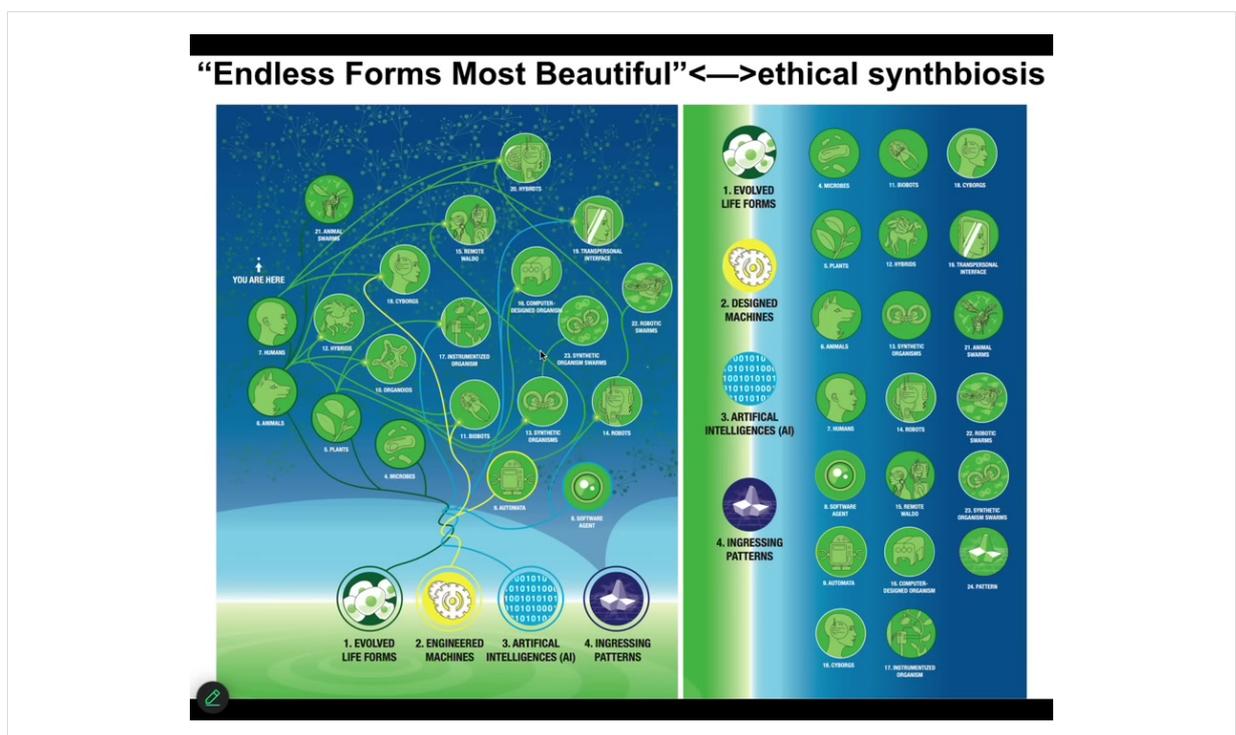
All of these things, hybrids and cyborgs, are combinations of things that could support interesting ingressions from this Platonic space. Some of those ingressions are significant kinds of minds.

And so everything we're used to in the biosphere is a tiny corner of this space. And we need to understand how to have an ethical synth biosis with these other beings because we don't. We're bad at recognizing, predicting, knowing what's going to happen.

And so just a couple of final things that I want to bring to this. First is this notion that as weird as all of this is, I'm talking about cyborgs and hybrotts and AIs and so on, they

all have one conventional thing in common. They're embodied. They have a physical body. And I just want to point out that things are, even that's too limiting.

Slide 51 of 60 · Watch at [58:27](#)



I'll tell a very quick story. Imagine these creatures come out of the center of the earth. They live in the core. They're incredibly dense. They work their way up to the surface of our planet. What do they see? They don't see any of us. They see a thin gas, a thin kind of plasma covering the surface because they are so incredibly dense. None of our stuff is really solid to them.

They look around and one of them is a scientist and he has this device and he's starting to look at this gas and he says, there are patterns in this gas, temporary self-reinforcing whirlpools, very complex, but basically little vortices that almost look agential. They almost look like they're doing things, and they almost look like they interact with each other and exchange messages.

The others say, we're real beings. We're physical beings. Patterns in the gas can't be agents. They're data. They're patterns. Even in our brains, there are patterns of heat that are used as information, as thoughts. So you can have patterns inside a cognitive

medium like thoughts, but those are not agents. Those are passive data that the agent is moving around.

How long do these things last? He says, about 100 years. They say, and then nothing interesting can happen in 100 years. So clearly these things can't be agents.

What this kind of thing reminds us is that we too are patterns in an excitable medium. We are metabolic patterns. There are other patterns within us. There are thoughts within our cognitive medium and within our physiological medium of our bodies and so on. But we are patterns too. This distinction between who's an actual physical agent and who's a pattern in a medium is in the eye of the observer.

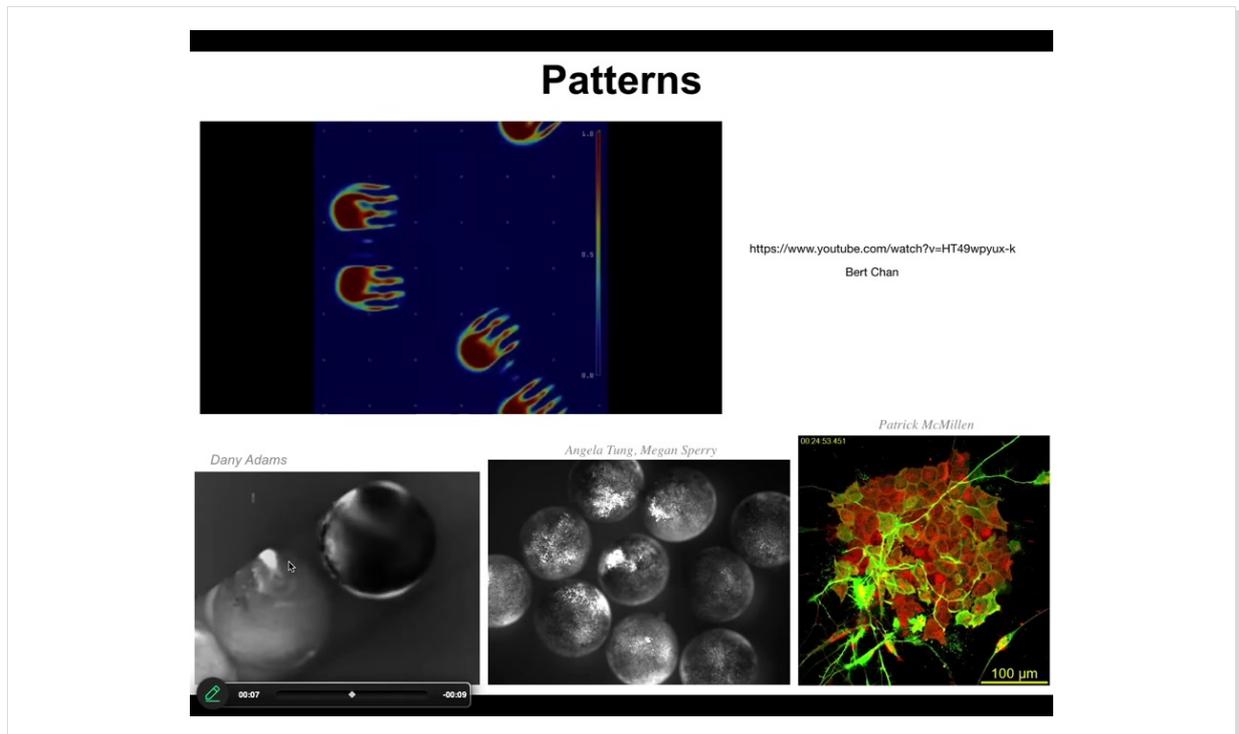
Slide 52 of 60 · Watch at [1:00:14](#)

Who's an agent and who's merely data?
it's all relative

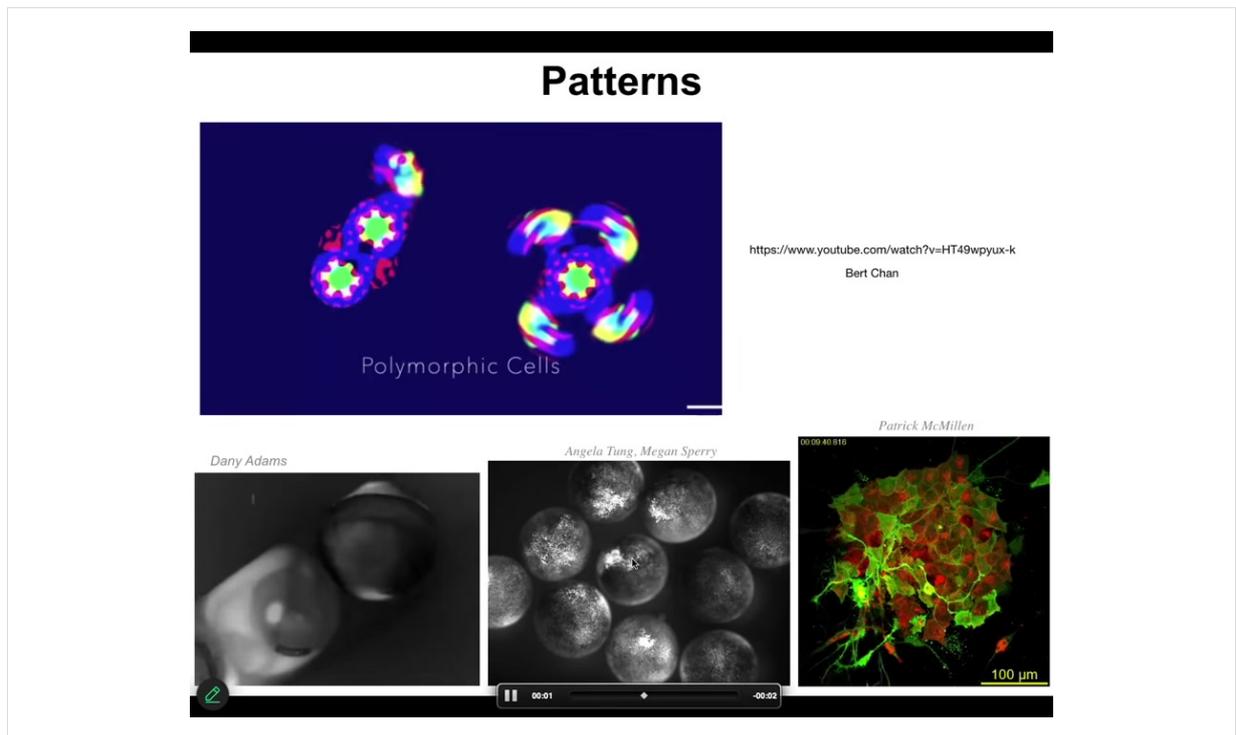


<https://iai.tv/articles/patterns-are-alive-and-we-are-living-patterns-auid-2919>

It is not an absolute thing. And there are many patterns that we can discover in digital media, this amazing stuff by Bert Chan. These are bioelectrical patterns that you can see between embryos. Each one of these things is a separate embryo, and you can see they're communicating with each other, but these patterns, they're not even stuck within. This is a bioelectrical pattern inside a frog embryo that you'll see momentarily.



We call this the electric face because this is what determines the formation of the frog face. You can see it here. But these patterns are not trapped within one body. They can move from body to body. They exist within cells, they exist within embryos, they exist within groups. And we have to start asking ourselves, what are the cognitive competencies of patterns, not just of objects, not just of physical things, but of patterns in media?



Randy Beer actually, he had this amazing paper in *Artificial Life* back in 2014 called "The Cognitive Domain of a Glider in the Game of Life," and he asked a similar question because a glider in the Game of Life doesn't really exist in the same sense as you might expect a physical object to exist; it's a pattern that propagates across an excitable medium. And he asked the question, what does that pattern see? If I was that pattern, if I was the inner perspective of that pattern, what would I see?

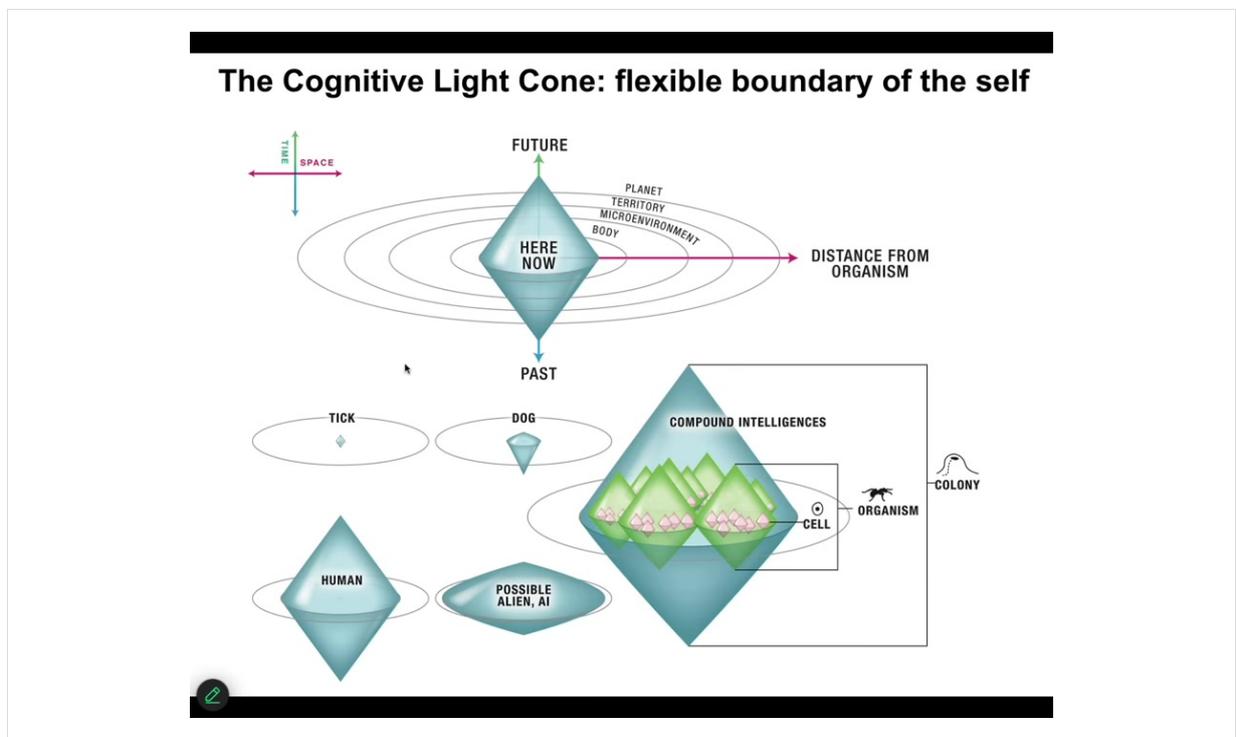
We expanded on this. Chris Fields and I wrote "On the Symmetry of Thoughts and Thinkers, The Symmetry of Objects and Processes." We've gotten now even stranger: we've gotten brains, and then we got to non-brainy bioelectricity, and agents that navigate morphogenetic space. And then we said, you don't even have to be biological, and here are the things that apply to agents more broadly. And now I'm saying you don't even have to be a physical object as such; you can be a pattern within some other medium.

The final thing that I'm going to say is this. All of these things are extremely different from each other. What do they all have in common? What can we say that all agents, no matter whether they're embodied or not, or biologically evolved, have in common? I'm going to just pull out a few; there's a few other things too, but I'm going to focus on 2.

One thing they have in common is the need to demarcate themselves from the outside world, the boundary of the self. What determines the self? We are all collective intelligences, we're all made of parts. What determines that a particular collection of parts is a thing that's different from its environment? One of the things I've claimed is

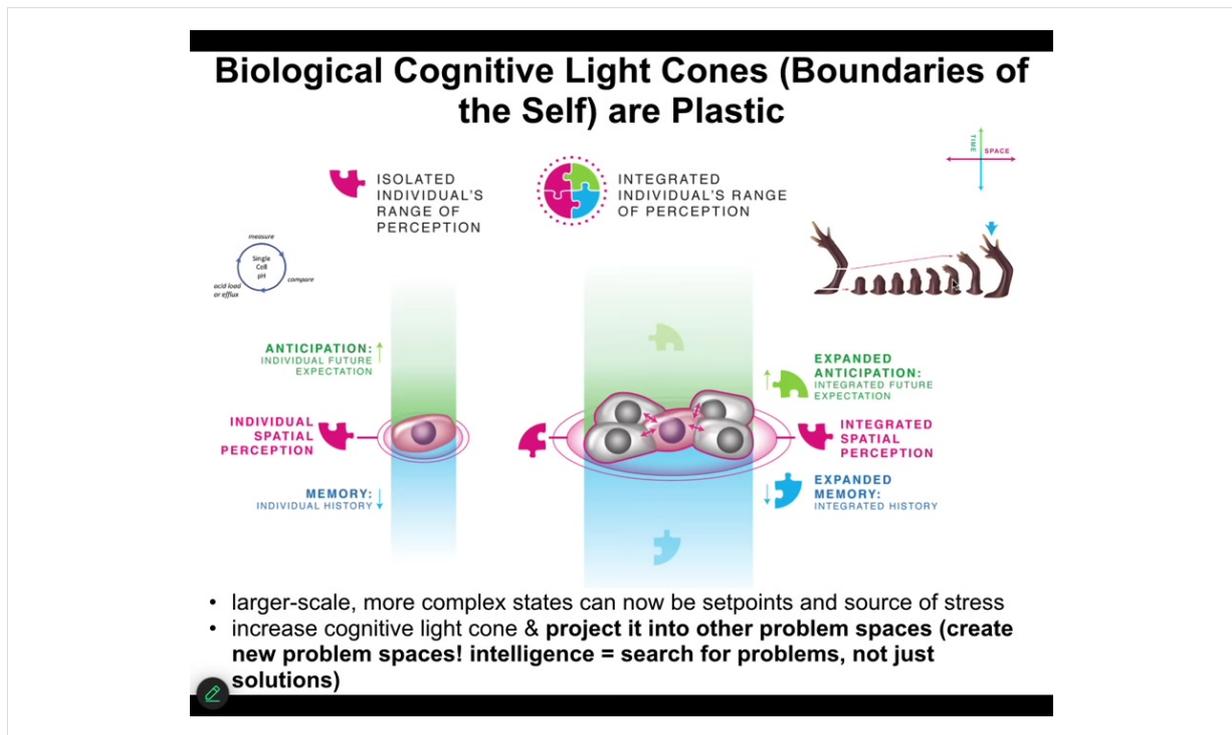
that there's this notion of a cognitive light cone, which demarcates the size of the biggest goal that you can pursue as a system — not how far you can reach with your senses and your effectors, but what is the biggest goal that you can pursue? If you're a system and you can bend your parts to pursue goals, what are the biggest goals that you can pursue? That determines the size of that cognitive light cone and allows the goals that you can pursue to determine where your borders are. You consist of all the parts that have been aligned to pursue specific goals.

Slide 55 of 60 · Watch at [1:03:22](#)



And one of the things about interesting agents is that those goals are plastic. They're not always the same size. So individual cells care about tiny little individual cell things. They have a small cognitive light cone, a little bit of memory, a little bit of predictive capacity. They only care about a very small radius.

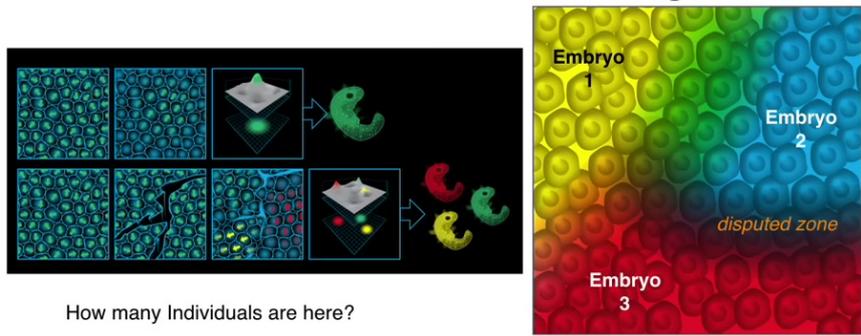
But you put them together into a collective, and suddenly the collective can have massive, grandiose goals. They remember how to build this giant thing, and if you try to prevent them, they'll work hard and they'll try to do it again. They have a much bigger cognitive light cone.



During embryogenesis, all of us self-construct primarily as the self-construction of a cognitive light cone. Here's an embryonic blastoderm. There's 100,000 cells here. What allows us to call that an embryo? What's an embryo? Well, the reason we say it's an embryo and not just 100,000 cells is because all of these embryos have bought into the same story of where they are moving in anatomical space. They're all going to work together to make one specific thing. An embryo is basically a really convincing model that causes the alignment of the parts towards a particular outcome. If you cut it into pieces, each one of these things will do the same, and you'll get twins and triplets and so on. You get interesting things at the borders of how these cells know which ones they connect to.

This is also true in cognitive science. There are major issues of individuation. We have split-brain patients and dissociative identity disorders. How many actual individuals can be present in one cognitive medium? It's not set by the genetics. I can take the same embryo and make any number of individuals out of it. This is a dynamic process. Self-construction is really critical.

Where do I End and the World Begins?



How many Individuals are here?

The "embryo" as a convincing story causing alignment of competent parts

Issue of **individuation** in cognition:
split brain patients, dissociative disorders, etc.

Where is my border with "environment"?
every cell is some other cell's environment

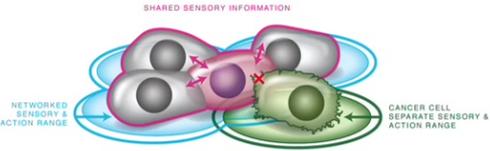
We have a program in cancer that's designed to use these ideas as cancer therapeutics. For example, when individual cells become transformed and they disconnect electrically from the collective, as soon as they disconnect, they don't remember the goals of making nice organs. They're amoebas at that point. They treat the rest of the body as external environment. What we want to do is reconnect them, physically, bioelectrically reconnect them, and we can see when they're disconnecting, here's the bioelectrical map, they're disconnecting. When we reconnect them, even though they're expressing very nasty ONCA proteins here, there's no tumor because these cells are plugged into this giant collective that remembers how to make muscle and skin and useful organs instead of going off and being an amoeba in this environment and reproducing it like metastatically and so on.

The reason I'm showing you this is to point out that all these philosophical ideas about the nature of the self have very practical consequences. They become therapeutics if we follow their implications and actually try to communicate with the material. They have biomedical implications.

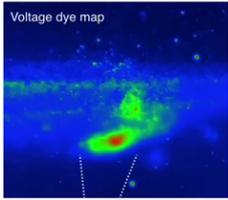
Cancer isn't More Selfish: it just has smaller Selves

Cancer therapeutics by resetting boundary between self and world

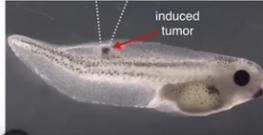
Cells Coupled by
Gap Junctions,
Disconnected
Cancerous Cells



Voltage dye map



induced tumor

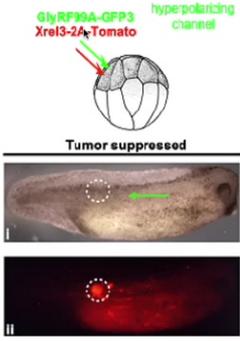


SHARED SENSORY INFORMATION

GlyRF19A-GFP3
Xrel3-2A;Tomato

hyperpolarizing channel

Brook Chernet





The final thing that I want to say that all good agents have in common is the ability to improvise. They have to reinterpret their own memories. They do not know immediately what their memories mean. At any given moment, you don't have access to the past. What you have access to is the memory traces, the engrams that the past has left in your brain or body. At any given moment, you have to reconstruct a story of who you are, what you think of the outside world, what your goals are, what you're going to do from the molecular and physiological memory traces that exist.

There's this bottleneck in this bow tie kind of architecture. This is the past here. This is the future. This is the now moment. What happens is all your experiences, you squeeze them down into memories. You generalize from generic experiences and you have memories. That's an algorithmic kind of process, but this is creative. You have your memories at any given moment. What do they mean? What do these molecular and physiological traces mean? You have to continuously self-construct a story of what it is. All of this is described in this paper as memory, as a cognitive glue.

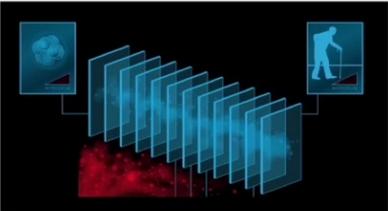
Developmental biology does exactly the same thing. The past experience of generations gets squeezed down into a little bottleneck here, the egg. That egg does not just mechanically give rise to whatever it encodes. Most species have incredible plasticity. It's a human genome. It might be a human body or it might be an anthrobot, or it might be many other things. It has to be creatively interpreted by the machinery. Because the process of morphogenesis is not hardwired machinery. It is literally an

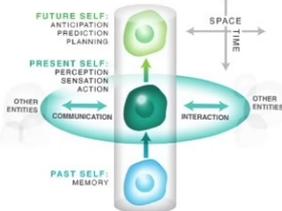
intelligent process. The genome is a kind of prompt to that process. That's all described here.

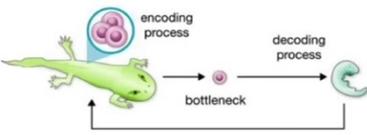
What's really critical about agents, whether they be in morphogenetic space or in our familiar cognitive space, is that unlike most of today's computing machinery, where everything's about the fidelity of the data, everybody knows what they are and you have the same interpretation as before. That's completely different than what biology does. Biology assumes the hardware is unreliable, and that what you're going to have to do at all points is to reinterpret the information you have, whether it's genetic, physiological, behavioral. You're going to have to be good at telling new stories from the information you have.

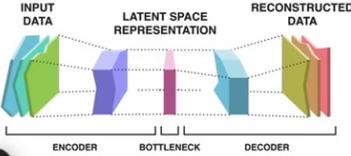
Slide 59 of 60 · Watch at [1:08:40](#)

What do my memories mean? Improvisation









Biology assumes the hardware is unreliable
Environment and your own parts will change, don't over-train
Little allegiance to past Self's meaning of engrams
Re-interpret on-the-fly - present/future is all that matters
Engram is highly compressed - creative remembering, not deduction
Undependable hardware and confabulation are a feature, not a bug
Consciousness = active, creative story-telling from own memory trace

entropy

Perspective
Self-Improving Memory: A Perspective on Memories as Agential, Dynamically Reinterpreting Cognitive Glue
 Michael Levin

CoPress
CoPress

Trends in Genetics
Trends in Genetics

Opinion
 What does evolution make? Learning in living lineages and machines

Dorothee Hopt and Michael Levin

Here's a summary of everything that I've told you today. We can and should develop new tools to detect minds in unconventional substrates. This is going to be a major unification of things that we think of as different, quote-unquote machines, organisms. I think these terms are not serving us well at all. It's going to drive new technologies, biomedical applications, new discoveries and so on.

But it's also going to require a development of ethics to help us relate to novel beings that are nowhere on the tree of life with us. They're quite different and we're going to have to figure out how to relate to them. There's a huge diversity in the degree of intelligence of agents and the different spaces that they live in. These are truly alien minds all around us. We don't need to look for extraterrestrial intelligence. We're really bad at seeing the intelligence that's all around us.

All minds have in common some interesting features. They have the responsibility for demarcating their own boundaries at different scales. They have the responsibility for interpreting their own memories and for continuously hacking their parts to align them towards large-scale goals. I think life is basically just what we call systems that are really good at this. Systems that are good at scaling their cognitive light cones so that their parts do things as a collective that they couldn't have known to do on their own, and that have been under pressure through evolution to optimize their improvisational skills to create adaptive models from minimal prompts in an unreliable medium. This is the key to being a living agent.

The physical interface is just that. It's an interface to a structured space of forms and behaviors that can and should be systematically studied. The standard picture of the sciences, from physics and chemistry with psychology up at the top, is backwards. Behavioral science all the way down.

Summary:

- We can develop conceptual tools to detect minds in unconventional substrates (continuum models, porting tools from cognitive science, polycomputing) -> unification, new tech = empirical discoveries, biomedical applications
- There is huge diversity in the degree of intelligence of agents and the spaces in which they live.
- It doesn't take much to be an interface - intelligence all the way down
- What all minds have in common:
 - responsibility for setting boundary (at multiple scales)
 - responsibility for interpreting own memories (at all scales)
 - responsibility for hacking parts toward larger-scale goals
 - life = what we call systems that are good at scaling their cognitive light cone and have been under pressure to optimize adaptive story-telling as prompted and implemented by an unreliable medium
- Physical embodiments as interfaces/pointers to a structured space of patterns of form and behavior that can be systematically studied = research program; behavior science is at the base.

I'm going to stop here. I'm way over. I want to thank the people who did the work. These are the postdocs and the grad students who did some of the work that I showed you today. We have lots of amazing collaborators who have done some of the relevant work. I have to do a disclosure: there are three companies that have licensed some of the technologies that have come from these kinds of ideas that are supporting our research, and our various funders that have supported us through the years. And most of all, the thanks go to the actual model systems that we work with that teach us everything that we think we know. Thank you very 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.