

# Cultivating Strategies

David I. Spivak

Finding the Right Abstractions  
Topos Institute  
May 2021

# Outline

## 1 Introduction

- Gathering together
- Strategies for our future

## 2 Abstracting life

## 3 Abstractions for the AI transition

## 4 Conclusion

## We are gathered here today...

The Long-Term Future Fund donated \$50k to convene this summit. Why?

- It brings together people with many skillsets and viewpoints.
- We exchange ideas, and the ones that fit us take root in us.
- If successful, we leave with new plants in our garden.

But what is this garden for?

- We want both ourselves and humanity to survive and flourish.
- But not everything about us needs to survive.
  - Some things can go: racism, wanton pollution, the vacuum tube.
  - Our values often improve, so we wouldn't want to lock those in.
  - The human form—our appendix, our height—isn't too important.
- So what is it that we actually want to live on?
  - What's a weed, and what's a plant?
  - Or as Owen once asked: What's cancer, and what's wisdom?

We are gathered at FRA to improve how we think about these things.

# Good abstractions

What are good abstractions in general?

- Good abstractions carry little content but hold a lot of structure.
  - You see this in good (concise, reusable, maintainable) code.
  - You also see this in category theory.
  - You see something analogous in matter itself.
- They should be dextrous, handling corner cases in stride.

What are good abstractions for thinking about survival and flourishing?

- Since we're doing it right now, they should fit this activity here.
- They should fit the origins of life and intelligence.

# Cultivating strategies

If I only knew what the title of my talk meant, mathematically, I'd be done.

- To me, cultivating strategies (skills, techniques) is central to life.
- My main purpose is to demonstrate to you what I mean.
  - It's what we're doing here at FRA.
  - And I'll explain the sense in which we've been doing it all along.
- I'll also briefly touch on my progress toward formalizing the idea.

I'm still missing an essential feature: what drives it.

- What moves us to cultivate strategies, to learn, to grow, to flourish?
- A common response is “we're driven to maximize utility”.
  - This puts a lot of content in some intrinsic ordered thing like  $\mathbb{R}$ .
  - I'll explain on the next slide why it completely misses my needs.
  - Much of this talk is an attempt to naturalize it.

## My gripe with utility

The “utility function” concept doesn’t yet fit in the garden I cultivate.

- Utility functions are defined on a measurable space. What is it?
  - Call it the *possibility space*.
  - Is it  $\mathbb{R}^{10000}$ ? What are the axes?
  - Perhaps it’s weird and stringy, or not a measurable space at all.
  - If this talk maximizes utility, imagine the complexity of the space.

The structure and formation process for these spaces is probably quite rich.

- To me, the space is primary, coming before the measures on them.
  - This space grew organically.
  - They grew out of interaction with parents, village, world.
- We should understand the shaping process for these spaces.

The term utility implies usefulness. Useful for what?

- Before we put a number on “usefulness”, let’s examine what it is.
- Let’s consider how it could arise in the first place.

# Outline

## 1 Introduction

## 2 **Abstracting life**

- Origins
- Survival strategies
- Channeling negentropy

## 3 Abstractions for the AI transition

## 4 Conclusion

# Instincts

Worms are born behaviorally complete; we're not.

- Humans are born with a few reflexes.
- Rooting reflex: when your cheek is brushed, turn and suck.
- Blinking reflex: when your eye is touched or you see bright light, blink.

But less is instinctual than you might think:

- Example: throw a pingpong ball at an infant's face and it won't blink.
- It's not instinctual to see walls as hard or even there!
- We build our whole world. Almost nothing comes with the package.
- We even create the distinction of self, what I claim as me.

What makes us willing and able to learn?

- What very basic thing in us makes this possible?
- What instinct leads us to structure our possibility space?
- We cultivate our understanding of the world and our place in it.



# Origin of life

Let's go back even further and consider the origins of life.

- Not much is known, so it's all about plausibility.
- I'm certainly no expert, so take this slide with a grain of salt.

Simple life forms exist in the context of potential differences. (Eric Smith)

- Life may have emerged in deep sea hydrothermal vents.
- $CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$ , high-energy  $\rightarrow$  low energy.
- The chemistry is just difficult enough that it requires life.
  - Living systems act as catalysts for particular exergonic reactions.
  - They can direct free energy toward nonspontaneous reactions.

Let's refer to what early life found as a *strategy* for dissipating energy.

# Genes

The mechanisms necessary to dissipate energy need to be kept.

- At some point we developed the ability to save them genetically.

Genes encode strategies for survival, enacted in real time.

- Genes are the code for protein production techniques.
- As your situation changes, the genes you express change with it.
- Proteins do everything, e.g. contract muscle, resonate with light.
- I'm referring to the way you deal with a situation as a strategy.

And genetics itself is a strategy that cultivates strategies.

- The conversion of DNA into proteins involves a ton of chemistry.
- Terribly complicated sequences of atomic interactions.
- Evolution made this protein-production chemistry legible as code.
- Now selection can act on the code rather than on atomic interactions.

# Schmidhuber's notion of beauty

Let's get back to us as experiential subjects (one who says "I").

- Schmidhuber's says we experience information as beautiful...
- ...when it helps us compress information we were already holding.
- This lightens our load, without losing anything we need.

Examples and similar concepts:

- Science: summarize experimental data, summarize theories.
- Category theory: compress mathematical information.
- The internet: no need for all those books and libraries.

Compression is a type of *strategy cultivation*.

- But again, what are all these strategies for?
- What is the "information we were already holding" about?

## Channeling negentropy

What's common between simplest lifeforms and most complex social orgs?

- It's plausible that the function of life is to dissipate free energy.
  - Ants use smell systems to locate the picnic scraps and devour it.
  - Humans use language systems to locate the oil underground.

Or perhaps we are here to produce entropy.

- We are constantly channeling negentropy into—organizing—ourselves.
- Not just any organization, but that which will help us do it again.
- But as we organize ourselves, we end up dissipating more entropy.
- Ilya Prigogine referred to life as a *dissipative structure*.

Perhaps an agent is the cultivation of strategies for entropy dissipation.

- In this way we contrast it with fire, and maybe even cancer.
- The cultivation itself is the work spent collecting, compressing, ...
- ... fine-tuning, and replicating the strategies that comprise the agent.

## Strategy replication

Dawkins talks about the “selfish gene”.

- Genes—protein production strategies—seem to want to replicate.
- Memes—concept production strategies—do too.
- Maybe it's more like “selfish strategy”? Good strategies are copied.

To that end, maybe we can naturalize the utility concept.

- Utility is just our/an agent's way of understanding what's good for it.
- “Good for it” is itself just a proxy for more agency.
- We suggested agency is cultivation of entropy-dissipation strategies.
- Utility is then just an understanding of how that process works.

The genes and strategies don't care so much about the individual agents.

- It's the ecosystem of experiments that really persists.
- Even the agent works on behalf of its most useful strategies.
- Bundles of strategies form and dissolve, but the strategies replicate.

This might give us a clue about what we can work towards.

# Outline

- 1 Introduction
- 2 Abstracting life
- 3 Abstractions for the AI transition**
  - Artifacts of intelligence
  - Mathematical abstractions
- 4 Conclusion

# The AI transition

I don't understand the term "artificial"; to me everything is natural.

- Other animals—mammals, crows, octopuses—use tools.
- Tools are natural. Language is natural. Water wheels are natural.
- We are part of nature.
- Evolution has always driven life to find more efficient strategies.

Rather than artificial, let's talk about *artifacts of intelligence*.

- Humans try to mimic intelligence they see in animals and people.
  - Example: "Computers" were originally people.
  - Turing explicitly designed machines to mimic their behavior.
  - We capture our understanding of life/intelligence in artifacts.
  - We can then run it continuously and at very fast rates.
  - I'll call these fast-running artifacts of intelligence "AI".
- This will change everything, but it's not divergent from evolution.

It's also "natural" to want this transition to go well for us!

# Collective intelligence

I imagine AI as interpenetrating human activity, not as individualized.

- Maybe there will be an individual AI machine with desires, maybe not.
- But we already see artifacts of intelligence all around us.
- Computers, databases, grammar checkers, recommender engines.
- These thread through our lives, cross cutting everything.

Email, zoom, and google docs lead to an ability to think-together.

- Joey quoted “it takes many brains to make a mind”.
- This gathertown platform lets us exchange info very fast.

What's emerging isn't a super-intelligent AI. It's collective intelligence.

- It can still be quite dangerous.
- Like money, it can concentrate human greed and fear.
- We mercilessly exploit the labor and bodies of people and animals.
- We could kill ourselves completely by inventing an amazing new virus.

We need math to understand collective intelligence and its effects.




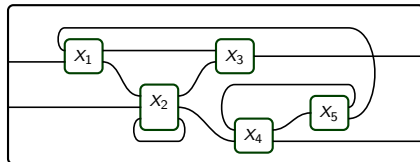
## A useful abstraction: dynamical systems

We always seem to be on the outside of things, studying them with radar.

- Light and sound and force reflect off of them, and that's what we get.
- But their activity suggests an internal state that we have no access to.

Taking this seriously, I like the open dynamical system model.

- Think of a thing as having an interface  by which it interacts...
- ...and internal states that are unseen, but that govern its responses.
- The interface may change form (close eyes) based on internal states.
- The interaction pattern between systems is like a wiring diagram...



- ...but think gathertown: our activities can change the wiring pattern.

All this has beautiful theory behind it, called *polynomial functors*.

# Polynomial functors

Again, the math of polynomial functors fits the above story quite well.

- The same formalism that defines interacting dynamical systems...
- ... also subsumes deep learning via gradient descent and backprop.
- A general formalism for compositional interacting systems.

If you're interested, you can see talks on youtube, or I'd love to discuss.

# Recruitment and vibing?

I wonder about how to mathematize *recruitment*.

- What's happening when I eat food, or a company hires a worker?
- Or when we use a tool, or even create one?
- When we understand another system, we can lead it (OODA loops).

I also wonder about how to mathematize *vibing*.

- People resonate with each other and can form teams.
- As Alex says, vibing offers an alternative to control.
  - Does the thermostat control the room temp, or vice versa?
  - Does the human breed the tulips and potatoes, or vice versa?

Can we see recruitment or vibing in the context of open dynamical systems?

# Outline

- 1 Introduction
- 2 Abstracting life
- 3 Abstractions for the AI transition
- 4 Conclusion**
  - Maintaining balance
  - Navigating the AI transition

## Serenity prayer

A version of Niebuhr's serenity prayer says:

*God, grant me the serenity to accept the things I cannot change,  
courage to change the things I can,  
and wisdom to know the difference.*

Abraham Lincoln is quoted as telling an advisor:

*My concern is not whether God is on our side;  
my greatest concern is to be on God's side, for God is always right.*

I don't want to control AI; I want to align with that which creates us both.

## Philosophy with a lifeline

Probably—based on empirical and scientific evidence—we're all gonna die.

- But we're here because we want *life* to be good.
- We enjoy this moment more when we're not actively killing ourselves.
- And some human activity is creating the poisons that can kill us.
- In his talk *Philosophy with a deadline*, Critch will encourage us to care.

But the AI transition is coming, like water rolling down a hill. What to do?

- Use philosophy to make a lifeline; a future compatible with us.
- Name the future: a plausible future to work towards.
- There's hope that elegant AGI, if first, would outcompete control AI.

# Navigating the AI transition

I want the most flexible abstractions for handling whatever is coming.

- As the world gets more complex, anomalies will become the norm.
- Intelligent artifacts multiply the complexity enormously.
- I'd like to be on a team that actively works to navigate through that.

I want to be involved with *making math for the mothership*.

- It should be both elegant and practical.
- It should assume no more compute than we have today.
- It should offer sense-making and organizational power.
- It should be about life, experience, agency, communication, ...
- ... systems, economics, ethics.

We should ground philosophy on solid math so we know what's being said.

## What do we want to live on?

As humanity evolves, what do you want to live on?

- Is it our body form? An insistence on that seems too nostalgic.
- Is it our history, our record of beautiful or amazing works?
- Consciousness is hard to define, but that's what I'd choose.

I think it's fair to hope that the best aspects of us live on.

- Let's try to articulate what about us really should live on.
  - Our contributions, what people appreciate about us, what works.
  - It's fairly natural for these to be replicated, cultivated.
- That's compatible with AI that wants to cultivate the best strategies.

If we aim for that pairing, I think it can work naturally without struggle.

*Thanks; looking forward to hearing your thoughts!*