

# Reflexive Organism Model

Uroš Jovanović

2025-10-17

## Contents

<b>How it all started? Ant colony as a superorganism</b>	<b>3</b>
<b>Superorganism as a fractal</b>	<b>4</b>
Forest as a superorganism . . . . .	4
<b>Reproduction unit problem</b>	<b>5</b>
Ecosystem as Superorganism Life Cycle . . . . .	7
<b>The case for the superorganism view</b>	<b>8</b>
<b>Superorganism as “I”, a poetic tale</b>	<b>9</b>
<b>Towards the superorganism model</b>	<b>10</b>
Colony as a superorganism . . . . .	11
Forest as a superorganism . . . . .	12
Unification, a general model . . . . .	14
What is this model all about? . . . . .	16
<b>Extending the general model with Assembly theory</b>	<b>19</b>
<b>Introducing the “I”, experience through reflection</b>	<b>24</b>
<b>Going full Boltzmann</b>	<b>27</b>
The Boltzmann Experience . . . . .	29
Oh no, what have we done? . . . . .	31
Tidy ants . . . . .	32
<b>Gravity</b>	<b>33</b>
<b>Revisiting electromagnetism</b>	<b>39</b>
The observer . . . . .	40
Information and fields . . . . .	41
<b>Hierarchical fields forever</b>	<b>42</b>
Biological evidence: hierarchy & smoothing . . . . .	43
Scaling to galaxies . . . . .	44
<b>Reflection</b>	<b>45</b>
<b>Communication in the hierarchy</b>	<b>45</b>
Temporal dynamics . . . . .	47
Operators . . . . .	50
So smooth . . . . .	52
<b>Cycles</b>	<b>52</b>
<b>Hierarchical time</b>	<b>53</b>
<b>Rewards</b>	<b>54</b>
Rewards through hierarchy . . . . .	55

Extinction as loss of reflexive coupling . . . . .	58
<b>Altruism vs. selfishness</b>	<b>58</b>
Beyond good and evil . . . . .	63
Minority report . . . . .	64
<b>Computational geometry</b>	<b>65</b>
Waves, rotations, spirals . . . . .	68
Forces . . . . .	69
<b>Organism as a spiral of potential self</b>	<b>70</b>
<b>Explorations</b>	<b>70</b>
Entropy revisited . . . . .	70
Arrow of time . . . . .	72
Parent-child information compression . . . . .	73
Antifragile . . . . .	74
<b>Individuality as dynamic invariant</b>	<b>75</b>
<b>Geometry of self-reference</b>	<b>77</b>
Example . . . . .	79
Organism's unified field tensor $T_{\mu\nu}$ . . . . .	81
Appendix to a $\mathcal{S}_{AT}$ model . . . . .	83
Compatibility drives integration . . . . .	83
<b>Sentient organism</b>	<b>84</b>
<b>Acknowledgements</b>	<b>89</b>
<b>Bibliography</b>	<b>90</b>

Copyright © 2025 Uroš Jovanović, CC BY-SA 4.0.

The document explores the concepts and themes of an experiment to define a self-aware, reflexive organism, that can describe itself as a superorganism composed of other organisms and is not bound to the number of hierarchical levels nor density of hierarchies. It is a first real try to formulate what I have been learning about known to me to be a sentient universe. I am calling it a reflective organism to note the evolution of the theory as well. It shows us enough to change the perspective we have about the complex systems and how they can be reasoned about.

Thinking within the framework of the model and applying it to basically everything is revealing to me a new view of the world to the point I think the model holds. Even if it might end up just as a proto-theory in the beginning. It feels so universal. And exactly because it is the theoretical framework, it needs to be applied to experiments and simulations, maybe even leading to the universal construct in the end. But that's where the fun begins.

There are many topics that have been explored as forked paths that have shaped this theory, however, not all are included yet (especially implications on society, computation). As someone with computer science background, I've explored mathematics, physics, chemistry, biology and social sciences from enthusiastic point of view. seeking the answer to the questions like *"How to make code feel alive?"* or *"What makes a group of something an entity by itself?"* or *"What are the conditions for a cooperative and prosperous society of individuals?"* . The knowledge that I gained over the years was just good enough to know how to fit together what I wanted to express in terms of "living complex organisms", however, I had to rely on LLMs to fill in the gaps. I knew what I wanted to express and what the components are and how they fit together, but for deriving details, I had to use various LLMs, seeking the one that would understand the instructions and produce the answer I was seeking, which lead to usually experimenting and stitching together various "variations" of the answers. A paper that feels multidisciplinary like the one presented here would require a decent size team of people from various background all having the same direction and intent of what they want to express. And for someone writing this alone, LLMs were of great help. Rewriting the paper from scratch to use my own words only would also deny the process by which it was created, and would probably lose important information that is stored in the emergent structure that it holds now.

The mathematical background follows standard textbook treatments. The ideas presented draw on a wide range of sources. See the bibliography for the key works that shaped my view of the world, like the physics series by Susskind, all of Wolfram's books, Kauffman's Investigations for its style of approaching the unknown, Deutsch's The Beginning of Infinity for its creative optimism, Wilson's books on ants and nature, Lane's books on energy in organisms, to name just a few.

Consider the paper as a seed and I invite anyone to participate and correct me wherever I made a loose assumption, or where more expressive formalism exists to say the same thing. Collectively making it better and more elegant, and more fit to be used in wide scope of applications in the future.

What my intentions with the paper are is to present a view of the world which to me fits better to what I observed lately living at the farm and dealing with large enough ecosystems. And which finally starts to provide the answers I have been seeking for a long time.

## How it all started? Ant colony as a superorganism

Although there already are theories about superorganisms, swarm intelligence and distributed agents, some taken from biology and some from computer science, they all have their own core perspective and principles. **Biology** explores emergent intelligence from simple individuals, mostly through stigmergy and collective decision-making. **Swarm Intelligence** abstracts these principles into *algorithms* and *robotics*, treating the swarm as a **distributed optimization machine**. **Distributed Cooperative Agents** look at things in more general sense, designed systems where agents can be arbitrarily complex, communicate explicitly, and learn individually or collectively, but often inspired by swarm/biological models.

On the contrary to this fragmented view, the goal was to show that there might be unifying view underlying them all rather than fragmented one that exists today. And the way to explore this was to look at a superorganism as a nested structure of organisms, where organisms on one level play the role of organs of their superorganism on the higher level. However the hierarchical structure was not enough. There has to be something more than mechanical interaction, something like a drive and reflection. We'll get to this requirement for something that shows "life force" later, for now let's just focus on structure and function.

And to start with, as always, we look at ant colonies. The mapping between the colony and the organism with organs will serve as an abstraction point from which the model for the superorganism is created with the goal to create a unifying language for all types of superorganisms, not just ants.

### A biological metaphor at the level of physiology:

- The **colony = an organism**
- **Castes (workers, soldiers, queen, etc.) = organs or tissues**
- **Food flows = circulatory system**
- **Pheromone networks = nervous system**

This analogy has actually been explored in *superorganism theory*, though usually in a more general way.

In this way, the queen is the reproductive system, workers are metabolic organs, soldiers are the immune system, brood are the future cells of the body, foragers being the sensory cells, the skin, the ears, and nurses, etc. are tissues performing specialized supportive tasks. Trails of food are like veins and arteries. Pheromone trails are the nervous system with fast and distributed signalling that coordinates the "body". Homeostasis works through  $CO_2$  levels, food reserves, humidity control, temperature control, etc. And what is important here is that the colony has a lifecycle, literally organismal ontogeny. A colony has a birth (founding queen), growth (expansion of workers/organs), maturity (optimized functioning), aging (decline in forager efficiency, queen senescence), and death (colony collapse).

Later on we'll see that the notion of an ontogeny, the life history, is important part to the definition of the model.

Looking at the colony from a perspective of an organism allows us to start looking at it from the systems perspective, starting to think about the colony not just from "ants as agents" but more like **ants as cells, colony as body**.

## Superorganism as a fractal

Let's take this even further. Let's assume that superorganisms are nested organisms across many levels. Not only are ant colonies such organisms, but also everything else up to ecosystems and biosphere. In other words, what we're looking for are signs of scale invariance.

However, to call ecosystems and biospheres a superorganism, we have to resolve the view of ecosystems and biosphere as complex adaptive systems but not as individuals, in other words, we have to address:

- **Reproductive unit problem.** Organisms and superorganisms have a *single channel of reproduction*, like germline (organisms), queen(s) (colonies). Ecosystems have no equivalent. Forests, coral reefs, or oceans don't reproduce as wholes. They persist and change, but they don't make new copies of themselves in a Darwinian sense. Without reproduction, you can't have **evolutionary individuality** (the "unit of selection").
- **Alignment of interests.** In organisms, all cells share the same genome. In superorganisms, all ants are kin (very high relatedness). In ecosystems, organisms are only loosely connected. Predators, prey, and parasites have conflicting evolutionary interests. This prevents us to see ecosystems from functioning as tightly integrated, cohesive individuals.
- **Integration strength.** Superorganisms regulate homeostasis (temperature, food, defense) as a whole. Ecosystems have feedback loops (nutrient cycles, predator-prey dynamics) but **much looser coupling**. If a keystone species is removed, the whole system may collapse, but that fragility shows it's not as integrated as a body.

In order to resolve this, let's look at another example. What would it mean to look at the forest as a superorganism?

## Forest as a superorganism

- **Tree as an organism.** A single tree is classically treated as one organism: roots, trunk, leaves are organs; xylem and phloem are circulation; chemical signaling and mycorrhizal interactions as information channels. But trees also have modularity; branches and roots can sometimes survive independently or be grafted. This makes them "semi-colonial" in structure, more like coral than like us.
- **Forest as a superorganism.** (Many ecologists, e.g. Suzanne Simard's work on "Mother Trees", have argued that **forests behave as superorganisms**).
  - **Mycorrhizal networks ("Wood Wide Web")**, fungi connect tree roots, transferring carbon, water, and chemical signals.

- \* Analogous to circulatory + nervous systems spanning the whole forest.
- **Division of roles:**
  - \* “Mother trees” as resource hubs (analogous to nutrient-producing organs).
  - \* Shade-tolerant vs. pioneer species as different functional “tissues.”
- **Collective homeostasis:**
  - \* Forests regulate microclimate, water cycling, and soil nutrients.
  - \* Like a giant body stabilizing internal conditions.
- **Collective memory:**
  - \* Long-lived trees + soil seed banks preserve forest resilience over centuries.

So from a **systems ecology** perspective, a forest can be seen as an emergent superorganism. However, just like with ecosystems, the hesitation comes from the following problems:

- **Reproductive unit problem:** a forest does not reproduce as a whole — individual trees reproduce. There’s no “germline” for the forest.
- **Alignment of interests:** trees can compete as much as they cooperate (for light, water, nutrients). Ant colonies mostly avoid this via kin selection (high genetic relatedness). Forests are genetically diverse.
- So, while forests show **organism-like integration**, they lack the *evolutionary individuality* that ants or termites have.

Some interesting cases can be found for example in **clonal forests (e.g., Pando in Utah)** where a giant aspen grove is genetically one organism, sharing a root system. Here the *forest literally is a single organism* — ~47,000 stems (trees) all genetically identical, fed by the same roots. Another one are **corals & fungi** where similar modular “superorganismal” systems are present as the colony is both individual and ecosystem.

We’ll deal with alignment of interests and organism-like integration later, for now, let’s focus on reproduction unit problem.

## Reproduction unit problem

The main problem to scale the idea of the superorganism hides in the reproduction unit problem, as in, making copies of the organism itself, replicating completely. Which for the forest or the ecosystem doesn’t seem to fit. However, what if we look at them from the perspective of growing units? What if they do reproduce, we just don’t recognize the “fetus” stage of them until they are fully developed? What happens if we don’t simply look for seeking organisms that are making discrete copies of themselves but instead view these organisms as growing, self-propagating units and thus reframing the reproduction as propagation of self-organizing units.

### Forests as Growing, Reproducing Units

- **Expansion as reproduction** A forest expands outward as seeds establish on its periphery. The “fetus stage” of a new forest is just a few pioneer

trees + microbes + fungi, but we don't call it a "forest embryo." After enough growth, we retroactively label it a "new forest."

- **Fission-like processes** Large forests sometimes fragment, and fragments grow into separate stable ecosystems (like a colony splitting). That's analogous to **asexual reproduction** in biology.
- **Succession as ontogeny** Bare ground → grasses → shrubs → pioneer trees → mature forest. This is the **developmental trajectory** (ontogeny) of a forest superorganism. Each stage is like embryonic, juvenile, mature phases of an organism.
- **Clonal analogy** In *Pando* (the giant clonal aspen forest), the forest really is one genetic organism. But even diverse forests replicate "functionally" as new patches inherit structure (soil, fungi, canopy layers) from existing ones.

**Ecosystems as "reproducing":** Coral reefs seed new reefs through larval dispersal. Grasslands spread via rhizomes and soil community inheritance. Rivers create new floodplain ecosystems that inherit nutrient cycles. Earth itself has long been argued (Gaia hypothesis) to maintain and "self-propagate" conditions for life, even across mass extinctions.

From this lens, **the reproductive act isn't always making a perfect copy — it's about propagating the *process*** that sustains the unit.

The issue with the reproduction unit might be due to our bias of how we look at these systems and why we take making a discrete copies as what reproduction of an organism is all about:

- **Human timescale bias:** forests take centuries to "reproduce," so we don't see the cycle clearly.
- **Boundary bias:** we like discrete bodies (organism, ant colony), but ecosystems are fuzzy-edged.
- **Genetic bias:** we anchor individuality to shared genomes, but ecosystems are multi-genomic consortia.

If we expand our lens from *replication of a thing* to *propagation of a process*, then forests and ecosystems can be seen as **superorganisms that reproduce. And here lies the most important part that will shape the superorganism model - namely that organisms are propagating processes!**

We can now see the forest or the colony from the perspective of a (simplified) superorganism life cycle, treating them as growing, reproducing units, showing ontogeny+reproduction at the superorganism level, just at very different timescales and with different organs. Lifecycle stages of a forest:

- **Embryo** → bare ground colonized by microbes, fungi, and pioneer plants.
- **Juvenile** → shrubs and early-succession trees take hold.
- **Adult** → mature forest with canopy layers and full soil biota.
- **Reproduction** → expansion into new patches, creating the "embryos" of future forests.

In a similar way, **Ant colonies** grow from a founding queen ("embryo") → worker expansion ("juvenile") → mature caste structure ("adult") → swarming into new colonies ("reproduction").



Now let's take this perspective and look at the ecosystems.

## Ecosystem as Superorganism Life Cycle

1. **Embryo Stage** – Colonization
  - Bare substrate colonized by microbes, fungi, lichens, algae.
  - Examples:
    - Forest: bare soil colonized by pioneer plants.
    - Reef: bare rock colonized by coralline algae.
    - Grassland: post-fire regrowth from seeds & rhizomes.
    - River floodplain: microbial mats after flooding.
2. **Juvenile Stage** – Early Succession
  - Pioneer species establish, stabilize substrate, enrich soil/nutrients.
  - High turnover, opportunistic species dominate.
  - Example: shrubs in forests, soft corals on reefs, annual grasses after fire.
3. **Adult Stage** – Maturity / Climax Community
  - Stable, diverse, multi-layered ecosystem structure emerges.
  - Efficient nutrient cycles, strong symbiotic networks (e.g., mycorrhizae, coral-algae symbiosis).
  - Homeostasis: regulates microclimate, nutrient balance, resilience.
4. **Reproductive Stage** – Expansion / Propagation
  - Ecosystem “reproduces” by **seeding new patches** or colonizing disturbed ground.
  - Examples:
    - Forest spreads seeds & fungi → new groves.
    - Reef releases larvae → new reef patches.
    - Grasslands spread rhizomes → new meadows.
    - Rivers create new oxbows & floodplains.
5. **Senescence & Death**
  - Disturbances (fire, storms, disease, collapse of keystone species) may cause ecosystem death.
  - But death often seeds new “embryos”:
    - Burned forest → regrowth.
    - Collapsed reef → rubble colonized again.
    - Grassland → succession reset.

**Organisms, colonies, ecosystems all share a life cycle: Embryo → Growth → Maturity → Reproduction → Death.** The difference is **scale, integration, and clarity of boundaries**, where cells/“normal” organisms have crisp boundaries, fast cycles, colonies take decades and have caste integration, while ecosystems take centuries/millennia, have fuzzy boundaries, but same structural pattern.

This lets us reframe ecosystems not as static “collections” but as **growing, reproducing bodies of processes**. They *do* have fetus stages — we just overlook them until they look “big enough.”

## The case for the superorganism view

- **Functional Integration:** Ant colonies, forests, reefs, etc. exhibit **organism-like division of labor**: some agents forage, some defend, some reproduce, some regulate. These roles are not reducible to individuals, they only make sense as *organs of a larger body*
- **Shared Physiology:** Colonies circulate food (trophallaxis), regulate temperature, and resist pathogens like immune systems. Forests share carbon/nutrients through mycorrhizal networks — like circulatory systems. These are **literal physiological processes**, not metaphors.
- **Collective Cognition:** Colonies and ecosystems make **decisions** (quorum sensing, nest choice, resource allocation). No individual ant/tree “knows” the solution, but the whole colony/forest does. That’s cognition at the superorganism level.
- **Life Cycle:** Colonies: queen founding → growth → maturity → swarming → death. Forests: succession → climax → expansion → collapse. Both follow an **ontogeny** like organisms. This is more than “emergence”; it’s a structured developmental program.
- **Evolutionary Individuality:** Ant colonies reproduce as colonies, not as individual ants. Clonal forests reproduce as forests. Selection acts on the **whole unit**, which is the hallmark of individuality in evolutionary theory.
- **Limits of Emergence View:** Saying “it’s just individuals interacting” misses the fact that **emergence produces new levels of individuality**. Cells → organisms → colonies → ecosystems → biosphere. At each stage, the whole has properties (metabolism, cognition, reproduction) that individuals alone *cannot* account for.

### Why the superorganisms feels more natural than the emergence view?

- When you watch ants, bees, or a forest, what stands out isn’t *individual cleverness*, but the **integration** — flows of food, information, protection, reproduction. It *behaves like a body*.
- The “emergence of individuals” view is technically correct, but it feels like it stops halfway; yes, the whole is built from individuals, but the whole has **new properties and functions** that are irreducible to individuals.
- In nature, evolution has repeatedly produced these higher-level individuals: cells → multicellular organisms → colonies. So calling them “just emergence” underplays what’s happening.

On the other hand, the **emergence view** keeps things rigorously tied to the measurable interactions of individuals. The goal is to show that we can achieve the same with treating such collections as superorganism as well.

In the end, the case for superorganism is in observation of what already exists. *A superorganism is more than the sum of its parts: it eats, thinks, grows, reproduces, and dies at the level of the colony or forest, not the individual ant or tree. We see circulatory flows, nervous signalling, immune defence, and life cycles that mirror organismal biology. Reducing this to ‘just individuals’ misses the central evolutionary fact: natural selection acts on these wholes. If you accept that a body is not just a bag of cells, then a colony is not just a bag of ants — it is a living superorganism.”*

## Superorganism as “I”, a poetic tale

Let’s move from assumptions and hypothesis to being a little bit more poetic, and we’ll make another leap in perception of organisms here before jumping into formal definitions. If colonies or forests are organisms, what it would feel like to be them? What would a day in a life of a colony feel like (from the perspective of a colony)?

**Colony** *The sun warms my skin — walls of soil and stone. Deep in my gut, I pulse steadily with new eggs, life flowing outward like blood. I stream through my tunnels, veins carrying energy to and from the world. At the forest edge, I extend my senses, tasting with antennae. When I find food, I whisper chemicals, a nerve impulse racing down my pathways. I respond, surging, feeding, drawing energy back to my core. At my borders, I stand guard, immune to threats. Inside, I grow new tissue, larvae that will carry me forward. By nightfall, I have eaten, defended, grown. No single cell knows what I know. I breathe, I feel, I think. I am alive.*

**Forest** *Mist beads across my canopy, lungs of countless leaves. My roots drink water, passing it through fungal capillaries. One limb hungers, another warns of attack; impulses travel my underground nerves, and I answer. I cool the air above, regulate the waters below. I shelter a microbiome of lives within me. My memory stretches in my rings, centuries of growth carried in my body. My children rise in my clearings where light returns to soil. By night, I breathe in rhythm with earth. I rest, I remember, I endure. I am alive.*

But the interesting part starts when you consider that these two organisms can exist one next to other, emphasizing symbiosis as co-perception. Each “I” experiences its own life through the presence of the other.

**The Colony in the Forest / The Forest with the Colony** *I move across the ground, my foragers flowing outward, streams of bodies tracing paths through light and shadow. I taste the air for sugars, the soil for hidden trails. Each step expands me, a thousand feet carrying me into the green vastness. I feel them as they enter me — tiny pulses moving through my soil, brushing past my roots. Their trails lay signals across my skin, faint but insistent, like new nerves being drawn through my body. I climb stems and trunks, finding nectar in flowers, sap in wounds, soft tissues to harvest. I stretch into your canopy, my limbs delicate but relentless. I mark my paths so I may return, laying invisible threads through your undergrowth. I sense those threads — pheromone traces, weaving into my own channels of scent and sap. I answer with sugars at my tips, with resin to seal my wounds. I shift growth toward sun and soil, and in that shifting, I shape where the ants may run. I return laden with food, feeding my inner gut, the queen deep inside me. Yet what I carry is your gift — your sap, your seeds, your honeydew. In carrying it, I carry you. I release seeds in their mandibles, letting them spread me farther than wind alone could reach. I let them prune my pests, drive away the mouths that chew too much of me. In carrying them, I carry myself. I am the colony, coursing through your veins of soil and bark. I am the forest, coursing through your trails and gut. Together, I move. Together, I endure.*

However, even superorganisms can be organs of a larger organism, where the

distinction between the colony and the forest dissolves, speaking as “I” that contains both.

### ***One Voice***

*I move across myself,  
roots and trails entwined,  
veins of soil, veins of bodies,  
nutrients flowing both ways.*

*I taste light through leaves,  
I taste sugar through mandibles,  
every sweetness returning to my gut,  
whether it is brood or seed.*

*I defend with soldiers,  
I defend with thorns and resin,  
immune pulses rising,  
no mouth consumes me unchecked.*

*I remember in rings,  
I remember in pheromone traces,  
centuries and seconds layered,  
all memory mine.*

*I breathe in canopy and tunnel,  
lungs of leaves, lungs of soil,  
air moving through me,  
life moving through me.*

*I grow in saplings,  
I grow in larvae,  
new shoots, new workers,  
all my future.*

*I scatter myself as seeds,  
I scatter myself as swarms,  
always expanding,  
always beginning again.*

*I am many,  
but I am one.  
I am forest, I am colony,  
I am alive.*

## **Towards the superorganism model**

We have now set the stage to start building models of superorganisms, and then abstract these specific models to get a more general one. Note that we’re still in the “changing the perception of what an organism is” stage and we’re more interested in the concepts that the model requires rather than focusing on strict and precise formulation. For now at least.

The models serve as a conceptual scaffold, describing just enough of the dynamics so that we can get the feeling of how to describe such dynamic systems as superorganisms. They serve more as a tool for exploring how these systems behave in terms of superorganisms.

## Colony as a superorganism

Ant colony is usually modelled through discrete elements, like agents and pheromones, and the dynamics between them, such as their relation causing discrete state changes, however, we're going to approach it from a different angle, for reason which will reveal itself later throughout the description of the model.

Let's consider a colony as continuum model, as a set of flows, spatial transport + sensing (chemotaxis) and state switching. The dynamics depends on the state of the colony, that is, the budget/reserves of energy it has. The foraging is all about material (food) flow, while how the transport happens is about pheromones. This all drives population as well.

Let the spatial habitat be a bounded domain  $\Omega \subset \mathbb{R}^2$  with boundary  $\partial\Omega$ . Time  $t \geq 0$ . One way to describe the colony is through the following properties:

- $s(x, t)$ ,  $r(x, t)$  — surface densities of **searching** and **returning** ants.
- $p(x, t)$  — **pheromone** concentration (short-term field, “nervous” substrate).
- $F(x, t)$  — **food** density in the environment.
- $E(t)$  — **nest energy**/resource stock (metabolic reserve).
- $\mathcal{N} \subset \Omega$  — nest region;  $\mathbf{n}(t) = \int_{\Omega} (s + r) dx$  — colony size.

Let's now take a look at the **dynamics (continuum chemotaxis/transport with deposition & depletion)**. Let  $D_s, D_r, D_p > 0$  be diffusion constants;  $\chi_p, \chi_f \geq 0$  chemotactic sensitivities;  $\lambda_p > 0$  evaporation;  $\gamma > 0$  food pickup;  $q_s, q_r \geq 0$  deposition rates;  $u_{\text{nest}}$  net production of new workers. We can describe the activity as follows:

$$\begin{aligned}
\partial_t s &= \nabla \cdot \left( D_s \nabla s - \chi_p s \nabla p - \chi_f s \nabla F \right) - \gamma s F + \underbrace{\kappa_s n_{\text{nest}} \mathbf{1}_{\mathcal{N}}}_{\text{leave nest}} + \underbrace{\rho r \mathbf{1}_{\mathcal{N}}}_{\text{drop food \& rejoin search}}, \\
\partial_t r &= \nabla \cdot \left( D_r \nabla r - \chi_p^{(r)} r \nabla p - \alpha_r r \nabla \psi \right) + \underbrace{\gamma s F}_{\text{pickup switches search} \rightarrow \text{return}} - \rho r \mathbf{1}_{\mathcal{N}}, \\
\partial_t p &= D_p \Delta p - \lambda_p p + q_s s + q_r r, \\
\partial_t F &= D_F \Delta F - \gamma s F, \\
\dot{E} &= \eta \int_{\Omega} \gamma s F dx - \mu \mathbf{n}(t), \\
\dot{n}_{\text{nest}} &= -\kappa_s \int_{\mathcal{N}} n_{\text{nest}} dx + \rho \int_{\mathcal{N}} r dx + u_{\text{nest}}(E).
\end{aligned}$$

with reflecting boundary on  $\partial\Omega$  for  $s, r, p, F$  (no-flux). Let's look at the **I/O (sensing & actuation)**:

- **Inputs**  $I(t)$  as external odor/food field  $F$  (taste/smell), hazard field  $H(x, t)$  (touch/pain), temperature  $T(x, t)$ . Sensing operator  $\Sigma_{\text{col}} : (F, H, T) \mapsto (\nabla F, \nabla p, \nabla H, T)$ , biasing motion via the advection terms above.
- **Outputs**  $O(t)$  as **pheromone writing**  $p$  (neural signaling), **food transport** flux to nest  $J_{\text{food}}(t) = \int_{\Omega} \gamma s F dx$ , and **excavation/construction** (optional extra state  $C(x, t)$  with  $\dot{C} = \xi r - \delta C$ ).

In the end let's look propose **reproduction (swarming)** mechanism. Define colony **fitness/viability** as  $\mathcal{V}(t) = E(t) - \alpha \text{latency}(t) - \beta \text{risk}(t)$ . When  $\mathcal{V} > \theta$  and  $\mathbf{n}(t) > \bar{n}$ , launch  $k$  **daughter nests** at locations  $y \in \Omega$  drawn from an **argmax of expected intake**  $y^* \in \arg \max_{y \in \Omega} \int_0^T \int_{\Omega} K(x - y) F(x, t) dx dt$ , with dispersal kernel  $K$ . Each daughter starts with initial  $s_0, r_0, p_0$  inherited from parental parameters ("genetic/cultural memory").

To observe the colony from this perspective doesn't require a lot of imagination. Let's look if we apply the same logic to something else.

## Forest as a superorganism

Let's try to apply the same logic of flows to the forest. We have a spatial forest with multiple tree guilds and a belowground network (mycorrhizal "wood-wide web") connecting trees. Again, we have material flow as water, nutrients, and carbon move through soil, plants, and the fungal network. Information flow with fast airborne VOC field and slower network signals coordinate defenses and allocation. Not unlike pheromone trails, we deka with structure, where mycelium (and biomass) reshape the very conduits (patch connectivity).

Let's look at the definitions. Domain  $\Omega$  as above; trees (or guilds)  $i = 1, \dots, m$ . A **mycorrhizal network**  $G = (V, E)$  connects tree root systems;  $L_G$  its graph Laplacian;  $w_{uv}$  edge conductances. Forest properties can be defined by:

- $b_i(x, t)$  — biomass density of guild  $i$  (leaves/wood).
- $W(x, t)$ ,  $N(x, t)$  — soil **water** and **nutrient** fields.
- $v(x, t)$  — airborne **VOC** signaling field (short-term "nervous" substrate aboveground).
- $M(x, t)$  — belowground **mycelial density** (conductive substrate).
- Node stocks  $C_u(t)$ ,  $\sigma_u(t)$  — carbon and signal at tree  $u \in V$  (for network transfer).

The core of the model is defined by the **dynamics (resource, growth, signaling)**. Let  $P_i(L, W, N)$  be photosynthesis;  $U_i(W, N)$  uptake;  $\phi_i$  allocation;  $m_i$  mortality;  $D$ . diffusions;  $\lambda_v$  VOC decay. The dynamics works as follows.

**Node ODEs:**

$$\begin{aligned}\dot{C}_u &= S_u(P_i, b_i) - R_u(C_u) - A_u(C_u) - G_u + \sum_v w_{uv}(C_v - C_u), \\ \dot{\sigma}_u &= s_u(H_u) - \lambda_{\sigma} \sigma_u + \sum_v w_{uv}(\sigma_v - \sigma_u),\end{aligned}$$

with  $G_u = \int \alpha_{g,i} \kappa_u(x) b_i dx$  (carbon used for growth in footprint).  $S_u$  is an **aggregation (sampling) operator** that turns a spatial field into the scalar

seen/owned by tree  $u$ . It is defined as  $S_u[f] = \int_{\Omega} \kappa_u(x) f(x) dx$ , where  $\kappa_u(x) \geq 0$  is a weighting kernel over  $u$ 's canopy/root footprint (often normalized so  $\int \kappa_u = 1$ ). In the carbon ODE it's applied to the local photosynthate production density  $P_i(L, W, N), b_i(x, t)$  (units: carbon  $\cdot$  area $^{-1}$  time $^{-1}$ ), so  $S_u(P_i b_i)$  has units carbon  $\cdot$  time $^{-1}$ — the photosynthesis influx into node  $u$ .  $R_u$  is the node's **respiration loss** (carbon cost) at tree  $u$ . It typically includes maintenance (and optionally growth) respiration, often modeled as a function of the carbon pool and temperature  $R_u(C_u, T) = r_0 C_u \theta^{T-T_{\text{ref}}}$  (Q(\*10)/Arrhenius-type modifier), or split as  $R_u = R_{m,u}(C_u, T) + R_{g,u}(G_u)$ . Units are carbon  $\cdot$  time $^{-1}$ , matching the other terms in  $\dot{C}_u$ .

**Biomass PDE (no physical diffusion):**

$$\partial_t b_i = \alpha_{g,i}, S_{u(i)}^{-1}(G_{u(i)}) - m_i b_i - h_i(b_i, H, v) - (\mathcal{K}_i * \text{fecundity}_i(b_i)),$$

with  $\mathcal{K}_i$  as the **seed-dispersal kernel** for guild  $i$  which tells you where offspring land (and possibly establish), while  $\text{fecundity}_i(b_i)$  tells you how many are produced.

**Soil & VOC:**

$$\begin{aligned} \partial_t W &= I_{\text{rain}} - \sum_i U_i^W(b_i, W) + \nabla \cdot (D_W \nabla W), \\ \partial_t N &= I_{\text{min}} + \text{decomp}(b) - \sum_i U_i^N(b_i, N) + \nabla \cdot (D_N \nabla N), \\ \partial_t v &= D_v \Delta v - \lambda_v v + \sum_i \epsilon_i \Psi_i(b_i, H). \end{aligned}$$

**Mycelium-edge coupling:**

$$\partial_t M = D_M \Delta M + \rho_M \sum_u \kappa_u C_u - \delta_M M, \quad w_{uv} = \bar{w}_{uv} \left[ \int_{\Gamma_{uv}} M, ds \right]^\beta,$$

states about how it spreads and how well the trees are connected.

**Seed dispersal / patch propagation**  $\partial_t b_i$  gets  $+ (\mathcal{K}_i * \text{fecundity}_i(b_i))$  with kernel  $\mathcal{K}_i(x) = \frac{1}{2\pi\sigma_i^2} e^{-\|x\|^2/2\sigma_i^2}$ .

Let's define now **I/O (sensing & actuation)** of a forest:

- **Inputs**  $I(t)$ :  $L(x, t)$  (light),  $W, N$  (touch/internal state of water/nutrients),  $H(x, t)$  (herbivory/pain),  $T(x, t)$  (temperature),  $F_{\text{fire}}$  (disturbance) with a sensing operator  $\Sigma_{\text{for}} : (L, W, N, H, T) \mapsto (P_i(L, W, N), H, \nabla W, \nabla N)$ .
- **Outputs**  $O(t)$ : (i) **VOC field**  $v$  (airborne signaling influencing herbivores, predators, ants) with **mycorrhizal transfers**  $\sum_v w_{uv}(C_v - C_u)$  (resource reallocation), **root exudates/litter** that modify  $N$ , **seed rain** via  $\mathcal{K}_i * \cdot$  (reproduction/expansion) and **canopy cooling** (microclimate actuation), modeled as  $Q_{\text{latent}}(x, t) = \alpha_{\text{ET}} b_{\text{leaf}}(x, t) f(W, T)$ , feeding back on  $T$ .

**Reproduction** only happens when the conditions are right. Hence, we need a **summary score** that tells, at time  $t$ , whether the forest is “doing well enough” to expand (or should contract/reset).

$$\mathcal{V}(t) = \underbrace{\int_{\Omega} \left[ \sum_i P_i(L, W, N), b_i - \sum_i m_i, b_i \right] dx}_{\text{net primary production (NPP) proxy}} - \beta \underbrace{fire_{risk}(t)}_{\text{disturbance penalty}} - \gamma \underbrace{water_{deficit}(t)}_{\text{stress penalty}}$$

$P_i b_i$  represents instantaneous **carbon gain rate** (photosynthesis per area  $\times$  biomass present) for guild  $i$ .  $m_i b_i$  os**background losses** (mortality/turnover) for guild  $i$ . The **integral over**  $\Omega$  sums gains–losses across space  $\rightarrow$  a landscape-scale net productivity rate.  $fire_{risk}(t)$  is a scalar proxy for how burn-prone the system is (higher = worse). An example: Fuel load  $\times$  dryness:

$$\int_{\Omega} \underbrace{B_{fine}(x, t)}_{\text{litter/grass}} \underbrace{[1 - \theta(W(x, t))]}_{\text{dryness}} dx. \text{ Or a precomputed fire-weather index averaged}$$

over  $\Omega$ . Scale with  $\beta > 0$  so units match the NPP term.  $water_{deficit}(t)$  as a scalar drought stress measure, e.g.  $\int_{\Omega} \max\{0, W^*(x) - W(x, t)\} dx$  or  $\int_{\Omega} (1 - f_W(W, T)) B_{leaf}(x, t) dx$ , and  $\gamma > 0$  weights how strongly deficit reduces viability.

When  $\mathcal{V} > \theta$  and a percolation/connectivity criterion holds (giant component of high-biomass cells), the forest **propagates** by seed kernels generating new persistent patches; when connectivity breaks under disturbance, senescence/reset occurs (succession restarts).

## Unification, a general model

Although the colony and the forest are very different in appearance and functions, when abstracted, they don’t seem that different. There are actors, which move on paths, and these paths are used to transfer signals between parts of the whole, composing the backbone structure of the whole. The actors sense the conditions within the whole as well as their surrounding and act as I/O. They both accumulate energy and the structure, the accumulations and the flows of signals and information define the state of the whole.

Let’s define a **Field-Coupled Superorganism** based on these observations that we learned from the two organisms and define it as a tuple:

$$\mathcal{S} = (\Omega, \mathcal{A}, \mathcal{G}, X(t), F(t), M(t), \Sigma, \Pi, \Upsilon, \mathcal{R}),$$

where:

- $\Omega$  is spatial domain.
- $\mathcal{A} = \{a_1, \dots, a_k\}$ , a set of **organ classes** (e.g., scouts/workers, tree guilds). Each has density/state  $x_j(\cdot, t)$ .
- $\mathcal{G}$  represents internal **networks** (trail graph, mycorrhizae) with Laplacian  $L_{\mathcal{G}}$ .
- $X(t)$  are concatenated **organ states** (PDE/ODE on  $\Omega$  and nodes  $V$ ).
- $F(t)$  are **fast shared fields** (“nervous” substrates; pheromone, VOC, electrical plant signals) governed by linear parabolic SPDEs  $\partial_t F = \mathcal{D}\Delta F - \Lambda F + S_F(X; I)$ , with diffusion  $\mathcal{D}$ , decay  $\Lambda$ , sources  $S_F$ .



- $M(t)$  is **slow memory/morphology** (architecture, nest/cavity structure, long-term archive parameters) with drift  $\partial_t M = \Phi(X, F, M)$ .
- $\Sigma$  is a **sensing operator** mapping environment  $I(t)$  and internal state to sensed signals  $S(t) = \Sigma(I, X, F)$ .
- $\Pi$  holds **policy/controller** producing actions  $U(t) = \Pi(S, M)$  (could be RL/optimal control/threshold rules).
- $\Upsilon$  is an **actuation operator** writing to environment and internal fields  $(X, F, M) \xrightarrow{\Upsilon(U)} (X', F', M')$  e.g., deposition, transfer, growth, emission.
- $\mathcal{R}$  is a **reproduction/senescence operator** that (i) spawns children  $\{\mathcal{S}_i\}$  by seeding initial states from  $M$  when a viability  $V(X, F, M; I)$  crosses a threshold; (ii) prunes/merges when viability drops.

**Coupled dynamics (compact form):**

$$\begin{aligned}\partial_t X &= \mathcal{F}(X, F, M; I) + \Upsilon_X(\Pi(\Sigma(I, X, F), M)), \\ \partial_t F &= \mathcal{D}\Delta F - \Lambda F + \Upsilon_F(\Pi(\Sigma(\cdot), M)), \\ \partial_t M &= \Phi(X, F, M), \\ O(t) &= \Gamma(X, F) \quad (\text{measured outputs to the outside world}).\end{aligned}$$

Here  $\Gamma$  selects observable actuations (transport fluxes, emissions, construction).

**I/O taxonomy (for any  $S$ ):**

- **Touch/pressure/pain:** local damage or force field  $D(x, t)$  sensed via  $\Sigma$  as  $D$  or  $\nabla D$ ; actuation  $O$  includes repair/defense flows.
- **Taste/smell:** scalar concentration fields  $C_k(x, t)$  sensed via  $C_k$  and  $\nabla C_k$ ; actuation includes emission fields, transport toward/away from gradients.
- **Proprioception/homeostasis:** internal stocks  $E, W, N, C$  compared to setpoints; actuation adjusts allocation, recruitment, aperture/ventilation, etc.

**“Oneness” (integration) metric:**

Let  $H$  be the interaction graph among organs (edges when coupled via  $F$  or  $L_G$ ). A simple structural index of integrated individuality is  $\mathcal{I}(S) = \lambda_2(L_H)$  (algebraic connectivity), and a dynamical one is **mutual information** between spatial partitions  $A, B \subset \Omega$ :  $\mathcal{I}_{\text{dyn}} = \frac{1}{T} \int_0^T I(X_A(t); X_B(t)) dt$ , both high when the system behaves as a single coordinated unit.

The following equations are a **minimal** realisation of the abstract super-organism template. Note that for the colony, we focused on *dynamic behavioural states* that any ant can occupy at a given moment. In many ant species *any* worker can become a forager or a carrier depending on local need. Modelling them as two *states* captures this flexibility without having to track a fixed caste hierarchy. In the continuum description they are the variables that actually drive the fast pheromone field  $p$  and the material flux of food.

- **Colony:**  $\mathcal{A} = \{\text{search, return}\}$ ;  $F = p$  (pheromone);  $\mathcal{G}$  is optional (explicit trail network once stabilized);  $\Pi$  implements quorum & recruitment;  $O = (p \text{ field}, J_{\text{food}}, C)$ .
- **Forest:**  $\mathcal{A} = \{\text{guilds}\}$ ;  $F = v$  (VOC) plus  $M$  (mycelium) split into slow (structure) and fast (conduction);  $\mathcal{G}$  is the mycorrhizal graph;  $\Pi$  allocates carbon/defense;  $O = (v, \text{transfers, seed rain}, Q_{\text{latent}})$ .

We’re now able to describe completely different systems with the same model.

## What is this model all about?

Let’s recap what we’ve done here. The Superorganism model is a generic, mechanistic framework for describing any biological or bio-inspired system in which:

- Multiple functional units (“organs”) coexist and interact (workers/queens, plant tissues, fungal hyphae, etc.).
- These organs exchange information through fast, spatially distributed fields (pheromones, chemicals, electric potentials) that obey diffusion-decay dynamics.
- They are also linked by an internal discrete network (trail graphs, mycorrhizal webs, vascular networks) whose topology influences how signals propagate.
- The whole system possesses a slow, memory-like component (nest architecture, tree branching pattern, caste history) that changes only on longer time scales and stores past experience or structural adaptation.
- A sensing  $\rightarrow$  decision  $\rightarrow$  actuation loop governs the behaviour: external stimuli are sensed, processed by a policy (hand-crafted rule, RL controller, optimal-control law), and translated into actions that modify the organs, fields, and memory.
- The system can reproduce or prune itself when viability crosses thresholds.

In essence, it is an abstract dynamical system, a set of coupled PDEs/ODEs plus a graph Laplacian, that captures how collective entities behave as unified wholes while being composed of many interacting parts. In other words, the core idea is that every superorganism has three fundamental layers:

1. **Organs (agents, tissues)** — the basic active units.
2. **Fields (signals)** — the fast flows of information that let parts coordinate.
3. **Memory (morphology)** — the slow, long-term structure that holds the system together.

And together, these layers let the system **sense, decide, act, and reproduce**.

Let’s dive into explanations and examples of the components of the model:

**Spatial Domain ( $\Omega$ )** defines the **arena** where agents interact and signals flow. Think of this as the “body” or “world” in which the superorganism lives. For ants, it’s the terrain of soil, nest, and food patches. For a forest, it’s the landscape of soil, roots, canopy, and air. For AI, it might be the network environment or problem space.

**Organs ( $\mathcal{A}$ )** divide labor so the whole system can do more than individuals. These are groups of individuals that perform specialized roles. In an ant colony: scouts, foragers, soldiers, nurses. In a forest: shade-tolerant trees, pioneer species, mycorrhizal fungi. In AI: scouts (search), workers (solve), couriers (route results).

**Networks ( $\mathcal{G}$ )** bind the parts together, carrying resources and signals. A superorganism is stitched together by **invisible wiring**. Ants use pheromone trails (like a nervous system). Trees use fungal networks to share carbon (like blood vessels). AI swarms use communication graphs or shared databases.

**Fast Fields ( $F$ )** allow **quick, flexible coordination** — the “nervous system.” Fast fields are **short-lived signals**. They diffuse, decay, and can be rewritten constantly. Pheromone plumes for ants. VOCs (volatile organic compounds) released by trees when attacked. Shared scratchpads or ephemeral memory in AI swarms.

**Slow Memory ( $M$ )** provide **stability, inheritance, and learning** across time. Slow memory is **long-term structure and record**. It changes slowly but anchors the system’s identity. The nest architecture of ants. The soil and tree architecture of forests. The knowledge base or archive in AI.

**Sensing Operator ( $\Sigma$ )** translate the outside world into internal signals. The system has to perceive its world. Ants sense food, temperature, and pheromone gradients. Trees sense light, water, herbivore attack. AI swarms sense tasks, feedback, anomalies.

**Policy ( $\Pi$ )** is the **decision-making engine** of the superorganism. Once it senses, the superorganism needs to decide what to do. Ants recruit more foragers to food if signals are strong. Forests shift carbon to stressed trees or invest in seeds. AI swarms allocate roles or increase exploration.

**Actuation ( $\Upsilon$ )** turn decisions into real-world change. This is how the superorganism affects the world. Ants lay pheromone and carry food. Trees release seeds and VOCs, or change water cycling. AI swarms write to memory, emit artifacts, call APIs.

**Reproduction Operator ( $\mathcal{R}$ )** treats the **whole system as an individual that grows, reproduces, and dies**. Superorganisms also have life cycles. Ant colonies swarm to make new colonies. Forests spread seeds and regrow after fire. AI swarms may clone themselves into new instances.

This model doesn’t just describe ants or forests. It gives a **general blueprint for “living wholes” made of many agents** — whether biological, ecological, or artificial.

It is a **unified framework across domains**. Existing models of colonies (ant PDEs, agent-based) and forests (succession, resource dynamics) are usually siloed. Here we’ve given a **common dynamical template** (states  $X$ , fast fields  $F$ , slow memory  $M$ , networks  $\mathcal{G}$ , sensing  $\Sigma$ , policy  $\Pi$ , actuation  $\Upsilon$ , reproduction  $\mathcal{R}$ ). That means you can study an *ant colony*, a *forest*, a *coral reef*, or even a *swarm of AI agents* with **the same mathematics**.

**Oneness quantified.** Biologists often argue *qualitatively* about whether colonies/forests are superorganisms. This model proposes **metrics of individuality**, *structural integration* with algebraic connectivity  $\lambda_2(L_H)$  and *dynamic integration*: mutual information between subsystems. This gives a **numerical yardstick** for “how much of an individual” a system is. That’s rarely formalized.

**Explicit I/O interface.** Organisms have senses and actuators; ecosystems are rarely modeled this way. Here, we explicitly map **input fields** (light, food, damage, heat, etc.)  $\rightarrow$  **sensing operator**  $\rightarrow$  **policy**  $\rightarrow$  **outputs** (pheromones, VOCs, seeds, construction). That allows us to treat a colony or forest like

a black-box cybernetic system, a controllable, trainable agent, which bridges **ecology**  $\leftrightarrow$  **control theory**  $\leftrightarrow$  **AI agentics**.

**Reproduction operator  $\mathcal{R}$ .** Classical models of ecosystems don’t treat the **propagation of the whole system** (e.g., forests giving rise to new forests). Our framework formalizes **when the unit reproduces**, not just when individuals do. This is critical for treating colonies or forests as Darwinian individuals, which has been conceptually sticky. This gives us a **life-cycle operator** for superorganisms, not just population turnover.

**Bridging scales.** By separating **fast fields** (pheromone, VOCs) from **slow memory/morphology** (nests, soil networks), we capture both, rapid collective decisions (minutes to days) and long-term development and inheritance (years to centuries). This dual-scale decomposition is new in superorganism modeling. Most models pick *either* short-term behavior *or* long-term succession, not both.

**Application to AI & MAS.** In multi-agent AI, most systems use explicit communication or central coordination. This model shows how to design swarms with **fast stigmergic channels** (ephemeral traces), **slow archives/morphology** (long-term system memory) and **reproduction** (system spawns new sub-swarms). It imports *ecological individuality* into AI design — not just swarms that optimize, but swarms that **live and reproduce**.

Summarizing model’s applicative possibilities:

- The model defines how to **measure individuality**. *Is a forest a super-organism?* With  $\mathcal{I}_S, \mathcal{I}_{dyn}$ , we can **quantify** how “integrated” it is. This helps settle conceptual debates with data.
- Generalization allows for **cross-domain transfer**. Insights from colonies can be directly tested in forests, or vice versa, because both use the same formal machinery. For example, pheromone decay  $\leftrightarrow$  VOC decay, nest building  $\leftrightarrow$  canopy construction.
- **Management of ecosystems.** If you model a forest as an I/O system, you can design **interventions** (inputs like prescribed fire, water, or adding ants) to steer system states predictably. Moves ecosystem management toward **control theory**, not just descriptive ecology.
- **Designing agentic AI ecosystems.** The model treats multi-agent AI as a *living unit* that grows, regulates itself, and reproduces. This could lead to AI systems that are **resilient, evolvable, and safe**, unlike brittle one-shot agent architectures.
- **Evolutionary transitions in individuality** as it provides a framework to study how “collections become individuals”, a central question in evolutionary biology. Could illuminate not just ants & forests, but also **AI collectives** as a new kind of evolutionary transition.

Let’s recap what the model introduces:

- **Unification** of colonies and forests (and potentially any superorganism/ecosystem) under a single template.
- **Formal I/O architecture** — treating them like agentic systems with senses, memory, and actuators.
- **Quantitative individuality metrics** ( $\lambda_2$ , mutual information).
- **Explicit reproduction operator** for the *whole system*.

- **Dual timescale design** (fast fields + slow memory).
- Application not just to biology but to **agentic AI design**.

## Extending the general model with Assembly theory

So far, the goal was to simply find the underlying language that can describe a collection of autonomous units working together as a whole, each having their own role at what the system as a whole decides and at the same time being supported by the system.

As the current model grows through its life cycle, what we’re considering as growth at the moment is simply just counting. Growth is considered as producing more of things, more of components. The model predicts some dynamics how this “more” happens in terms of laws of the life cycle, but in essence, it is simply about counting more units. More ants, more biomass, more agents.

The life cycle at the core of what a superorganism is already hints that there should be more to mere growth through numbers. Note also that the model contains components of memory, where fast decisions are storing their actuation into slow storage. We already have history, what we miss might be proper direction of construction. Thus it feels natural to extend the model through the lens of assembly theory (AT).

Let’s quickly recap key ideas behind the AT. Assembly theory (from Lee Cronin and colleagues) is a framework to describe **how complex objects come into being**.

- Every object can be seen as a sequence of **assembly steps** (putting together smaller building blocks).
- The **assembly index (AI)** represents the minimum number of steps needed to construct the object.
- Objects with **high assembly index** reflect directed, purposeful processes (like life and technology), not just random chance.

What we need to add to the theory that would fit the general model we’re extending are **assembly pathways**, **assembly index** and **growth intent**. Assembly pathways can be seen as a sequence of construction events. Something like *Larva* → *Worker* → *Nest* → *Trail* → *Foraging System* or *Soil* → *Pioneer* → *Mycorrhizal Net* → *Canopy* → *Dispersal Network*. Assembly index (AI) is all about the depth of this pathway.

We are slightly simplifying the assembly pathways here for the sake of clarity. The order shown represents a coarse-grained view of how the building blocks at the organismal level, represented by the set  $\mathcal{A}$ , evolve through successive joining operations. Although we depict transitions such as *Larva* → *Worker* as single steps in the pathway, in reality this reflects the set of changes produced by the queen and larva through many iterations of coupled dynamics before the *Worker* state is reached. The number of distinct assembly steps required to produce such an element is captured in its assembly index (AI). The structural information of how it was built is retained in the memory field  $M$  as well as

in the organisms history network  $\mathcal{G}$ . These stored structures allow the reuse of previously established pathways as encoded, ordered joining operators. When the *Worker* state emerges from the *Larva* state can also be described within AT as the formation of a stable, reusable unit, a bottom-up construction that simultaneously satisfies the top-down requirements imposed by the definition of the parent organism.

Growth intent is more interesting. In biology, we consider intent as equivalent to fitness, pushing toward higher assembly pathways to improve survival. However, in the context of a superorganism the intent can be seen as goal functions that reward assembly steps leading to better integration and capability. While Assembly Theory itself does not posit an intrinsic optimization principle, we introduce a **growth directive**

$$\max_{\Pi} \mathbb{E}[AI(t+T) \mid AI(t)]$$

Growth now has direction towards higher-order structures. It is not conscious purpose, but directed assembly trajectory encoded in system dynamics. Colony doesn't just add numbers, it builds functional stages. Forests don't just add biomass, it builds layers (soil  $\rightarrow$  understory  $\rightarrow$  canopy  $\rightarrow$  seed dispersal). Assembly pathways constrain growth, hence the superorganism follows predictable life cycles. AT gives us a language to talk about purposeful growth in superorganisms.

Where does it fit in the general model? It's all about updates to the policy  $\Pi$  in such a way that it should act to increase the assembly index, ensuring the superorganism grows toward higher integration and capability.

We enrich the model with two new components:

1. **Assembly Pathway**  $\mathcal{P}(t)$ 
  - The sequence of construction steps leading from simple to complex forms.
  - Example (colony): Egg  $\rightarrow$  Worker  $\rightarrow$  Nest  $\rightarrow$  Trail  $\rightarrow$  Foraging Network
  - Example (forest): Soil microbes  $\rightarrow$  Pioneers  $\rightarrow$  Mycorrhizal Net  $\rightarrow$  Canopy  $\rightarrow$  Seed Rain
2. **Assembly Index**  $AI(t)$ 
  - The minimum number of non-trivial assembly steps required to build the current system configuration.
  - Low  $AI$ : small fragments, little integration.
  - High  $AI$ : complex, integrated system (superorganism fully expressed).

We now have the abstract model:

$$\mathcal{S}_{AT} = (\Omega, \mathcal{A}, \mathcal{G}, X(t), F(t), M(t), \Sigma, \Pi, \Upsilon, \mathcal{R}, \mathcal{P}(t), AI(t))$$

The extensions allow us to define growth law and intent principle. **Growth Law:**

$$\partial_t AI = G(X, F, M, \Pi; I) - L(D),$$

where  $G$  represents growth term, ie. assembly of new structures from existing ones, directed by policy  $\Pi$ .  $L$  is loss term. ie. disturbances  $D$ , senescence, fragmentation. And **Intent Principle** where policies  $\Pi$  are guided not only by survival but also by assembly depth maximization:

$$\Pi^* = \arg \max_{\Pi} \mathbb{E}[AI(t + \Delta t) \mid AI(t)]$$

This is the **assembly-driven intent**; the superorganism acts to climb its assembly pathway.

Growth law is defined slightly differently from what the assembly theory uses. Let's make a bridge between our ODE and AT's graph definition. In the model,  $\mathcal{G}$  represents internal network which also represents the order in which construction events have happened. We can treat  $\mathcal{G}$  as the *carrier* of all possible assembly pathways and compute the index from it:

$$AI(t) = \underbrace{\mathcal{I}[\mathcal{G}(t)]}_{\text{graph functional}} = \min_{\text{directed paths } p \subseteq \mathcal{G}(t)} |p|.$$

How the growth term  $G$  modifies the graph? The general model already contains an actuation operator  $\Upsilon$  that writes actions into the fast fields, the slow memory field and the network. A convenient way to make this explicit is:

$$\dot{\mathcal{G}} = \Gamma(X, F, M, \Pi; I),$$

where  $\Gamma$  adds or strengthens edges (or creates new nodes) according to the current policy  $\Pi$  (because  $\Pi$  is itself a function of AI, the actuation is *goal-directed*).

In concrete terms, an ant colony may either increase the conductance  $w_{uv}$  of a trail edge when many ants repeatedly traverse it (fast-field flux  $\rightarrow$  feedback current  $j^\mu \rightarrow$  memory update in  $M \rightarrow \Gamma$  raises  $w_{uv}$ ), or create a *new* edge between two previously unconnected nests once the policy predicts that doing so will add one more step to the assembly pathway (e.g. “nest  $\rightarrow$  new foraging tunnel”).

Because  $\Gamma$  is **explicitly** a function of  $(X, F, M, \Pi)$ , we can write the *growth term* in the AI ODE as the **chain rule**:

$$G = \frac{\partial \mathcal{I}}{\partial \mathcal{G}} : \dot{\mathcal{G}} = \frac{\partial \mathcal{I}}{\partial \mathcal{G}} : \Gamma(X, F, M, \Pi; I).$$

The colon “:” denotes the double-contraction of a tensor (the gradient of the functional) with the graph-update matrix. In words: *AI grows in proportion to how many new admissible assembly steps the policy actually creates.*

What about the loss term  $L$ ? Disturbances  $D$  (e.g. fire, predation, pathogen attack) are modelled as **edge-removal** or node-death events  $\dot{\mathcal{G}}_{\text{loss}} = -\Lambda(D)$ , with  $\Lambda$  a non-negative operator that deletes edges (or lowers conductances to zero). The corresponding contribution to the AI dynamics is

$$L = -\frac{\partial \mathcal{I}}{\partial \mathcal{G}} : \dot{\mathcal{G}}_{\text{loss}} = \frac{\partial \mathcal{I}}{\partial \mathcal{G}} : \Lambda(D).$$

Thus the **loss term** in the ODE is simply the *negative* effect of graph degradation on the assembly index.

Substituting the two in the original growth law gives a fully **graph-consistent** AI evolution:

$$\partial_t AI = \frac{\partial \mathcal{I}}{\partial \mathcal{G}} : \left[ \Gamma(X, F, M, \Pi; I) - \Lambda(D) \right].$$

Because the Assembly Index in the AT is *defined on the graph*, any change in the graph automatically changes AI. Conversely, the ODE tells us exactly how a desired change in AI must be implemented as an edge-addition/removal operation. **The growth term  $G$  is nothing but “add edges that increase the shortest admissible path length”.**

Let’s look at how this would work in the case of an ant colony. We write the growth term as a **graph-update operator**  $\Gamma$  (the rate at which the internal network  $\mathcal{G}$  gains nodes/edges) plus any *direct* production of new agents that do not require an edge (e.g. queen egg laying, seed release). Using the chain rule we have

$$G = \underbrace{\frac{\partial \mathcal{I}}{\partial \mathcal{G}} : \Gamma(X, F, M, \Pi)}_{\text{network-driven AI increase}} + \underbrace{g_{\text{prod}}(X, F, M, \Pi; I)}_{\text{direct agent production}}.$$

We need to define the graph operator (let’s call it  $\Gamma_{\text{ant}}$ ). The internal network  $\mathcal{G}$  for ants is the **trail graph** whose edge weight  $w_{uv}$  measures how strongly two locations are linked by a pheromone trail. A biologically plausible rule that turns fast pheromone flux into permanent trail reinforcement is

$$\dot{w}_{uv} = \underbrace{\alpha_P \int_{\Omega_{uv}} P(\mathbf{x}, t) d\mathbf{x}}_{\text{pheromone-driven reinforcement}} - \underbrace{\beta_w w_{uv}}_{\text{natural decay}}$$

The integral is over the spatial segment that connects node  $u$  to node  $v$ ; it captures how much pheromone has passed, while  $\alpha_P$  converts a *fast* flux into a *slow* increase of the memory field; the decay term models trail erosion.

If a **new foraging site** appears (detected by high food cue  $I_{\text{food}}$ ), the policy creates a brand-new edge:

$$w_{u,\text{new}} = \begin{cases} \gamma_f I_{\text{food}}(\mathbf{x}_{\text{new}}) & \text{if } P > P_{\text{th}}, \\ 0 & \text{otherwise} \end{cases}$$



Thus  $\Gamma_{\text{ant}}$  requires both. Ants increase their **individual count** independently of trails:

$$g_{\text{prod}}^{\text{ant}} = \underbrace{\eta_Q f_Q(I)}_{\text{queen egg-laying rate}} + \underbrace{\eta_W f_W(P, I)}_{\substack{\text{worker recruitment} \\ \text{(pheromone-stimulated)}}} - \underbrace{\delta_U U}_{\text{mortality}}.$$

$f_Q(I)$  grows with food and temperature and  $f_W$  rises when pheromone concentration exceeds a threshold (workers are recruited to reinforce successful trails). Putting everything together, the **ant-colony growth term** reads

$$G_{\text{ant}} = \frac{\partial \mathcal{I}}{\partial \mathcal{G}} : \Gamma_{\text{ant}}(X, F, M, \Pi) + g_{\text{prod}}^{\text{ant}}(X, F, M, \Pi; I).$$

Because  $\Gamma_{\text{ant}}$  adds edges (or strengthens them) that increase the shortest directed assembly pathway in  $\mathcal{G}$ , the first term directly raises the Assembly Index, exactly what the Intent Principle demands.

We now have **directedness in growth**. *The general model treats growth as a passive counting process*: more ants, more biomass. By inserting an *intent* that rewards increasing *AI*, we endow the system with a *goal-driven trajectory*. The policy  $\Pi$  is no longer merely “survive” or “explore”; it now has to choose actions that are *valuable in terms of assembly depth*. We introduce **explicit representation of life-cycle stages**. Life cycles (colony founding, forest succession, etc.) become *explicit variables*. They can be compared, aligned, and even transferred across systems. For instance, a tree model can inherit the “soil  $\rightarrow$  pioneer  $\rightarrow$  canopy” pathway from a plant-based template, encoded in the seed. We can express **quantifiable purpose without consciousness**. Assembly theory tells us that high *AI* is a signature of purposeful construction (life, technology). By maximizing expected future *AI* we are effectively *encoding a fitness function* that prefers structured, integrated outcomes – the same thing natural selection would do, but expressed mathematically rather than through ad-hoc survival criteria. **Coupling structure and dynamics** by the growth law as *AI* ties the *fast* state variables ( $X, F$ ) and the *policy* directly to a slow, structural metric. This is a step beyond the usual “state  $\rightarrow$  action  $\rightarrow$  next state” loop: we now have a *meta-loop* where the outcome of actions feeds back into the very notion of what counts as “growth”. **Predictive constraints**, because  $P(t)$  imposes an ordering on feasible assembly events, the model automatically generates *predictions about which structures can appear next*. In a forest model this would say “after pioneer shrubs the next viable step is mycorrhizal networking”, not just “random growth”. A scalar *AI* gives us potentially universal yardstick, allowing **comparability across domains**. Two very different superorganisms (a fungal mat vs an ant colony) can now be compared by their *assembly indices* and *pathways*. However, only if we first coarse-grain each system onto a common graph representation and, when needed, weight or normalise the edges so that one “step” has the same *effective cost* in every domain. This gives a potential for common language to talk about, say, the “complexity” of a honeycomb versus that of a root-hive system.

Adding AT, we have turned our general model from *mechanistic description of numbers and fields* into a *generative theory of purposeful construction*.

## Introducing the “I”, experience through reflection

Let’s make another extension to the model. So far, the model has intent, but it is completely mechanical. The model already predicts a feedback loop where slow memory is read after the actuation process and the policy that defines the decision the organism makes. Let’s explore a bit more what this feedback loop could mean. What if we would consider another purpose behind it all, which we’d call experience. Experience is based on a feedback loop that evaluates what has been created or disintegrated. This actuation into slow memory can be seen as the expression of the organism, self-realization, and self-awareness of what has been created, done. Let’s extend the model so that this experience of self-realization becomes part of the intent and the driving force of the organism. Thus what we’re trying to do is to create a model where experience and self-awareness become not side-effects, but core drivers of purpose.

In order to include these ideas, we need to add a new axis to the model. It’s no longer just about building structures but also experiencing the structure through feedback, where experience is defined as fast-slow loop closed inward. In other words, the organism senses what it has built/dissolved and reflects back into its self-model. We’re going to reuse the idea from the assembly part and analogously add an **experience index**,  $\mathcal{E}(t)$ . The new model now becomes:

$$\mathcal{S}_{AT} = (\Omega, \mathcal{A}, \mathcal{G}, X(t), F(t), M(t), \Sigma, \Pi, \Upsilon, \mathcal{R}, \mathcal{P}(t), AI(t), \mathcal{E}(t))$$

Just as the assembly index provides a quantitative measure of its structural complexity, so does the experience index  $\mathcal{E}(t)$  quantifies the *feedback richness* between fast and slow layers, as in, how much of what is written into slow memory is also *sensed and looped back* into decision-making.

We can observe this already. In humans as sensory reflection, emotional states and even self-narratives. In ecosystems we can observe resilience feedback like how forest fire scars change species assembly. Or even something we take as ordinary; a *nutrient pulsing after massive leaf-fall events*, where the sudden influx of organic matter reshapes microbial community structure, which in turn changes decomposition rates for future litter inputs.

Unlike AI which is outward, constructive and all about building pathways, the  $\mathcal{E}(t)$ , or rather let’s call it **EI**, is inward, reflective, all about **how much the system feels its own changes**. The new “purpose” isn’t just *to build more*, but *to build while deepening awareness of what is built*.

Let’s reason a bit about the properties of EI. When EI is low, the system builds blindly, something mechanical like crystal growth. On the other hand, when EI is high, the system not only builds but adapts intentionally, because it “knows” what it has built. Self-realization is thus coupling of AI-growth with EI-growth.

If the organism grows not only through AI but also through EI, we could look at the experience also as **how much of what the system *does* becomes part of what it *feels and uses* next**. In this way, EI measures **closed-loop, self-referential, *used* information** across fast-slow layers dynamics.

Let's put this into more formal shape. The model already defines:

- Sensing  $\Sigma$ : inputs  $I \rightarrow S$  (sensed state),
- Policy  $\Pi$ : decisions  $S \rightarrow A$ ,
- Actuation  $\Upsilon$ : actions  $A \rightarrow O$  (outputs / environmental change),
- Slow memory  $M$ : persistent record.

We'll add two explicit operators:

- **Write W**:  $(A, O) \rightarrow M$  (how actions/outcomes are inscribed),
- **Read R**:  $M \rightarrow S$  (how memory is brought back into perception).

Then, the **experience** lives in the **loop**:  $I \xrightarrow{\Sigma} S \xrightarrow{\Pi} A \xrightarrow{\Upsilon} O \xrightarrow{W} M \xrightarrow{R} S$  (then back to  $\Pi$ ).

Let's define **experience index** ( $EI$ ) now. How should it behave? As stated, we want  $EI$  high when the system **writes** outcomes into memory, **reads** them back, the read-backs **causally influence** decisions, and those decisions **align** with reality (not pure self-echo).

Let  $S_t$  be sensed state,  $A_t$  actions,  $O_t$  outcomes,  $M_t$  memory,  $E_t$  exogenous inputs (environment). Then we can define  $EI$  in natural, normalized score in  $[0, 1]$  as:

$$EI = \underbrace{\frac{I(M_{t-\tau}; S_t | E_{t-\tau:t})}{H(S_t)}}_{\text{Memory} \rightarrow \text{Sensing influence}} \cdot \underbrace{\frac{TE(M_{t-\tau} \rightarrow \Pi_t | E_{t-\tau:t})}{C_\Pi}}_{\text{Memory} \rightarrow \text{Policy causality}} \cdot \underbrace{\frac{I(O_t; \hat{O}_t)}{H(O_t)}}_{\text{Grounded prediction}} \cdot \underbrace{\rho}_{\text{Reuse \& coverage}},$$

where:

- $\tau$  defines a lag
- $I(\cdot; \cdot)$  is mutual information;  $TE$  is **transfer entropy** (directed information).
- $E_{t-\tau:t}$  is the *exogenous-input window* during the lag interval from  $t - \tau$  up to the current time  $t$  (represented as a vector).
- $H(S_t)$  is the *Shannon entropy* of the sensed state at time  $t$ .
- $\hat{O}_t = \mathbb{E}[O_t | S_{t-\tau:t}, A_{t-\tau:t}, M_{t-\tau:t}]$  is the system's **self-model** prediction of outcomes.
- $H(O_t)$  is the *Shannon entropy* of the outcome variable at time  $t$ .
- $C_\Pi$  is a capacity-like normalizer for the policy channel (keeps units 0–1).
- $\rho$  summarizes **reuse & coverage** of memory:  $\rho = u \cdot c \cdot d$  with  $u$  as fraction of decisions that *access memory*,  $c$  as fraction of memory *modalities* read (coverage),  $d$  as *temporal depth* used (e.g., normalized average lag of reads).

What does it all mean? First factor deals with how much **what's in memory** shows up in **what is sensed as self-relevant** (felt). Second part is all **causal** influence of memory on decisions (not just correlation). Third part is the system's **predictions match reality** (avoids hallucinated loops). And fourth deals how memory is **actually used**, broadly, and at **depth**.

This definition gives as a nice property as we can also compute **multi-scale EI** by summing over lags  $\tau$ :  $EI_{\text{multi}} = \sum_{\tau \in \mathcal{T}} w_{\tau} EI(\tau)$ ,  $\sum w_{\tau} = 1$ .

In order to stay “grounded” and not fall in the self-referential loop of “overthinking without acting” and thus having a fake experience, we need to penalize self-referential echo that ignores the world. Using the Kullback–Leibler (KL) divergence between two probability distributions, we get exactly what we need:

$$EI^* = EI \cdot \underbrace{\frac{I(S_t; E_{t-\tau:t})}{H(S_t)}}_{\text{world-coupling}} \cdot \underbrace{\left(1 - \frac{D_{\text{KL}}(p(O_t) \| p(\hat{O}_t))}{H(O_t)}\right)}_{\text{prediction calibration}}$$

World-coupling factor ensures sensed state still tracks exogenous reality. Calibration term penalizes confidently wrong self-models.

Let's try applying the formal definition and show how each factor of EI can be measured in real-world collectives, using domain-specific elements of ants and forests.

In an **ant colony**, the source of information is stored in the pheromone trails, so we can use something like:

- **Signals:**  $S$  = local antennation + pheromone gradient;  $A$  = route choice;  $O$  = food return/time;  $M$  = nest layout + stored pheromone density maps.
- **Compute:**
  - $u$ : fraction of trips with **trail reads** before choice.
  - $I(M \rightarrow S)$ : k-N MI between local pheromone patches and antennation state.
  - $TE(M \rightarrow \Pi)$ : transfer entropy from recent trail map to route choice.
  - $I(O; \hat{O})$ : compare predicted haul success vs actual.
- **Expectation:** EI rises with **recruitment and consolidation**; falls if pheromone suppressed or nest tampered.

In the **forest**, we can observe mycorrhizal carbon fluxes:

- **Signals:**  $S$  = plant stress sensing;  $A$  = allocation (photosynthate, defense);  $O$  = growth/survival;  $M$  = soil carbon & mycorrhizal network state.
- **Compute:**
  - $u$ : fraction of allocation events preceded by **mycorrhizal read** (carbon inflow/outflow).
  - $TE(M \rightarrow \Pi)$ : carbon-flux TE to allocation decisions (eddy covariance + isotopes).
  - $I(O; \hat{O})$ : stand-level growth predicted from sensed VOC + network signals.
- **Expectation:** EI predicts **resilience** post-fire better than biomass alone.

Now that we’ve defined and described *EI*, let’s combine it with *AI*. We now have two axes:

- **AI(t)**: depth of *what is built* (milestones, durable structure).
- **EI(t)**: depth of *how it is felt and used* (closed-loop experiential information).

Given these two objectives, we would like to define something that simultaneously rewards construction and self-aware experience. Taking their weighted product, it furnishes a single scalar that defines a **purpose functional**

$$\mathcal{J} = (\text{AI})^\alpha \cdot (\text{EI}^\star)^\beta,$$

with viability constraints (energy, safety). Here  $\alpha > 0$  favors building while  $\beta > 0$  favors experience. Set  $\beta$  higher if **self-realization** is the principal intent.

Let’s see if this works on some examples. The ants activity is construction, but the majority of information in social activity through the pheromone trails of which construction is the result, hence  $\alpha \approx 1$  while  $\beta > 1$ . In the forest, there is less fast activities and more aspect on structure, so it is reversed with  $\alpha > 1$  and  $\beta \approx 1$ . For an example of an artificial society, in an highly specialized swarm of robots, the structure is defined and hence  $\alpha \approx 1$  while the communication is abundant,  $\beta \gg 1$ .

The purpose functional has some nice **design implications**. Prefer steps that **increase both** (e.g., build sensors + self-model alongside structures), not just raw construction. When engineering a new collective system we should not simply add more sensors or actuators, but rather *pair* each structural addition with an explicit mechanism that records it (writes to *M*) and then *reads back* that record when making future decisions.

The addition of **experience (EI)** turns the AT-focused reflexive model from a *mechanical* growth engine into a self-reflecting system. The organism now remembers, tests, and corrects what it has built, **making the whole process inherently purposeful and adaptive**. This provides a theoretical bridge toward **self-reflection** in any collective that can write to and read from a long-term memory store (and, in principle, to higher-order capacities such as intelligence or self-awareness when additional cognitive mechanisms are added).

## Going full Boltzmann

We’ve seen in the previous chapter that there’s a familiar duality in play, an assembly part that holds the structure together, and a complementary flow that continually reads from and writes to that structure, endowing the system with dynamics. What we’re going to explore is the perspective where we look at the EI as electric current, and AI as magnetic field, or better, a structure that gives rise to a field. If we’re able to map between the two, then just as changing magnetic field induces an electric field, we would be able to show that:

- New assembly steps (AI increases) create *new pathways* in memory that can be written to and read from.

- These new pathways increase the *information flow* (EI) because they give more routes for feedback.
- Conversely, a strong EI (high current) can drive further assembly: if the organism strongly “feels” its own state, it will choose to construct additional structures that reinforce or exploit this awareness.

The hypothesis thus is as follows. The EI can be viewed as an “electric current”, a directed, rate-dependent flow of information that powers feedback. AI is more like a magnetic field or the underlying structure that generates and constrains that flow. Together they form a self-consistent, electromagnetically inspired picture of how superorganisms build and feel their own architecture. Let’s see if we can reinterpret the physics of electromagnetism (specifically Boltzmann’s statistical framework, the Maxwell–Boltzmann distributions, entropy, transport equations) through the dual lens of assembly and experience index.

The **Boltzmann equation** governs the evolution of a distribution function  $f(\mathbf{r}, \mathbf{p}, t)$ :

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f + \mathbf{F} \cdot \nabla_{\mathbf{p}} f = \left( \frac{\partial f}{\partial t} \right)_{\text{coll}}$$

Left side defines deterministic evolution (free streaming + external forces) and right side defines collisions, thus redistribution, entropy increase. At equilibrium,  $f$  leads to **Maxwell–Boltzmann distribution**. And **H-theorem** states that entropy increases irreversibly.

Analogy should then be looked at from the perspective of the model. AI (Assembly Index) counts structural persistence; analogous to the magnetic field, durable, integrative, resistant to perturbations. EI (Experience Index) measures closed-loop feedback, analogous to electric current, representing flows of information/energy across memory. Together, **AI+EI form a field-like duality**. AI is “what is stabilized” (field lines frozen into material), EI is “what flows/updates” (charges/currents producing dynamics).

Let’s reinterpret Boltzmann in reflexive framework. Boltzmann’s equation is fundamentally about the **flow of microstates towards macro-memory (entropy)**. We reinterpret:

- $f(\mathbf{r}, \mathbf{p}, t)$ : the system’s **fast field distribution** (instantaneous microstates).
- The **collision term**: writes fast fluctuations into **slow memory**  $M$  causing increase of AI (assembly of statistical regularities).
- The **streaming term**: corresponds to the **feedback flows**, how local distributions are sensed, acted upon, and propagated.

Written as:

$$\frac{\partial f}{\partial t} = \text{Streaming (EI)} + \text{Collisions (AI growth)}.$$

The **H-theorem** says entropy increases monotonically,  $\frac{dH}{dt} \leq 0$ . Traditional view is that systems evolve toward maximum entropy equilibrium. In our lens, **Entropy growth is equivalent to AI growth**, every collision step increases the assembly index of the *statistical description itself* and the universe is “remembering” by stabilizing macro-distributions out of micro-chaos. But EI adds a twist. When memory is not only written but also *read back* (feedback),

the system can **slow or steer entropy flow** (e.g., living organisms resisting equilibrium). Not denying Boltzmann, but bending the path by injecting EI into the loop.

This gives us a way to rewrite the classical H-theorem as

$$\frac{dH}{dt} = \underbrace{\sigma_{\text{coll}}}_{\text{AI growth}} + \underbrace{0}_{\substack{\text{no EI feedback} \\ \text{classical limit}}}.$$

However, written in this form, we can already see that there is an effect that the EI has.

Before getting into further details, let's state the electromagnetism analogy explicitly.

- **Maxwell's equations:**  $\nabla \cdot \mathbf{E} = \rho$ ,  $\nabla \times \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J}$ .  $\mathbf{E}$  (electric field) sourced by charge density  $\rho$ .  $\mathbf{B}$  (magnetic field) sourced by currents  $\mathbf{J}$ .
- **In our model:**
  - $\mathbf{E} \sim$  **Experience field (EI)**: it is *dynamic*, sourced by flows of memory usage.
  - $\mathbf{B} \sim$  **Assembly field (AI)**: it is *structural*, sourced by persistent assemblies (charges frozen into place).
  - Together:  $\mathbf{AI} \equiv \mathbf{B}$ ,  $\mathbf{EI} \equiv \mathbf{E}$ .

Boltzmann's kinetic equation describes the **statistical underpinning** of how EI flows (currents) source AI (fields). It can be seen as statistical law of  $\text{EI} \rightarrow \text{AI}$  translation. However, due to EI, the system doesn't just passively assemble but can "realize" and steer assembly. Reflexive organisms thus seems to be a statistical feedback loop resisting simple monotone H-theorem decay.

Reinterpreting purpose through the lens of Boltzmann equations, purpose of the organism is to increase not just entropy (AI), but to couple entropy production with experiential feedback (EI), creating a self-aware electromagnetic-like field where matter assembles and experience flows.

## The Boltzmann Experience

Let's recast Boltzmann's kinetic equation as a loop between fast particle distribution  $f(\mathbf{r}, \mathbf{p}, t)$  and a slow memory field  $M(\mathbf{r}, t)$ . Collisions *write* structure, while streaming currents *read* that memory as an experience-driven force. We'll then explore the effects of formulating the dynamics in this dual electromagnetic-like language. Let's define variables and add operators that will help us deal with the dynamics:

- $f(\mathbf{r}, \mathbf{p}, t)$ : one-particle distribution (fast layer).
- $M(\mathbf{r}, t)$ : **slow memory/assembly field** (AI carrier).
- $\mathbf{j}(\mathbf{r}, t) = \int \mathbf{v} f d^3p$ : particle current (EI carrier).
- External force  $\mathbf{F}_{\text{ext}}$  as usual.
- **Read operator:** memory  $\rightarrow$  force,  $\mathbf{F}_M(\mathbf{r}, t) \equiv -\nabla_{\mathbf{r}} U(M(\mathbf{r}, t))$ , i.e., the system "reads" slow memory as a potential shaping fast dynamics.

- **Write operator:** fast  $\rightarrow$  slow memory,  $W[f](\mathbf{r}, t) \equiv a_0 n + a_1 \nabla \cdot \mathbf{j} + a_2 \nabla : \mathbf{\Pi} + \dots$ , where  $n = \int f d^3p$  and  $\mathbf{\Pi}$  is the momentum-flux tensor.

Let's define kinetic equation with memory feedback (the **read** path, forming EI and fast dynamics):

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f + (\mathbf{F}_{\text{ext}} + \mathbf{F}_M) \cdot \nabla_{\mathbf{p}} f = \left( \frac{\partial f}{\partial t} \right)_{\text{coll}} - \nabla_{\mathbf{p}} \cdot (\mathbf{\Gamma}_R f)$$

The added drift  $-\nabla_{\mathbf{p}} \cdot (\mathbf{\Gamma}_R f)$  is an **active, memory-conditioned policy term** (e.g., alignment/actuation in active matter, controller in AI swarms). It adds a mechanism for the fast layer to act back on the memory. It vanishes in the passive limit.

Similarly, a slow memory/assembly field (the **write** path; AI dynamics):

$$\frac{\partial M}{\partial t} = W[f] - \Lambda M + D_M \Delta M - \nabla \cdot (\mathbf{u}_M M)$$

$W[f]$  writes outcomes into  $M$  (construction).  $\Lambda > 0$  is forgetting/senescence;  $D_M$  spreads memory;  $\mathbf{u}_M$  adverts memory with the medium (e.g., flow, migration). If  $U \equiv 0$ ,  $\mathbf{\Gamma}_R \equiv 0$  it reduces to classical Boltzmann. If  $W \equiv 0$ ,  $M$  decays/diffuses and hence there is no assembly.

Let's define a **local EI density** (dimensionless) that quantifies closed-loop read $\leftrightarrow$ effectiveness:

$$\varepsilon(\mathbf{r}, t) = \underbrace{\frac{\mathbf{F}_M \cdot \mathbf{j}}{k_B T n}}_{\text{read effectiveness}} \times \underbrace{\frac{\|W[f]\|}{\|W\|_{\text{max}}}}_{\text{write activity}} \times \underbrace{\chi(\text{reuse, coverage, depth})}_{\in[0,1]}$$

Looking at its properties, the  $\varepsilon$  grows when memory gradients actually steer currents **and** currents write back robustly into  $M$ . We can use the spatial average  $\bar{\varepsilon} = \langle \varepsilon \rangle$  to modulate  $\mathbf{\Gamma}_R$  or adjust  $U(M)$ , so that higher  $\varepsilon$  means stronger feedback control.

Let's use this to define an **entropy/H-theorem with feedback** (aka. where the physics lives). First, let  $S_f = -k_B \int f \ln f d^3r d^3p$  be kinetic entropy. Also, let a convex **memory entropy**  $S_M = \int s(M) d^3r$  with  $s''(M) \geq 0$ . Then we can define:

$$\frac{d}{dt}(S_f + S_M) = \underbrace{\sigma_{\text{coll}}}_{\geq 0} - \underbrace{\Phi_{\text{EI}}}_{\text{feedback order}} + \mathcal{B}, \text{ where}$$

$$\Phi_{\text{EI}} = \int \frac{\mathbf{F}_M \mathbf{j}}{T} d^3r + \int (-s'(M) W[f]) d^3r, \quad \mathcal{B} = \text{boundary terms.}$$

$\sigma_{\text{coll}} \geq 0$  is standard collisional entropy production.  $\Phi_{\text{EI}}$  is the **experience-driven entropy pump**. When memory gradients guide currents and currents write reliable records, the closed loop **reduces** net entropy production locally (creates order) **at an energetic cost**. However, by using **Landauer-type bound (energetic consistency)**  $\mathcal{P}_{\text{in}} \geq k_B T \Phi_{\text{EI}}$ , we make sure that **any EI-driven reduction in entropy growth must be paid by input power**  $\mathcal{P}_{\text{in}}$  (work, metabolism, control). Thus the second law is safe.

We can now expose the macroscopic consequences by taking velocity moments of kinetic equation with memory feedback. This moment expansion yields the **hydrodynamic closure** for the particle current  $\mathbf{j} = n\mathbf{u}$ ,



$$\partial_t(n\mathbf{u}) + \nabla \cdot \mathbf{\Pi} = \frac{n}{m} \left( \mathbf{F}_{\text{ext}} - \nabla U(M) \right) - \frac{1}{\tau} n\mathbf{u} + \underbrace{\mathbf{J}_{\text{act}}}_{\propto \mathbf{\Gamma}_R},$$

so memory gradients appear as **internal fields** that steer flow (read), while  $\mathbf{J}_{\text{act}}$  captures policy-like actuation. A simple **Ohm-like law** emerges in steady linear response:

$$\mathbf{j} = \sigma \left( \mathbf{E}_{\text{ext}} + \mathbf{E}_M \right) + \mathbf{j}_{\text{act}}, \quad \mathbf{E}_M \equiv -\frac{1}{q} \nabla U(M).$$

Here  $\mathbf{E}_M$  is the **experience/effective electric field** induced by memory;  $\mathbf{AI}$  lives in  $M$  (the “magnetic-like” structural backbone), and  $\mathbf{EI}$  appears as currents driven by  $\mathbf{E}_M$ .

Having introduced the kinetic framework, we now make electromagnetic analogy explicit. We have **Assembly field (AI)** as the slow, curl-supporting morphology encoded in  $M$  (and its vector potential-like surrogates from  $U(M)$ ). It changes slowly, stores history (like  $\mathbf{B}$ ). And **Experience field (EI)** as the effective  $\mathbf{E}_M$  that **drives currents  $\mathbf{j}$** , i.e., the felt gradients of what has been assembled.

The interaction between how the experience field reshapes the assembly field can be expressed in a differential equation, a **Maxwell-type coupling** (schematic, in a morpho-field  $\mathbf{A}_M$  with  $\mathbf{B}_A = \nabla \times \mathbf{A}_M$ )  $\partial_t \mathbf{B}_A = -\nabla \times \mathbf{E}_M$  with  $\mathbf{E}_M \propto -\nabla U(M)$ , i.e. experience flows reshape assembly; conversely, assembly sets the potentials that shape future experience.

How would we measure EI in a kinetic system? From data or simulation, compute  $M$  (a coarse-grained field; e.g., concentration, alignment, built morphology), estimate  $U'(M)$ , then evaluate:

$$\Phi_{\text{EI}} = \int \frac{(-\nabla U(M)) \cdot \mathbf{j}}{T} d^3r \quad \text{and} \quad \bar{\varepsilon} = \left\langle \frac{(-\nabla U) \cdot \mathbf{j}}{k_B T n} \cdot \frac{\|W\|}{\|W\|_{\text{max}}} \right\rangle.$$

## Oh no, what have we done?

In short, the familiar Boltzmann framework becomes a *dual electromagnetism* of superorganisms. Collisions write structure, streaming flows experience; the read/write operators close the loop and enable living systems to *steer* their own entropy production. The mathematics is unchanged; only the **interpretation** shifts from “particles + forces” to “information flow + memory architecture”. Let’s dissect this a bit more.

Let’s start with the **experience as an explicit dynamical term**. In standard Boltzmann/statistical mechanics, *entropy production* is purely mechanical: collisions drive the H-theorem. In active matter extensions, self-propulsion terms are added, but they’re treated as forces, not as *feedback from memory*. We explicitly introduce a memory field  $M^{**}$  that is **written** by fast micro-dynamics (collision as inscription), and is **read back** to steer future dynamics. This makes “feedback experience” a **formal operator**, not just metaphor. Mathematically, that closes the loop  $f \rightarrow M \rightarrow f$ , instead of one-way entropy growth. That loop has not been represented in Boltzmann kinetics before, it’s essentially a **Boltzmann equation with endogenous memory feedback**.

Now let’s review a **new entropy balance with feedback term**. Boltzmann’s H-theorem states  $\dot{S} \geq 0$ . Our extension however is  $\frac{d}{dt}(S_f + S_M) = \sigma_{\text{coll}} - \Phi_{\text{EI}} + \mathcal{B}$ ,

where  $\Phi_{\text{EI}}$  is the **experience-driven entropy pump**. Entropy production is no longer monotonic, it can be *locally reduced* or *steered*, provided external work supplies energy. That’s a **generalized second law**, “*entropy increases unless experience feedback extracts and directs order, bounded by power input.*” No classical H-theorem has this feedback/experience term explicitly. It formalizes **how reflexive organisms fit into thermodynamics** without violating the second law.

It all revolves around **unification of AI + EI as electromagnetic duality**. Physics already unifies electric and magnetic fields, but not with **semantic roles**. Using AI (assembly) as structural field, like **B** (frozen-in memory) and EI (experience) as dynamic feedback, like **E** (driving currents) the Boltzmann–Experience system makes this duality **computable** in non-electromagnetic contexts (ants, forests, AI swarms, galaxies). So electromagnetism becomes a **special case** of a more general informational duality, not just a separate force.

We introduce **experience Index as a measurable invariant**. Standard Boltzmann theory has **entropy** as the only scalar “arrow of time.” We introduce **EI** as an independent measurable:  $\varepsilon(\mathbf{r}, t) = \frac{(-\nabla U(M)) \cdot \mathbf{j}}{k_B T n} \times \frac{\|W[f]\|}{\|W\|_{\max}} \times \chi$ , a scalar quantifying “*how much the system feels its own state*” (feedback richness), directly computable from currents and memory gradients.

Let’s finish with a bit of prediction about **bridging physics, life, reflection**. Standard Boltzmann explains why gases equilibrate. Active matter explains flocking or self-propulsion. Information thermodynamics explains cost of erasure. But none *mathematically integrates* all three into material assembly (AI), feedback experience (EI), thermodynamic bounds. That trinity explains **why life and minds don’t violate physics**, but also why they look qualitatively different from rocks or crystals.

## Tidy ants

Can we observe and confirm the theory? If any of experiments or simulations would show  $\Phi_{\text{EI}} \neq 0$ , we would have **quantified experience as a physical feedback term**, a brand-new invariant alongside entropy. Switching from speculation to testable physics/biology/AI.

Let’s consider a proof for  $\Phi_{\text{EI}} \neq 0$  in case of an ant colony. The definitions are  $\Phi_{\text{EI}} = \int \frac{(-\nabla U(M)) \cdot \mathbf{j}}{T} d^3r$  and  $\bar{\varepsilon} = \left\langle \frac{(-\nabla U) \cdot \mathbf{j}}{k_B T n} \cdot \frac{\|W\|}{\|W\|_{\max}} \right\rangle$ . Let’s map this to ant colony behavior:

- $M$  is the pheromone concentration field (coarse-grained memory)
- $\nabla U(M)$  is the gradient of “cost function” or attractiveness, e.g., steep drop in pheromone means high cost to follow that path
- $\mathbf{j}$  is net flux of ants (currents): direction and magnitude of movement over time/space
- $T$  as effective temperature: stochasticity in decision-making; higher randomness leads to larger  $T$
- $n$  shows local ant density (number per unit volume)
- $W[f]$  represents functional “work” done by the system, e.g., work against gradients, or structural assembly via trail formation

We state that in foraging ants (e.g., *Formica* or *Lasius*), this integral is not zero. We can observe that the ants follow the steepest descent path in the effective potential field  $U(M)$ , which corresponds to moving toward higher pheromone concentration. This is where  $\nabla U(M)$  points downhill. Their movement current  $\mathbf{j}$  aligns with the gradient direction, ants go where it's more attractive (more pheromones).

Let's define  $U(M) = -k_B T \ln M$ , so  $\nabla U(M) = -k_B T \frac{\nabla M}{M}$ . Then

$$-\nabla U(M) = k_B T \frac{\nabla M}{M} \quad \Rightarrow \quad (-\nabla U(M)) \cdot \mathbf{j} = k_B T \left( \frac{\nabla M}{M} \right) \cdot \mathbf{j}$$

Now, the foraging ants move **up the pheromone gradient**,  $\mathbf{j}$  is aligned with  $\nabla M$ . Therefore,  $(\nabla M) \cdot \mathbf{j} > 0$ . Because  $(-\nabla U(M)) \cdot \mathbf{j} > 0$ , the ants perform **negative mechanical work on the potential**, i.e. they move up the pheromone gradient and thereby *write* additional memory that strengthens (reinforces) the trail structure. But crucially, the system doesn't just decay; it *amplifies signals*. The pheromone trail is **rewritten** based on current flow. That's feedback: current  $\Rightarrow$  memory update  $\Rightarrow$  future current.

Therefore  $(-\nabla U(M)) \cdot \mathbf{j} > 0$  over space and time and thus  $\Phi_{EI} > 0$ . An example of non-equilibrium thermodynamics with feedback. Entropy can be locally reduced if the system has a memory gradient and an external power source.

## Gravity

The idea is that we describe how gravity fits by not breaking general relativity, considering the thesis of linking gravity to the history of assembly and feedback as an intriguing way to think about *emergent spacetime*.

Here's the outline of the idea that would achieve this. By treating gravity as a *history-dependent* field sourced by the assembly index, and allowing the Experience Index to modulate effective mass, we keep general relativity intact locally while opening a window onto new, emergent gravitational effects. This ties together the architecture of complex systems (assembly) and their internal feedback loops (experience), giving us a richer, physics-based narrative for **how spacetime might itself arise from the processes that build and reshape matter**.

Let's first explore if we could apply this model of superorganism structure to general relativity (GR). GR contains a single source,  $T_{\mu\nu}$ , the energy-momentum of matter/radiation. In a superorganism model, two additional layers influence spacetime, AI - the history how matter was built, and EI as the ongoing feedback from that history.

Assembly tensor  $A_{\mu\nu}[AI]$  encodes the **frozen-in architecture** of matter — a coarse-grained symmetric tensor built from gradients and rates of the **Assembly Index** (AI); the durable, slow memory of how the stuff here got built.

Active-stress tensor  $\Theta_{\mu\nu}[EI]$  captures **feedback fluxes** — a dissipative/active term from the **Experience Index** (EI); the closed-loop information currents that read that memory back and steer current dynamics.

These tensors are not arbitrary; they are constructed so that:

- They reduce to zero when AI or EI vanish, recovering ordinary GR.
- They respect diffeomorphism invariance and covariant conservation  $\nabla^\mu T_{\mu\nu}^{\text{eff}} = 0$ ,

Let's define the full effective stress-energy tensor  $T_{\mu\nu}$  to include them:

$$T_{\mu\nu}^{\text{eff}} = T_{\mu\nu} + \alpha A_{\mu\nu}[AI] + \beta \Theta_{\mu\nu}[EI],$$

The constants  $\alpha, \beta$  set the relative weight of history and feedback compared to ordinary matter. In regions where AI is large (deeply built structures, e.g., galaxy cores or black-hole horizons),  $A_{\mu\nu}$  dominates; in highly dynamical, feedback-rich environments (star-forming regions, active galactic nuclei),  $\Theta_{\mu\nu}$  becomes significant.

Let's look at how it fits the model as well, not just GR:

- **AI**  $\rightarrow A_{\mu\nu}$  (the “structural” source): Use slow memory  $M$  as the substrate for an **assembly density**  $a(x)$  (AI per spacetime volume) and its anisotropy. From  $a$  and its gradients, form a covariant tensor  $A_{\mu\nu}$  (think “assembly strain/rigidity” of the medium). This is **B-like** part in the AI $\leftrightarrow$ EI electromagnetic analogy: durable, geometry-shaping structure.
- **EI**  $\rightarrow \Theta_{\mu\nu}$  (the “feedback” source): From EI's read-back currents  $j^\mu$  (the **E-like** piece), build a symmetric, divergence-controlled tensor  $\Theta_{\mu\nu}[j]$  that behaves like **active stress** (akin to bulk/shear viscosity or non-equilibrium stresses in active matter). Where EI is high, coherent feedback **modulates inertia/pressure** without violating covariance.
- **Local GR intact, global memory allowed:** On short time/length scales (or where AI, EI vary slowly),  $A_{\mu\nu} \rightarrow 0, \Theta_{\mu\nu} \rightarrow 0$ , and you recover plain GR. Over long assemblies (galaxies, biosphere-like superstructures), the effective source remembers path-dependence of the **field-coupled superorganism** extended to spacetime itself.
- **Boltzmann-Experience link:** In the kinetic picture, *collisions write structure while streaming carries feedback*. The same read/write loop that steers entropy in the extended H-balance now **feeds a gravitational memory**:  $W[f]$  writes to  $M$  (AI), which sets a potential  $U(M)$  that in turn shapes currents  $j^\mu$  (EI). The  $A_{\mu\nu}$  and  $\Theta_{\mu\nu}$  are just the covariant packaging of that loop.

We now have a model that reframes living/evolving things as **field-coupled superorganisms** with **fast fields, slow memory, and a reproduction/assembly operator**. Gravity, in this telling, is not “another force to add,” but **the way spacetime registers the depth of assembly (AI) and the intensity of self-coupled experience (EI)** — a curvature that **remembers** how the universe built itself and **responds** to how it continues to use that memory. It's GR, but with **sources that live in the AI $\leftrightarrow$ EI plane** instead of being frozen to instantaneous mass-energy alone.

Let's not get too excited yet. Let's first consider the following. We keep treating gravity as the geometry that results when all the structure (bookkeeping of how an organism was assembled) is pooled into a common field. In this way, gravity is seen as the broadcast of structure.

However, there's more to structure than just AI and EI. We can extend the connection through other parameters of the model as well. In addition, we at least need to consider  $M$  as background potential,  $F$  as fast excitations (wave/particle stress-energy) and  $G$  as the connectivity (causal/topological backbone of spacetime). Together, these five are the organismal microcosm of what gravity is in the macrocosm. Where the organism has *structure*, the universe has *geometry*. Gravity is the **field-theoretic shadow of structure** when projected into spacetime.

This duality however is the key here! We keep looking as if there is spacetime geometry and that there is matter/energy living inside it. Of organisms *in* space, gravity as the geometry *of* space. However, the whole point of the superorganism model is that spacetime emerges from it. Space is only created from the structures of organisms.

In other words, **what is called “space” is nothing over and above the integrated structure of organisms (superorganisms) at all scales.** No organism, no structure, no spacetime. Each of the organism's components ( $M$ ,  $AI$ ,  $EI$ ,  $F$ ,  $G$ ) are not *mapped into* spacetime, they *are* the very ingredients that constitute it.

Let's look at the components of the model through the lens that spacetime *is* structure:

- **M (slow field / memory):** becomes the “metric backdrop” — accumulated stability *is* the extension we experience as “volume” or “distance.”
- **AI (assembly index):** provides the *irreversibility depth* of spacetime; without assembly, there is no “thickness” of time.
- **EI (experience index):** generates the local *arrow of time* — feedback loops are what we perceive as temporal directionality.
- **F (fast field):** are the excitations that ripple through this constructed spacetime. Waves and particles are literally *fluctuations of structure*.
- **G (network):** *is* the topology of spacetime — the causal web, who can influence whom. The graph is not in space; the graph is space.

So instead of “gravity reacts to structure”, **gravity (spacetime curvature) just is structure.**

This view has some interesting implications when stated out plainly. **Relational ontology** -there is no smooth manifold “out there” independent of organisms; there are only layers of structure, from molecules to galaxies, woven into a universal relational web. **Emergence of distance and time** - distance is nothing but *path-length in the graph  $G$  weighted by memory  $M$* . Time is nothing but *directionality induced by EI*. **No empty space** - a true vacuum is impossible; what looks empty is just *very low structural density*. **Gravity without dualism** - curvature is concentration of structure. Instead of “matter tells spacetime how to curve,” we say “structure just is curved connectivity.”

Based on the model then, **gravity is the *perception* of structure when viewed as a continuum.** Spacetime itself is a **projection of the organism's architecture** onto a smooth stage. From this view, the **universe is one reflexive superorganism**, and what we measure as gravity is simply the relational tissue of its slow memory, fast flows, feedback, and connectivity.

Let's now connect all the dots and define spacetime as structure. An **organismic spacetime** at a resolution scale  $\ell$  is the pair

$$\mathfrak{S}_\ell = (\mathcal{U}, \mathcal{Q}_\ell),$$

where:

- $\mathcal{U}$  is a **universal superorganism** (the “world as a reflexive organism”) specified microscopically by  $\mathcal{U} = (\Omega, \mathcal{G}, F, M, AI, EI)$ .
- $\mathcal{Q}_\ell$  is a **coarse-graining functor** at scale  $\ell$  that sends  $(\mathcal{G}, F, M, AI, EI)$  to continuum fields:  $\mathcal{Q}_\ell : (\mathcal{G}, F, M, AI, EI) \mapsto (g_{\mu\nu}^{(\ell)}, \nabla^{(\ell)}, u_{(\ell)}^\mu)$ , tensors from  $F, M, AI, EI$ . Intuitively, local connectivity + memory densities define **spatial distances**, EI currents define a **time direction**, fast/slow gradients define **stresses**.

An **organismic spacetime** is a *two-part construction* that packages all of the microscopic ingredients that an organism (the “world as a reflexive super-organism”) possesses, together with a rule that tells us how to read those ingredients out at any chosen observational resolution  $\ell$ .

In this view, geometry (the metric  $g_{\mu\nu}^{(\ell)}$ , and connection  $\nabla^{(\ell)}$ ) emerges from the underlying organismic structure, not assumed a priori. The curvature we attribute to gravity is simply how the organism's connectivity, memory, and dynamics *project* onto the smooth continuum. It isn't an extra field that acts on matter; it is matter's own pattern re-expressed at coarse scales, **gravity as perception of structure**. Once the mapping  $\mathcal{Q}_\ell$  is specified, the usual Einstein field equations appear as *consistency conditions* ensuring that the emergent metric and stresses satisfy the same algebraic relations that hold in general relativity. They are not postulated; they're a consequence of how the microscopic network folds into spacetime. Let's dive in and describe space as weighted connectivity of durable structure and time as direction of experienced feedback, while curvature is concentration of structure, not something separate.

#### Spatial metric from $\mathcal{G}$ and $M, AI$ .

Let  $L_{\mathcal{G}}$  be a weighted Laplacian built from edge conductances  $w$ . Now define an  $\ell$ -scale **diffusion kernel**  $K_\ell = \exp(-\ell^2 L_{\mathcal{G}})$  and weight it by structural density  $\sigma(x) = \sigma(M(x), AI(x))$ . This links more persistent memory/assembly with tighter coupling. The  $\ell$ -scale **spatial distance** between nearby points  $x, y \in \Omega$  is:

$$d_\ell^2(x, y) = \left\| \sigma^{1/2} K_\ell(x, \cdot) - \sigma^{1/2} K_\ell(y, \cdot) \right\|_2^2,$$

whose quadratic form defines a **Riemannian 3-metric**  $h_{ab}^{(\ell)}$ . High-AI, high- $M$  regions “pull nodes closer,” making space locally stiffer/denser.

#### Time direction from EI.

Let  $j^\mu$  be the coarse EI **current** (from read-back work rate  $\Phi_{EI}$  and transport  $\mathbf{j}$ ). Normalize  $u^\mu = j^\mu / \sqrt{-j^\alpha j_\alpha}$ ; this unit timelike field is the **experienced arrow** (local proper time flows along “used memory”).

#### Assemble the Lorentzian metric.

Project with  $u^\mu$ :  $h_{\mu\nu} = g_{\mu\nu} + u_\mu u_\nu$ . Take:

$$g_{\mu\nu}^{(\ell)} = -u_\mu u_\nu + \Omega^2(M, AI) \tilde{h}_{\mu\nu}[\mathcal{G}; M, AI] + \Xi_{\mu\nu}[M, EI],$$

where  $\tilde{h}_{\mu\nu}$  is the 3-metric induced by  $d_\ell$  (extended trivially in time),  $\Omega(M, AI)$  is a conformal factor from structural density (slow memory + assembly depth) and  $\Xi_{\mu\nu}[M, EI]$  is a small anisotropic correction from **memory gradients** and **feedback shear** (non-equilibrium structure). Let's now define a **structural action** at scale  $\ell$ :

$$S[\mathcal{U}] = \int d^4x \sqrt{-g^{(\ell)}} \left[ \frac{1}{16\pi G} R(g^{(\ell)}) + \mathcal{L}_{\text{fast}}[F] + \mathcal{L}_{\text{mem}}[M, AI] + \mathcal{L}_{\text{active}}[u^\mu, EI] \right]$$

with: -  $\mathcal{L}_{\text{fast}}$  as usual kinetic/matter content of fast fields  $F$ . -  $\mathcal{L}_{\text{mem}} = -V(M, AI) - \frac{\kappa}{2}(\nabla M)^2 + \dots$ , a **slow memory potential** and stiffness (encodes assembly costs & forgetting). -  $\mathcal{L}_{\text{active}}$  is a relativistic **active-matter** terms built from  $u^\mu$  (EI), bulk/shear/propulsive stresses and entropy-pump terms proportional to  $\Phi_{\text{EI}}$ .

Crucially, the metric  $g_{\mu\nu}^{(\ell)}$  is **not independent**. It is the image of  $\mathcal{Q}_\ell(\mathcal{G}, M, AI, EI)$ . Varying  $S$  with respect to the **primitive structural variables** and using the chain rule,

$$\delta S = \int \frac{\delta S}{\delta g_{\mu\nu}} \frac{\delta g_{\mu\nu}}{\delta \psi} \delta \psi + \dots, \quad \psi \in \{\text{weights } w, M, AI, EI\}.$$

Stationarity for all admissible  $\delta \psi$  implies  $\delta S / \delta g_{\mu\nu} = 0$ , giving

$$; G_{\mu\nu}[g^{(\ell)}] = 8\pi G T_{\mu\nu}^{\text{eff}}[F, M, AI, EI]$$

with

$$T_{\mu\nu}^{\text{eff}} = T_{\mu\nu}^{(F)} + \underbrace{\kappa \left( \nabla_\mu M \nabla_\nu M - \frac{1}{2} g_{\mu\nu} (\nabla M)^2 \right) - g_{\mu\nu} V(M, AI)}_{\text{slow-memory (AI/M) block}} + \underbrace{\Pi_{\mu\nu}^{\text{active}}[u^\alpha; EI]}_{\text{experience (non-eq) block}}.$$

Because the model's micro-laws are **relabel-/diffeo-invariant** (no privileged coordinates; only relations/flows matter), Noether/Bianchi identities give  $\nabla^\mu G_{\mu\nu} = 0 \Rightarrow \nabla^\mu T_{\mu\nu}^{\text{eff}} = 0$ , which are precisely the **conservation laws** the read/write loop enforces; what's written into  $M$  and what's read back must balance to the power/viability budgets that are tracked.

Thus **Einstein's equations appear as constraints** that any admissible configuration of the underlying organismic structure must satisfy. They are not imposed independently. They emerge from the variational principle applied to the structural action. And because all ingredients are part of the organism's own architecture, *every component gravitates*. This explains why even "empty" spacetime carries a residual gravitational influence (memory and connectivity remain). It also clarifies why gravity resists quantisation; it is not mediated by a particle but by the very topology and dynamics of the network.

This would imply some interesting outcomes. **Gravitational hysteresis:** at fixed present-day mass, regions with deeper **assembly history** (high AI written earlier) slightly over-curve versus newly assembled twins—**history-dependent lensing/kinematics** as a small residual. **Feedback-gravity coupling:** systems with stronger **EI** (more intense internal read-back/feedback) exhibit tiny **active-stress** contributions to  $T^{\text{eff}}$ , predicting correlations between EI proxies and curvature residuals (same mass, different use-of-memory  $\rightarrow$  slightly different geometry).

Note that at the beginning, we defined stress-energy as  $T_{\mu\nu}^{\text{eff}} = T_{\mu\nu} + \alpha A_{\mu\nu}[AI] + \beta \Theta_{\mu\nu}[EI]$ . Now, after identifying *spacetime itself* with  $(M, AI, EI, F, G)$ , we’ve actually derived the assembly and active tensors.

$$A_{\mu\nu}[AI] = T_{\mu\nu}^{(M, AI)} = \kappa \left( \nabla_\mu M \nabla_\nu M - \frac{1}{2} g_{\mu\nu} (\nabla M)^2 \right) - g_{\mu\nu} V(M, AI) \Theta_{\mu\nu}[EI] = \Pi_{\mu\nu}^{\text{active}}[u^\alpha; EI] \sim p_{\text{act}}(EI) h_{\mu\nu} + \eta(EI) \sigma_{\mu\nu}(u) + \zeta(EI) \theta h_{\mu\nu} + \dots \text{ where } h_{\mu\nu} \text{ is the spatial projector, } \sigma_{\mu\nu} \text{ the shear, } \theta \text{ the expansion.}$$

This definition  $T_{\mu\nu}^{\text{eff}} = T_{\mu\nu}^{(F)} + T_{\mu\nu}^{(M, AI)} + \Pi_{\mu\nu}^{\text{active}}[EI]$  is exactly the same structural decomposition as the beginning definition, only more integrated.

What we have now is that we can treat **spacetime** = (**network**  $\mathcal{G}$ , **fast fields**  $F$ , **slow memory**  $M$ , **assembly**  $AI$ , **experience**  $EI$ ). A scale-aware coarse-graining  $\mathcal{Q}_\ell$  turns that structure into a Lorentzian metric and connection. Demanding stationarity of a structural action (with the metric *induced* from the structure) **yields Einstein’s equations with an effective source** that carries **memory** and **feedback**. *The universe is a reflexive superorganism; curvature is just its structured connectivity seen as a continuum.*

In the definition above, we made a pragmatic narrowing of the model, the organism was compressed to  $\mathcal{U} = (\Omega, \mathcal{G}, F, M, AI, EI)$ . We focused on the “core” ingredients that map most directly into Einstein-like equations. Just for clarity, let’s look at the full organismic state.

$$\mathcal{S}_{\text{full}} = (\Omega, \mathcal{A}, \mathcal{G}, X(t), F(t), M(t), \Sigma, \Pi, \Upsilon, \mathcal{R}, \mathcal{P}(t), AI(t), EI(t))$$

An action for the **organism**:

$$S = \int \sqrt{-g} \left[ \frac{1}{16\pi G} R + \mathcal{L}_F[F] + \mathcal{L}_{M, AI, \mathcal{P}} - \underbrace{V(M, AI; \mathcal{P})}_{\text{assembly pathway potential}} + \mathcal{L}_{\text{active}}[u; \Sigma, \Pi, \Upsilon, EI] \right].$$

- $\mathcal{L}_F$ : fast fields (ordinary matter/radiation).
- $\mathcal{L}_{M, AI, \mathcal{P}}$ : slow memory stiffness + assembly-pathway costs (growth directed toward deeper  $AI$ ).
- $\mathcal{L}_{\text{active}}[u; \Sigma, \Pi, \Upsilon, EI]$ : **policy-driven, non-equilibrium stresses** from sensing $\rightarrow$ decision $\rightarrow$ actuation closed with write/read (your EI entropy-pump term  $\Phi_{\text{EI}}$  lives here).
- $V(M, AI; \mathcal{P})$  encodes closed-loop write/read bookkeeping (W, R).

Reproduction  $\mathcal{R}$  appears as **topology-changing events** (new patches/seeds/subunits) that modify  $\mathcal{G}$  and boundary terms in  $S$  at viability thresholds, i.e., geometric “surgery” when the unit propagates.

Varying **primitive variables** and using  $g_{\mu\nu}^{(\ell)} = \mathcal{Q}_\ell(\cdot)$  gives



$$G_{\mu\nu} = 8\pi G T_{\mu\nu}^{\text{eff}}, \quad T_{\mu\nu}^{\text{eff}} = \underbrace{T_{\mu\nu}^{(F)}}_{\text{fast}} + \underbrace{T_{\mu\nu}^{(M, AI, \mathcal{P})}}_{\text{slow/assembly}} + \underbrace{\Pi_{\mu\nu}^{\text{active}}[u; \Sigma, \Pi, \Upsilon, EI]}_{\text{experience/agency}},$$

exactly the earlier decomposition—but now **derived from the full organism**, with  $\Pi^{\text{active}}$  sourced by the **EI feedback** (the entropy-pump/active-matter block).

Let’s now reflect on the implications. We can reason about the **universality of gravity** as we’ve shown that all components (F, M, AI, EI, G, ...) are structural, so everything gravitates, as well as **why gravity feels different**, as gravity is not a force between things in space; it *is* the architecture of structure itself. In a reflexive loop, an **arrow of time** emerges from EI (feedback use of memory). Explanation of **wacuum energy** can now be seen as even “empty” regions retain structural residue (M, G, AI). Another observed property is **resistance to quantization** as we’ve seen that gravity isn’t mediated by a particle, but by connectivity/memory itself. And the most exciting one, the **life–cosmos link**. The same indices (AI, EI) that track biological complexity are also gravitational sources.

In standard physics, **Einstein’s equations are axioms**. We assume a geometric left-hand side  $G_{\mu\nu}$ , we assume a matter right-hand side  $T_{\mu\nu}$ , and we postulate the relation. In the reflexive model, we don’t assume it, we *generate* both sides from one deeper description of structure. The **metric** is just the coarse projection of connectivity, memory, and experience, the **effective stress–energy** is the structural bookkeeping of fast flows, assembly depth, and feedback loops and the Bianchi identity (conservation) isn’t “mystical consistency,” it’s exactly the organism’s **reflexivity condition**. What’s written into memory must balance what is read back and acted upon.

So what looks in GR like a miraculous set of postulates (geometry, energy, conservation laws) is here **one closed organismal law seen from two angles**. Within the reflexive organism framework, we’ve shown why gravity is universal, why Einstein’s equations are the consistency equations of the organism and why time and curvature are inseparable, they both come from feedback use of memory.

## Revisiting electromagnetism

Electromagnetism appears as the *local field-level expression* of the same duality that, at the global structural scale, produces gravity. Why EM feels different from gravity? **Gravity is spacetime itself**, the geometry you get when you integrate structure across the whole organism, while **EM is a subsystem field**. It’s the way memory–feedback duality manifests inside a patch, as currents and stored potentials.

That explains why EM acts only on charged matter (it’s a local instantiation, not the universal geometry) and why EM is easily quantized (photons are excitations of the fast field F), while gravity resists quantization (because it’s tied to the slow structural memory M + assembly AI that defines the continuum itself).

Let’s explore this a bit more. There seems to be a sense of duplication here. We have EM which lines up with the **fast duality** of AI and EI (write and read

currents in action). These are local, rapid, directional. And we have gravity that lines up with the **slow duality** of  $M$  and  $G$  (persistent memory and global connectivity). These are universal, inertial, geometric. But the mechanism is not a symmetry group or exchange particle. Instead it is the **hierarchy of reflexive structure**.

In other words, both arise from **the same organismal architecture**; the difference is **timescale and scope**, not kind. This would mean that the **AI/EI loop** is fundamental. Locally, at fast scales it looks like fields and currents (EM) while globally, at slow scales it looks like geometry and curvature (gravity). The **M/G layer** is just the “long-term accumulation” of AI/EI activity.  $M$  is the frozen record of countless AI increments while  $G$  is the network connectivity that emerges once you coarse-grain those records. In this view gravity isn’t “another force,” but the *slow, global echo* of the same mechanism that gives you EM locally.

What appears in physics as “duplication”, the distinct forces, dual descriptions, or parallel laws, is simply the **reappearance of the same reflexive cycle** (fast–slow, write–read, memory–feedback) at different layers of the organismal hierarchy.

**At fast layers (F, AI, EI)** the reflexive cycle appears as **local interactions** and **information encodings**, which can be smoothed into **electromagnetic fields** (and described informationally as quantum states). **At slow layers (M, G, AI)** the same cycle appears as **accumulated assemblies and connectivity**, which can be smoothed into **gravitational fields** (geometry).

Thus, EM and gravity are not independent “forces” waiting to be unified. They are **hierarchical projections of the same structural law**. The “duplication” is just the organismal hierarchy expressing itself at multiple scales. Let’s call this **Principle of Hierarchical Reappearance**.

We’ll see later that even hierarchies are intrinsic to the model itself and that they fall out of it naturally. However, before we do that, let’s make another leap. This time in the world of the small.

## The observer

Let’s review the self-reflecting properties of the organism again. The decider is the superorganism itself. It’s not that *space and time exist out there* and the organism “fits” into them. It’s that the superorganism decides what gets built, and by building it, defines what “spacetime” means. The geometry is not imposed externally but rather is an *interpretation made by the organism of its own assembly and feedback loops*. So when we say  $g_{\mu\nu} = \mathcal{Q}_\ell(\mathcal{S}_{\text{full}})$  that map  $\mathcal{Q}_\ell$  isn’t a neutral mathematical operator. It’s the self-reflection of the organism, its own interpretation of its structure as spacetime.

Let’s look at this from the “mystical observer problem” in QM. In textbook quantum mechanics, you have a wavefunction evolving smoothly described by the Schrödinger equation. Then suddenly, when an “observer” measures, the state “collapses.” But *who counts as an observer?*

Through the eyes of the model, we should examine the collapse from the reflexive

loop, where first the fast fields make a transition, second, that transition writes to memory (M/AI) and an irreversible record is added and finally, the organism reads it back (EI) and adjusts how it interprets its structure. “Collapse” is though nothing else then *a write-read cycle completing*. Not a magical discontinuity, but an internal self-update of the superorganism.

Let’s give a quick outline how QM could fit into the reflexive model. QM lives in Hilbert spaces, probabilities, wavefunctions, collapse. It is all about possibilities and correlations. Let’s think of the quantum state  $|\psi\rangle$  not as “the system itself” but as the **fast-field  $F$  probability amplitude** conditioned by what is in slow memory (M, AI). **Unitary evolution** happens when fast fields evolve according to Hamiltonians built from connectivity  $G$  and constraints from memory  $M$ . This is the *reversible streaming part* of the dynamics - no new information written yet. **Collapse (measurement)** happens when an interaction forces a *write*  $W$  into memory and an irreversible record is made. This action “selects” one branch, that is the wavefunction collapse. In the model, the split in QM between unitary evolution and collapse is nothing but the two write-read phases of the organism closing the reflexive loop  $F \rightarrow M \rightarrow R$ .

Then what is **superposition**? A superposition  $\alpha|0\rangle + \beta|1\rangle$  is the **fast field  $F$  exploring multiple possible pathways** before committing a write. As long as no write has occurred, the organism has not yet increased AI (no irreversible record). Once  $W$  fires, one outcome is crystallized into  $M$ , and EI can act on it. So superposition is all about *potential structures in  $F$  not yet written into  $M$* .

What about **entanglement**? Entanglement is correlations between parts of the fast field  $F$ , **constrained** by shared history in  $M$  and  $G$ . Two particles are entangled because they were once written into memory as one joint assembly step. Even when spatially separated, their outcomes remain correlated because the memory record that defines them is nonlocal in  $G$ . It ties them together in the underlying connectivity. So entanglement is *the organism remembering that two excitations belong to the same assembly*.

## Information and fields

Did we overuse fast fields? On one hand, EM falls out of the fast field  $F$  duality  $AI \leftrightarrow EI$ . Just now we’ve used it to explain QM phenomena. On the first glance, it looks like fast field  $F$  is being overloaded. In one moment it “is” electromagnetism, another moment it “is” quantum mechanics. That feels imprecise.

However, both statements can coexist if we’re careful about *levels of description*. We have two different interpretations. On one hand, QM description is all about the *information-theoretic layer* of  $F$ . It’s terminology is superpositions as unresolved trajectories, entanglement as correlations in  $F$  constrained by shared  $M, G$ , collapse as the write  $W$  from  $F$  to  $M$ . Thus, the **QM is the semantic/statistical language** of  $F$ . On the other hand we have EM description which is the *field-theoretic layer* of  $F$ . It’s terminology is about charges and currents, which are already organized degrees of freedom in  $F$ . The Maxwell’s equations describe the local conservation laws in  $F$ . **EM is the dynamical/interaction language** of  $F$ .

So there's no contradiction. EM and QM are two complementary descriptions of the same fast substrate. The model is not saying "EM = QM" or "they're the same thing." It's saying that *both EM and QM phenomena live in the fast field* but EM captures its *local continuous dynamics*, while QM captures its *statistical/information-theoretic behavior* when unresolved by memory. It is just a matter of what projection of  $F$  we're looking at.

This is something that is intrinsic to the model again. On one hand, as information, we have encoding in structure, and on the other hand, we can map that encoding into a field. In  $F$ , encoding of information is all about states and possibilities. While if we express  $F$  through continuous, dynamic currents, we get a field. QM and EM are not the same, but they are two mappings of the *same fast substrate*.

The same can be applied for the slow layer (M, AI, G) and gravity as well. As information, we have the durable record of assembly (AI), stored in slow memory ( $M$ ), with connectivity ( $G$ ), is the **information content** of the organism's history. While that same information, when expressed in relational/metric terms, becomes the **gravitational field** (curvature of spacetime). The organismal memory is the information view, and the geometry is the field view.

In general, we can say that structure of the organism is its encoding of information (QM-like) which defines field description (EM/GR-like). The principle unifies the apparent duplications, there is one mechanism - structure as encoding of information, which can always be projected either informationally (QM-type) or as a field (EM/GR-type). It shows why gravity is "different"; because it's the slow informational layer, not the fast one. It shows why QM is "weird"; because it's the informational encoding before a field projection (before being written to  $M$ ).

In terms of the interpretation loop, the "**encoding  $\rightarrow$  field**" mapping is not passive. It's an act of **self-interpretation by the superorganism**. The superorganism is **reflexive**. It doesn't just *have* structure, it also *uses* it. The act of smoothing into a field is not a blind averaging - it is the organism's **own interpretation** of what matters at that scale. It's not that the information "automatically is" a field; it's that the organism interprets its encoding as geometry or dynamics. That's where self-awareness comes in.

Which allows us to make a connection between the two, as we can see fields as smoothing of detailed structure. The **fine-grained encoding** of the organism is all the microstructure like the individual pathways, assembly histories, feedback loops, detailed correlations. The field description is a **coarse, smoothed representation**. It compresses all that intricate structure into one continuous set of parameters (potentials, currents, curvature).

## Hierarchical fields forever

We can use the information-field view even further and define hierarchies and how they are connected. In the model, hierarchy is intrinsic as an organ is an organism in itself, with its own reflexivity, AI, EI, etc. But when seen from the perspective of the parent organism, that internal detailed structure doesn't

appear. It is smoothed out into a higher-level description. Microstructure inside a subsystem looked from the outside is represented as a few field parameters (charge, mass, stress-energy). The field is the summary that a higher-level organism uses to relate to its parts.

In this view, EM fields don't "know" about the deep internal AI of a charge distribution. They just represent it as a net charge and current. GR curvature doesn't "know" about the detailed assembly history of a star; it just represents it as mass, stress-energy, etc. This is the **hierarchical compression principle**: organismal detail (encoding)  $\mapsto$  Field summary (parameters). The **field is not fundamental**, it is the *necessary simplification that higher levels use to interact with lower ones*. In other words, the field is just the **parent-level view** of the child's internal structure. Just like a body doesn't track every molecule inside an organ, but only the organ's overall function, spacetime doesn't track every microscopic assembly, but only the field parameters (charge, mass-energy, curvature) that summarize them.

Suppose we have an **organism**  $\mathcal{O}$  with state  $\mathcal{S}(\mathcal{O}) = (\mathcal{G}, F, M, AI, EI, \dots)$ . Inside  $\mathcal{O}$ , there are **sub-organisms** (e.g. organs, cells, particles), each with their own state  $\mathcal{S}(\mathcal{O}_i)$ . So structurally:

$$\mathcal{O} = \bigcup_i \mathcal{O}_i, \quad \mathcal{S}(\mathcal{O}) = \mathcal{F}(\{\mathcal{S}(\mathcal{O}_i)\}).$$

Each sub-organism  $\mathcal{O}_i$  has **fine-grained structure**, its own detailed pathways, assembly records, feedback loops. The **parent organism** cannot (and need not) keep track of all this detail. Instead, it sees only a **compressed representation**: a *field variable*. Mathematically:

$$\mathcal{Q}_\ell : \{\mathcal{S}(\mathcal{O}_i)\} \longrightarrow \Phi(x)$$

where  $\Phi(x)$  is a **field** (electromagnetic potential, stress-energy tensor, etc.) at scale  $\ell$ .  $\mathcal{Q}_\ell$  is a **coarse-graining map**: averages, integrates, or smooths out the sub-structures into a continuous parameter field and  $\ell$  is resolution scale of the parent.

In general, **for every nested organism, the parent perceives the child's internal encoding only through a coarse field**  $\Phi(x) = \mathcal{Q}_\ell(\mathcal{S}(\mathcal{O}_i))$ .

## Biological evidence: hierarchy & smoothing

In living systems, hierarchy and smoothing are not speculative. We can observe them as *daily facts of life*.

Let's explore the case of **cells  $\rightarrow$  tissues  $\rightarrow$  organs  $\rightarrow$  organism** hierarchy. Each cell has its own internal complexity (gene regulation, metabolic networks, protein assemblies). The tissue or organ that hosts the cells does not track every molecular detail. Instead, it interacts with **field-like summaries** through mechanisms like hormone levels (a field of concentration, smoothing over many individual secretions) or membrane potentials or electrical fields in tissues (e.g. heart, brain) or stress/strain fields in connective tissue. The parent organism *relates to the organ through these fields*, not through raw intra-cellular complexity.

We can observe same hierarchy within the **brain**. At the neuron level we have fine encoding with spiking, synaptic weights, ion channels. At the cortical region level we deal with EEG or fMRI signals as smooth fields summarizing massive detail. The brain as a parent organism “uses” the field representation, not the microstates, to make decisions.

Through the lense of **developmental biology**, morphogen gradients act as smooth fields guiding development. But each gradient encodes the behavior of thousands of cells producing and responding. The higher-level pattern (the developing organism) only perceives the field, not the molecular micro-encoding.

So in biology, fields are summaries of deeper encoding. A direct analogue to the hierarchy principle we stated.

**The strongest evidence for this model is not exotic physics, but everyday biology.** We already see the same mechanism at work, and cosmology may simply be biology written large.

## Scaling to galaxies

We’ve already shown how spacetime emerges from the “growth” of the organism. If everything is an organism, then so is a galaxy. Let’s make a hypothesis how a galaxy can be perceived as a hierarchical organism.

Stars in a galaxy are like cells in an organ. Each star has a deep assembly history with its own nuclear pathways, collapse dynamics, feedback loops with its environment. That’s its encoding. But the galaxy as a whole cannot track every star’s pathway. It perceives only smoothed summaries like mass distribution (gravitational potential) and electromagnetic fields (from averaged stellar winds, plasma flows) or chemical abundances (summaries of stellar life cycles). Hence we have Gravity as field summary through which the galaxy experiences the collective “weight” of its stars as a smooth gravitational field. That field is the compressed summary of countless assembly histories, equivalent to how the organism perceives organ function as hormone levels or electrical potentials. The EM is also a field summary. Interstellar plasmas are chaotic at micro-scale (individual charged particles). Yet galaxies exhibit coherent EM fields at kiloparsec scales, guiding cosmic rays and shaping gas flows. Here again, the galaxy only “sees” the smoothed field summary of deeper micro-encodings.

What’s the “beat” of a galaxy then, it’s reflexive loop? In the write phase, each star’s fast activity  $F_i$  and slow assembly record  $(M_i, AI_i)$  inject information into the galactic memory via the feedback current  $j^\mu$ . In the read/experience phase, the galaxy “senses” its own field  $\Phi_{\text{gal}}$  through the motion of gas clouds, orbital dynamics of stars, and propagation of cosmic rays; this constitutes the EI for the galactic organism. In the update phase, the sensed fields feed back into star-formation rates, supernova feedback, and large-scale magnetic amplification, closing the reflexive loop.

We argue that in every observed living system higher-level coordination depends on field-like summaries of nested complexity. If galaxies are themselves organism-like structures, then the way they “use” gravity and EM as smooth fields is *exactly parallel* to how an organism uses hormones, potentials, or morphogen gradients.

## Reflection

In the history of physics, gravity and electromagnetism were treated as fundamental givens, equations written onto a blank stage of spacetime. Quantum mechanics added its own layer of mystery, insisting that nature somehow needs an “observer” to decide.

But in the **superorganism model**, none of this is primitive. **Fields** are not axioms but rather the **smoothing of detail**, the way a parent organism relates to the internal encodings of its parts. **Gravity** is the coarse geometry of slow assemblies (M, AI, G). **Electromagnetism** is the coarse dynamics of fast flows (F, EI). **Quantum states** are the fine encodings themselves, before they are smoothed into a field description. **Observation** is simply the superorganism’s reflexivity, writing and reading its own state.

In living systems, higher-level organisms never track the raw encoding of their parts. They interact through smoothed summaries of hormones, electrical potentials, gradients.

Galaxies, too, do not track every assembly detail of stars and plasmas. They perceive their gravitational potentials and magnetic fields, the smoothed summaries of immense complexity.

Through the lens of the model’s most extreme, **black holes** demonstrate this principle with brutal clarity. Whatever detailed structure went into forming them, the information from stellar dynamics, nuclear burning, accretion pathways, is erased from the outside view. What seems to remain is a **field summary** like mass, spin, and charge. The black hole’s gravitational field is the most radical smoothing imaginable. Every detail of the organismal encoding is compressed. Whether additional “memory” remains hidden depends on how far the model is extended to incorporate quantum feedback loops (EI/AI) and the associated **information-preserving** mechanisms that contemporary physics is still trying to pin down.

Thus, from cells to organs, from stars to galaxies, and even in black holes, the same law holds: **fields are the smoothed face of structure, the summary an organism offers to its parent.**

What this tells us is that **the universe and life do not just share metaphors, they share mechanisms.** The way an organism relates to its organs is the same way spacetime relates to its subsystems. So the strange dualities of physics - fields vs. particles, GR vs. QM, local vs. universal - begin to dissolve. They are not different domains to be patched together. They are simply the *two faces of the same reflexive cycle*, encoding and field.

In this light, the universe is not a dead stage where laws play out, but a living structure, continuously writing, reading, compressing, and interpreting itself. **Biology was never the exception - it was the clue.**

## Communication in the hierarchy

We’ve already set the stage for hierarchies by nesting organisms as well as using the fields as the coarse view of the parent about the detailed structure of the

child organism. Let's dive a bit deeper how such communication would take place and the implications the hierarchy has on propagation of the reflexive loop.

What we're interested in is the loop between parent and child, and then from child to parent again. Especially in terms of the model, like fast fields, and the write operation, building of AI and assembly paths, then read operation and measuring EI. But not just from the view of one organism like before, but specifically about the interaction and communication between the hierarchy of organisms.

We'll explore the communication and the reflexive loop first on the case of two levels, the parent and the child, the children being the organs of the parent.

Let's start with a clean and elegant recipe, and then move to include the framework of the model, that sits inside a richer state/update system that enforces conservation and memory.

We begin with *reflexive loop* between a **parent organism**  $P$  and its set of **child organisms (organs)**  $\{O_i\}_{i=1}^N$ . Let's define the terminology:

- $P$  represents parent super-organism (e.g. a whole organism, an organ system, or even a galaxy)
- $O_i$  is a child sub-organism (an organ, cell cluster, etc.)
- $\mathcal{S}(O_i)$  holds fine-grained state of  $O_i$ , all molecular / cellular / structural details
- $\Phi_P(x)$  is a coarse field that  $P$  perceives as the “summary” of its children (e.g. a stress-energy tensor, EM potential...)
- $\mathcal{Q}_\ell^{(P)} : \{\mathcal{S}(O_i)\} \rightarrow \Phi_P(x)$  is a coarse-graining map for  $P$  at resolution scale  $\ell$ .
- $\Psi_{O_i}(t)$  is the encoding (information, quantum state, etc.) that child  $O_i$  writes back to its parent.
- $W_P : \{\Phi_P(x)\} \rightarrow \mathcal{E}_P$  is the “write” operation of the parent, from its field it produces an *encoding*  $\mathcal{E}_P$  (e.g., a neural spike train, a gene-expression program).
- $R_P : \mathcal{E}_P \rightarrow \Phi_P(x)$  is the “read” operation of the parent, it decodes back into a field.

The steps in the loop are as follows:

1. **Parent  $\rightarrow$  children (write)**
  1. Start with the **parent's coarse description**. The parent already holds a field  $\Phi_P(x)$ .
  2. **Encode the field:**  $W_P(\Phi_P) = \mathcal{E}_P$ . This step is fast (e.g., neural firing, hormonal release) and produces an *encoding* that can be sent to all children.
  3. **Broadcast to children:** Each child receives  $\mathcal{E}_P$ . The child interprets it according to its own internal rules.
2. **Child  $\rightarrow$  parent (write)**
  1. **Each child processes the parent's encoding:**  $O_i$  applies a *local* mapping  $F_i(\mathcal{E}_P)$  that transforms the received signal into its own fine-grained update  $\Delta\mathcal{S}(O_i)$
  2. **Update child state:**  $\mathcal{S}_{new}(O_i) = \mathcal{S}_{old}(O_i) + \Delta\mathcal{S}(O_i)$ .



3. **Encode back to parent:** Each child produces an encoding of its updated state  $\Psi_{O_i} = G_i(\mathcal{S}_{new}(O_i))$ , where  $G_i$  is a *write* map that compresses the fine structure into a message.
3. **Parent reads children's messages**
  1. **Collect all child encodings:**  $\{\Psi_{O_i}\}_{i=1}^N$ .
  2. **Coarse-grain into a new field:** Apply the parent's coarse-graining map  $\Phi_P^{new}(x) = \mathcal{Q}_\ell^{(P)}(\{\Psi_{O_i}\}_{i=1}^N)$ . This is precisely the mapping where the parent *cannot* keep all fine details; it only perceives a smoothed field.
  3. **Optional refinement:** If desired, feed  $\Phi_P^{new}$  back through  $W_P$  to generate a new encoding  $\mathcal{E}_P^{new}$ , closing the loop.

**Iterative Reflexive Cycle:** the whole process can be written as an iteration:

$$\begin{aligned}\mathcal{E}_P^{(k)} &= W_P(\Phi_P^{(k)}), \\ \Psi_{O_i}^{(k)} &= G_i(F_i(\mathcal{E}_P^{(k)})), \\ \Phi_P^{(k+1)}(x) &= \mathcal{Q}_\ell^{(P)}(\{\Psi_{O_i}^{(k)}\}_{i=1}^N).\end{aligned}$$

Here  $k$  indexes discrete time steps or “cycles”.

The reflexive cycle defines the fast-slow dynamics. The *fast* part is the encoding/decoding between parent and child (steps 1–2). And the *slow* part is the coarse-graining  $\mathcal{Q}_\ell^{(P)}$ , which integrates over many cycles and produces the emergent field.

Note that we're not limited to two layers only. If  $P$  itself has sub-parents, the same recursion applies: treat each parent as a child of its own parent and use the same maps. The hierarchy is therefore **self-similar**; every level uses the same fast–slow reflexive machinery.

## Temporal dynamics

Let's try to sketch out a mathematically explicit model that takes the above algorithmic definition of the iterative reflexive cycle one step further by replacing the informal recursion with a set of coupled differential-equations and graph-based operators that can be analysed (or simulated) in a single framework.

The essential layers for every organism  $P$  are organs, fast and slow fields. Organs  $\mathcal{A}_P = \{O_i^P\}_{i=1}^{N_P}$  represent discrete functional units (cells, tissues, colonies). Each has an internal state  $x_i^P(t) \in \mathbb{R}^{d_x}$ . Fast fields  $\phi_P(x, t) \in \mathbb{R}^{d_\phi}$  are spatially distributed signals (pheromones, electric potential, etc.) that obey a diffusion-decay PDE. Slow memory  $M_P(t) \in \mathcal{M}_P \subset \mathbb{R}^{d_M}$  holds the structural variables (architecture, caste distribution, vascular network). These are updated only on long time scales.

The key idea is that each layer feeds into the next via linear/non-linear operators that are *self-similar* across scales, i.e., the same functional form can be used for a child organism  $O_i^P$  and for its own parent  $P$ .

We start with the agent dynamics in the fast field. Each organ  $O_i^P$  evolves according to an ODE that depends on its own internal state, on the local value of

the fast field  $\phi_P$ , and on the influence of other organs through a *communication graph*  $\mathcal{G}_P = (V, E)$ .

Let  $L_{\mathcal{G}_P}$  be the graph Laplacian of  $\mathcal{G}_P$ . The **evolution of each organ's state** is governed by

$$\dot{x}_i^P = f(x_i^P, \phi_P(r_i, t), (L_{\mathcal{G}_P} x^P)_i) + \sigma \eta_i(t)$$

where:

- $r_i$  is the physical location of organ  $O_i^P$ ;
- $\phi_P(r_i, t)$  is the field sampled at that point;
- $(L_{\mathcal{G}_P} x^P)_i = \sum_j w_{ij}(x_j^P - x_i^P)$  encodes diffusive coupling on the graph;
- $f$  may be a nonlinear function (e.g., logistic growth, threshold firing);
- $\eta_i(t)$  is Gaussian noise and  $\sigma$  its strength.

This way we capture *fast* interactions within the organism while still allowing for spatially extended signaling via  $\phi_P$ . The Laplacian term is a discrete analogue of diffusion, ensuring that local organ states influence each other through the communication network (e.g., vascular or mycelial pathways).

Next, let's look at the **fast field dynamics**. The field obeys a *reaction-diffusion* equation whose source term aggregates contributions from all organs:

$$\partial_t \phi_P(x, t) = D \Delta \phi_P - \kappa \phi_P + S(\{x_i^P(t)\}_{i=1}^{N_P}),$$

with  $D$  is the diffusion coefficient,  $\kappa$  a decay rate and  $S$  is a *source* operator that maps organ states to field production (e.g., secretion of pheromone proportional to metabolic activity). A convenient choice for  $S$  could be  $S(\{x_i^P\}) = \sum_{i=1}^{N_P} g(x_i^P) \delta(x - r_i)$ , where  $g : \mathbb{R}^{d_x} \rightarrow \mathbb{R}^{d_\phi}$  converts organ activity into field output, and  $\delta$  is the Dirac delta localised at each organ.

Next in the big three is the definition of the **slow memory dynamics**. The slow component  $M_P(t)$  aggregates past experiences and structural changes. It evolves according to a *discrete* update that integrates over many fast cycles:

$$M_P(k+1) = \mathcal{Q}_P(M_P(k), \bar{x}^P(k), \bar{\phi}_P(k))$$

where:

- $k$  indexes *coarse-grained* time steps (e.g., days, generations);
- $\bar{x}^P(k) = \frac{1}{T} \int_{t_k}^{t_{k+1}} x^P(t) dt$  is a time-averaged (coarse-grained) value of the fast variable  $x^P(t)$  and similarly
- $\bar{\phi}_P(k)(x) = \frac{1}{T} \int_{t_k}^{t_{k+1}} \phi_P(x, t) dt$  for the field;

- $\mathcal{Q}_P$  is a *coarse-graining operator* that may perform averaging, thresholding, or structural adaptation (e.g., adding/removing nodes in the communication graph).

The memory update can be written as an ODE with a very small time constant, for example  $\tau_M \dot{M}_P = -M_P + \mathcal{F}(\bar{x}^P, \bar{\phi}_P)$ ,  $\tau_M \gg 1$ .

Let's consider a more general **hierarchical coupling**, not only an organism with organs, but an organism  $P$  that itself contains sub-organisms. Let's define two operators based on the location of their activity. On the parent level, we have an **upward operator**  $\mathcal{U}$ , which aggregates the fields of children  $\phi_P^{\text{in}}(x, t) = \sum_{Q \in \text{sub}(P)} w_Q \phi_Q(x, t)$ , where  $w_Q$  are weighting factors (e.g., proportional to the size of child  $Q$ ). The children receive a coarse-grained field from the parent via **downward operator**  $\mathcal{D}$  and inject them into their own:  $\phi_Q^{\text{out}}(x, t) = \mathcal{D}(M_P(t), \phi_P(x, t))$ . These operators are chosen to preserve the *self-similar* structure: if  $P$  is a single organ,  $\mathcal{D}$  and  $\mathcal{U}$  reduce to identity.

Collecting it all for every organism in the hierarchy yields a **coupled multi-scale dynamical system**:

$$\begin{cases} \dot{x}_i^P = f(x_i^P, \phi_P(r_i, t), (L_{\mathcal{G}_P} x^P)_i) + \sigma \eta_i(t), \\ \partial_t \phi_P(x, t) = D \Delta \phi_P - \kappa \phi_P + \sum_{j=1}^{N_P} g(x_j^P) \delta(x - r_j), \\ \tau_M \dot{M}_P = -M_P + \mathcal{F}(\bar{x}^P, \bar{\phi}_P), \end{cases} \quad \text{with hierarchical coupling via } \mathcal{D}, \mathcal{U}.$$

Let's investigate the definitions a bit. We defined *spatial realism* through diffusion equation for fields rather than abstract “encoding”. *Graph coupling* captures non-local interactions without resorting to continuous averaging. The scalar  $\tau_M > 0$  is a **time-constant** (or *memory lag*) that determines how rapidly the slow memory can respond to changes in the fast variables, or,  $\tau_M$  is the *characteristic time* over which slow memory  $M_P$  integrates fast fluctuations. Its meaning and practical implications are:

- **Separation of scales.** If  $\tau_M \gg T$  (the period over which we average the fast dynamics), the memory updates only after many fast cycles, guaranteeing that it reflects a *coarse* or *integrated* view. This is exactly what a coarse-graining map  $\mathcal{Q}_\ell$  would produce in a discrete-time formulation.
- **Rate of adaptation.** A small  $\tau_M$  means the organism can quickly incorporate new information (high plasticity), whereas a large  $\tau_M$  corresponds to a more *rigid* memory that changes slowly. For example, an old tree whose trunk structure is largely immutable.
- **Stability vs. responsiveness.** In control-theoretic terms,  $\tau_M$  plays the role of a low-pass filter. It damps high-frequency fluctuations in  $x^P, \phi_P$ . Too small and the memory will chase noise, while too large and the system will be sluggish or even unstable if the fast dynamics drive rapid changes.
- **Relation to biological time scales.** In a cell,  $\tau_M$  might correspond to transcription-translation turnover (hours). In an organism it could map onto developmental timescales (days–years). For cosmological structures,  $\tau_M$  could be of the order of millions of years, reflecting the slow evolution of stellar populations or galactic morphology.

- **Mathematical convenience.** By writing the memory update as a first-order linear ODE with constant coefficient  $\tau_M^{-1}$ , we can solve it analytically (exponential decay) and embed it in larger numerical schemes without extra stiffness.

## Operators

Let’s review the above not from the point of definition of fields but the flow of information. In a multi-scale system the *communication* between levels is usually formalised with two complementary operators that carry information down from a higher-level “parent” to its lower-level “children”, and up from children back to the parent. The idea is that each level can only see a coarse, smoothed representation of what happens below it, while still being able to influence those sub-systems through a field that encapsulates the parent’s state.

Let’s explain a bit more about the **upward** and **downward** operators. The naming convention follows the **hierarchical position of the operator**, not the direction of information flow. That is, **upward** and **downward** refer not to the *direction* in which the signal physically travels, but to the *side of the hierarchy that owns or initiates the operation*:

### $\mathcal{U}$ – Upward operator

- Acts at a *parent* level.
- Takes the children’s local encodings  $\{\Psi_{O_i}\}$ , aggregates them via  $\mathcal{Q}_\ell^{(P)}$ , and produces a new coarse field  $\Phi_P^{\text{new}}$ .
- The resulting field is then *sent down* to the children.
- Hence “upward” refers to the operator’s origin (parent) moving up the hierarchy, not the direction of its output.

### $\mathcal{D}$ – Downward operator

- Acts at a *child* level.
- Takes the parent’s field  $\Phi_P$ , possibly weighted or filtered, and injects it as an external source into the child’s dynamics.
- The field is therefore moving **down** from parent to child, so we call this operator “downward”.

In short:

```

Parent (higher level)
|
U  <-- upward operator: aggregates children {\textrightarrow} new parent field
|
Child (lower level)
|
D  <-- downward operator: pushes parent field into child dynamics

```

The parent *creates* and *sends* the field  $\rightarrow$  this is what the parent calls “writing” or using the **upward** operator. The child *takes* that field and *injects* it into its dynamics  $\rightarrow$  this is what the child calls “applying the downward operator”. The **same field** is produced by  $\mathcal{U}$  at the parent side and consumed via  $\mathcal{D}$  at the child side. The naming convention simply keeps track of *who owns* each step in the reflexive cycle.

### Downward coupling (parent to children)

The parent sends its *coarse-grained* field to its children via an **downward operator**  $\mathcal{D}$ .

The *downward* operator  $\mathcal{D}$  thus supplies a field that each child  $O_i$  receives from its immediate parent  $P$ :

$$\phi_{O_i}^{\text{in}}(x, t) = \mathcal{D}(\Phi_P(x, t)) = w_P \Phi_P(x, t),$$

where  $w_P$  is a weight (e.g., proportional to the child’s size or coupling strength). In practice, this term appears in the child’s field equation:

$$\partial_t \phi_{O_i}(x, t) = D \Delta \phi_{O_i} - \kappa \phi_{O_i} + \sum_{j=1}^{N_{O_i}} g(x_j^{O_i}) \delta(x - r_j) + \phi_{O_i}^{\text{in}}(x, t).$$

Thus the downward operator acts as an *external source* that is added to the field dynamics of every child. We denote the field generated by the downward operator as  $\phi_Q^{\text{in}}$  when we view it from the child’s perspective, and equivalently as  $\phi_Q^{\text{out}}$  when we treat it as the output of the parent’s downward step. The two symbols refer to the same quantity:  $\phi_Q^{\text{in}} = \phi_Q^{\text{out}}$ .

### Upward coupling (children to parent)

After a child has produced its encoding  $\Psi_{O_i}$  (which depends on its own fast state), the parent aggregates all of them into a new field:

$$\Phi_P^{\text{new}}(x) = \mathcal{U}(\{\Psi_{O_i}\}_{i=1}^{N_P}) = \mathcal{Q}_\ell^{(P)}(\{\Psi_{O_i}\}).$$

Equation is simply the third line of the reflexive cycle, now labeled as  $\mathcal{U}$ . It becomes a *boundary condition* for the parent’s field equation:

$$\partial_t \phi_P(x, t) = D \Delta \phi_P - \kappa \phi_P + \sum_{j=1}^{N_P} g(x_j^P) \delta(x - r_j) + \underbrace{(\Phi_P^{\text{new}}(x) - \Phi_P(x, t))}_{\text{upward feedback term}}.$$

In many implementations the upward term simply *replaces* the old field value at discrete update steps (i.e.,  $\phi_P \leftarrow \Phi_P^{\text{new}}$ ). The key point is that  $\mathcal{U}$  injects the children’s coarse information back into the parent’s dynamics.

### Coupled multi-scale dynamics

When you stack these operators across many levels, you obtain a **coupled multi-scale dynamical system**:

$$\dot{\phi}_P = F_P(\phi_P, \phi_P^{\text{in}}, M_P),$$

$$\dot{\phi}_Q = F_Q(\phi_Q, \phi_Q^{\text{out}}, M_Q),$$

with  $F_P, F_Q$  being the intrinsic dynamics of each level. The hierarchy thus becomes a reflexive loop: the parent’s field shapes its children, while the aggregated feedback from all children reshapes the parent.

## So smooth

When it comes to smoothing, we need to clarify something first. It might be mistakenly interpreted, making the field as a passive mathematical construct imposed from outside, that the smoothing is just an arbitrary averaging. That is not the case. The smoothing is *a choice made by the system itself*.

**The parent organism actively decides which features of its child’s micro-encoding to keep and which to discard. The parent organism selects those micro-patterns that are predictive or controllable at its own scale and encodes them as a continuous field.**

This decision can be understood as an *information-theoretic compression* that preserves what is functionally relevant for higher-level control, while throwing away noise or redundancy. It mirrors how a brain uses attention: it “smooths” the flood of sensory data into a manageable set of variables (e.g., the location of an object) and then feeds that back to the body.

The *choice* of which details to keep and which to discard is part of what defines a *coarse-graining* operator  $\mathcal{Q}_\ell$ , so it **is** a decision made by the parent organism (or its governing dynamics), leading to something like  $\Phi(x) = \mathcal{Q}_\ell(\mathcal{S}(\mathcal{O}_i))$ .

For example, in a living system the *mechanism* is guided by functional needs: “I only care about hormone concentration, not individual secretion events.” That intent is encoded in how  $\mathcal{Q}_\ell$  is built (e.g., a low-pass filter that ignores high-frequency spikes).

So while the equations look like simple averages, they’re really shorthand for a *policy* that the parent organism follows to decide which aspects of its child’s micro-encoding matter at that level.

## Cycles

Let’s make an explicit algorithmic description of a reflexive cycle within a hierarchy using the upward and downward operators. It ties neatly the dynamics of a fast-slow cycle to include the hierarchical flow of information. An iteration of the fast-slow reflexive loop can be defined via operators as follows:

1. **Parent aggregates children via  $\mathcal{U}$ .** The parent collects all child encodings  $\{\Psi_{\mathcal{O}_i}\}$  and applies its coarse-graining map  $\Phi_P^{\text{new}}(x, t) = \mathcal{U}(\{\Psi_{\mathcal{O}_i}\}) = \mathcal{Q}_\ell^{(P)}(\{\Psi_{\mathcal{O}_i}\})$ .
2. **Parent pushes the new field down.** Each **child receives the updated parent field through  $\mathcal{D}$**  as an external input:  $\phi_{\mathcal{O}_i}^{\text{in}}(x, t) = \mathcal{D}_i(\Phi_P^{\text{new}}) = \sum_{Q \in \text{sub}(P)} w_Q \Phi_P^{\text{new}}$ .
3. **Child follows the guidance,** the received field  $\phi_{\mathcal{O}_i}^{\text{in}}$  enters its fast dynamics.
4. **Child’s trajectory generates new data** like sensor readings, actions, currents. These are written into the child’s local memory, the “record” part of the EI.
5. **Child writes its record to the parent.** The *write* operation that the parent will later read. In practice this means each child contributes a term

$\Psi_{O_i}$  back to the pool that the next  $\mathcal{U}$  call will aggregate.

The focus of such description of the cycle however is all about reading the values of the fields and processing them. It is the view from the point of the experience itself.

Why the distinction? Because the cycles can be interpreted from different perspectives. In the case of the structure creating cycle, we can talk about spacetime, and thus time itself. And in the case of the experience processing cycle, we can focus on what it means to experience, to define how we shape the structure and why.

## Hierarchical time

The property of “time” has already arisen throughout the model at various occasions. We’ve already seen how it affects the fast-slow cycle, or coarse-graining between hierarchical structures. In this section we explore the properties of time in the framework of the reflexive organism.

Let’s first look at time outside the framework. A *clock* is a physical system whose internal state changes at a regular rate (e.g., atomic transition). The rate of change of the clock’s internal state is determined by physical constants such as  $c$  or  $\hbar$  (constants are not themselves time rates). *Proper time* is determined solely by the spacetime metric  $g_{\mu\nu}$ :  $d\tau^2 = -g_{\mu\nu}dx^\mu dx^\nu / c^2$ . It is a *geometric* quantity, independent of any internal dynamics. An *arrow of time* is usually given by entropy increase (second law) or the cosmological initial condition. No intrinsic link to the *activity* of the system that experiences it.

In this view, *time is an external scaffold* on which physical processes unfold. The scaffold itself does not change in response to what happens within it. However, the reflexive model reshapes the understanding of time.

The *rate* at which a system can update its slow memory  $M$  becomes an intrinsic dynamical variable. Proper time for the organism is now a function of  $\tau_M$ . EI quantifies how strongly a system’s fast dynamics are fed back into its own memory. A high EI means that the system *feels* its own state, thereby altering its effective metric  $g_{\mu\nu}^{(\ell)}$  through  $\mathcal{Q}_\ell(M, \text{AI})$ . The flow of time is therefore self-modulated by how much the organism has learned about itself. Proper time emerges from a closed loop: fast dynamics generate a write to memory, which in turn reshapes the slow field that drives the next fast step. This cyclical process can be formally described as a *delay differential equation* for  $g_{\mu\nu}$  and hence for  $\tau$ . Correlations between spatially separated parts are maintained because their joint record resides in the same memory. Thus, entangled subsystems share a *common time reference* that is not simply dictated by spacetime geometry but by the *shared assembly history*.

Furthermore, reflexive model introduces a hierarchy of organisms, each with their own memory timescale and makes the whole concept of time even more interesting. **Time is not a single universal parameter** but a *hierarchically nested sequence of local proper times*, each governed by the memory dynamics of that level. Let’s dive into this fascinating property.

First, let us write down the core equations for one organism  $O$  (this could be any member of the hierarchy; the notation is identical).

- **Fast dynamics**  $\dot{x}_F^\mu(t) = v^\mu(x_F, j^\nu(t))$ , where  $x_F$  are the fast variables (e.g., charge densities, field amplitudes) and  $j^\nu$  are the read-back currents that feed into EI.
- **Memory update**  $\dot{M}(t) = -\frac{1}{\tau_M}(M(t) - M_{\text{eq}}[x_F])$ . The equilibrium  $M_{\text{eq}}$  is a functional of the fast variables; it captures how the organism “stores” the recent history of its internal dynamics.
- **Emergent metric**  $g_{\mu\nu}^{(\ell)}(x, t) = \mathcal{Q}_\ell(M(t), AI(t))$ . The map  $\mathcal{Q}_\ell$  is the organism’s *self-reflection*, it interprets the current state of memory and assembly as a coarse-grained geometry. Importantly,  $g_{\mu\nu}^{(\ell)}$  is *time-dependent* because both  $M(t)$  and  $AI(t)$  evolve.
- **Proper time along an organism’s worldline**  $d\tau_O = \frac{1}{c} \sqrt{-g_{\mu\nu}^{(\ell)}(x, t) dx^\mu dx^\nu}$ . Since  $g_{\mu\nu}^{(\ell)}$  varies on the timescale  $\tau_M$ , the rate at which proper time elapses for  $O$  is modulated by how fast its memory can update.

Now, consider a **parent organism**  $P$  and one of its **child organs**  $O_i$ . Each has its own reflexive loop, but they are coupled in two ways. One, the organ’s fast dynamics generate *read-back currents*  $j_i^\mu(t)$ . These currents are summed over all organs and fed into the parent’s memory  $M_P$ . Thus, the parent “writes” its geometry from the aggregate of its children. Second, the parent’s emergent metric  $g_{\mu\nu}^{(\ell)}(P)$ , updated on a slower timescale  $\tau_{M,P}$ , is broadcast to each organ as part of the *slow* background field that modulates their fast dynamics. In effect, the child reads its environment from the parent’s coarse-grained geometry.

Mathematically we write:

$$\begin{aligned}\dot{M}_P(t) &= -\frac{1}{\tau_{M,P}} (M_P - M_{\text{eq}}[j_i^\mu]), \\ g_{\mu\nu}^{(\ell)}(P, t) &= \mathcal{Q}_\ell(M_P(t), AI_P(t)), \\ \dot{x}_{i,F}^\alpha(t) &= v^\alpha(x_{i,F}, j_i^\beta, g_{\mu\nu}^{(\ell)}(P, t)).\end{aligned}$$

The *time* that each organ experiences is thus a **function of the parent’s memory** and vice versa. Because  $\tau_{M,P}$  is usually much larger than  $\tau_{M,i}$ , the parent’s proper time drifts slowly relative to its organs, which can adapt rapidly.

## Rewards

Let’s address another mechanism which is usually considered external to the organism, the subject of rewards. And just like so many things so far, we’re going to show that the reflexive model turns an external property into an intrinsic one.

We’ve already talked about the experience through reflection process  $I \xrightarrow{\Sigma} S \xrightarrow{\Pi} A \xrightarrow{\Upsilon} O \xrightarrow{W} M \xrightarrow{R} S$  (then back to  $\Pi$ ). And defined the experience index EI for it. We later corrected it for world-coupling and prediction calibration into  $\text{EI}^*$ . We’ve also already defined the purpose function  $\mathcal{J} = (\text{AI})^\alpha \cdot (\text{EI}^*)^\beta$ , where the organism defines how much it favors building and experience. However, experiences and assemblies vary. Some are more “rewarding” than others (we’ve



marked this with a scalar, the higher it is for EI, the more “rewarding” the experience). What it means to be “rewarding” is the theme of this section.

Rewards are mostly seen as external feedback, imposed as optimization criterion. For example, in RL a reward is external scalar signal  $R_t$  used to train a policy  $\Pi^* = \arg \max_{\Pi} \mathbb{E} \left[ \sum_t \gamma^t R_t \right]$ .

In the reflexive organism model, the “reward” is not external, it’s **how much the loop is closed**. So instead of  $R_t$ , we use the **instantaneous experience index**  $EI_t$ :

$$\Pi^* = \arg \max_{\Pi} \mathbb{E}_{\Pi} \left[ \sum_{t=0}^{\infty} \gamma^t EI_t \right].$$

where

$$EI_t = \underbrace{\frac{I(M_{t-\tau}; S_t \mid E_{t-\tau:t})}{H(S_t)}}_{\text{Memory} \rightarrow \text{sensing}} \cdot \underbrace{\frac{TE(M_{t-\tau} \rightarrow \Pi_t \mid E_{t-\tau:t})}{C_{\Pi}}}_{\text{Memory} \rightarrow \text{policy}} \cdot \underbrace{\frac{I(O_t; \hat{O}_t)}{H(O_t)}}_{\text{Prediction groundedness}} \cdot \underbrace{\rho_t}_{\text{reuse/coverage}}.$$

So the reward isn’t a number handed down externally, it’s calculated internally from **information flows** between memory, sensing, policy, and outcomes. This is the **EI-RL principle** to optimize for trajectories that maximize reflexive experience over time.

However, this is only one part of the loop. We need to include AI as well. This could produce a **log-product** reward metric, something like:

$$r_t = \underbrace{\alpha \log(AI_t + \epsilon)}_{\text{build}} + \underbrace{\beta \log(EI_t^* + \epsilon)}_{\text{experience}} \Rightarrow \max_{\Pi} \mathbb{E}_{\Pi} \left[ \sum_t \gamma^t r_t \right].$$

Or **power-product** like  $R_{\text{episode}} = \prod_t (AI_t^{\alpha} (EI_t^*)^{\beta})^{\gamma^t}$ .

An example of reward for practical drop-in is presented in  $r_t$  as  $r_t = \alpha \log(AI_t + \epsilon) + \beta \log(EI_t \cdot \frac{I(S_t; E_{t-\tau:t})}{H(S_t)} \cdot (1 - \frac{D_{\text{KL}}(p(O_t) \parallel p(\hat{O}_t))}{H(O_t)}) + \epsilon) - \lambda c_t$ .

## Rewards through hierarchy

Let’s look at the propagation of rewards through hierarchy now. We’ll look at the coupling metric that measures what a “good” hierarchy means. We’ll base it on the upward and downward operators introduced earlier. Let’s give an overview how to define hierarchical rewards and then let’s dive into each coupling measure to better understand what we’re trying to achieve.

The intent is to reuse the upward  $\mathcal{U}$  and downward  $\mathcal{D}$  operators that couple parents and children, and to reward *both* the unit’s internal reflexivity (AI/EI) and (the **quality of coupling** along  $\mathcal{U}, \mathcal{D}$ . The operators (and the coupled dynamics) are already defined as

$$\begin{aligned}\phi_{\mathcal{O}_i}^{\text{out}} &= \mathcal{U}(M_P, \phi_P), \quad \phi_P^{\text{in}} = \sum_{\mathcal{O}_i \in \mathcal{O}} w_{\mathcal{O}_i} \phi_{\mathcal{O}_i}, \\ \dot{\phi}_P &= F_P(\phi_P, \phi_P^{\text{in}}, M_P), \quad \dot{\phi}_{\mathcal{O}_i} = F_{\mathcal{O}_i}(\phi_{\mathcal{O}_i}, \phi_{\mathcal{O}_i}^{\text{out}}, M_{\mathcal{O}_i}).\end{aligned}$$

These realize the “parent shapes child, children reshape parent” reflexive loop. We’ll now define coupling quality at each level.

Looking at **downward** responsiveness from the parent to a child, what we’re interested in is whether the parent’s broadcast of  $\phi_P^{\text{out}}$  actually enter the child’s sensing and *causally* shape its policy. Thus the **downward quality** is defined as

$$C_{\downarrow}(\mathcal{O}_i | P)_t = \frac{I(S_{\mathcal{O}_i, t}; \phi_{P, t}^{\text{out}})}{H(S_{\mathcal{O}_i, t})} \cdot \frac{TE(\phi_{P, t-\tau:t}^{\text{out}} \rightarrow \Pi_{\mathcal{O}_i, t})}{\mathcal{C}_{\Pi_{\mathcal{O}_i}}}$$

The first factor measures readability, it deals with how much the children actually see the parent. The second factor measures usefulness and is all about how much the parent changes the child’s decisions. So  $C_{\downarrow}$  is a **dimensionless quality score** for the *downward* interface.

Now, let’s do the other way around and look at **upward** fidelity, that is, do the children’s aggregated messages match the parent’s predicted/desired in-field and inform the parent’s policy. The **upward quality** is thus defined as

$$C_{\uparrow}(P | \mathcal{O})_t = \left(1 - \frac{D(\phi_{P, t}^{\text{in}} \parallel \hat{\phi}_{P, t}^{\text{pred}})}{Z}\right) \cdot \frac{TE(\{\phi_{Q_j, t-\tau:t}\} \rightarrow \Pi_{P, t})}{\mathcal{C}_{\Pi_P}}$$

The first factor is about model calibration, it deals with how much the children match what the parent thinks they will send.  $Z$  is the normalization constant or the upper bound, something like  $Z = \max_t D(\phi_{P, t}^{\text{in}} \parallel \hat{\phi}_{P, t}^{\text{pred}})$  and  $\mathcal{C}_{\Pi_P}$  is the policy execution cost. The second factor is about casual usefulness and deals with how much the children’s past messages cause the parent’s decisions. In other words,  $C_{\uparrow}$  is the score for “*My children tell me exactly what I expected and what they tell me actually changes what I do.*”

Now let’s define local rewards that use AI, EI as well as coupling quality. We need to define them at each level. A reward for the child  $Q$  under parent  $P$  is defined as follows:

$$r_{\mathcal{O}_i, t} = \alpha_{\ell} \log(AI_{\mathcal{O}_i, t} + \varepsilon) + \beta_{\ell} \log(EI_{\mathcal{O}_i, t} + \varepsilon) + \kappa_{\ell} \log(\mathcal{C}_{\downarrow}(\mathcal{O}_i | P)_t + \varepsilon) - \lambda_{\ell} c_{\mathcal{O}_i, t}$$

Let’s look at the parent  $P$  over children  $\mathcal{O}$ :

$$\begin{aligned}r_{P, t} &= \alpha_{\ell+1} \log(AI_{P, t} + \varepsilon) + \beta_{\ell+1} \log(EI_{P, t} + \varepsilon) + \\ \xi_{\ell} \sum_{\mathcal{O}_j \in \mathcal{O}} w_j \log(EI_{\mathcal{O}_j, t} + \varepsilon) &+ \mu_{\ell} \log(C_{\uparrow}(P | \mathcal{O})_t + \varepsilon) - \lambda_{\ell+1} c_{P, t}\end{aligned}$$

So the parent is rewarded for its own purpose, for his children being reflexive (a log-mean), and for **faithful, policy-relevant aggregation** from children. Which enables us now to define a **multi-level objective**:

$$\max_{\{\Pi\}} \mathbb{E} \left[ \sum_{t=0}^{\infty} \gamma^t \left( \sum_{\ell=0}^L W_{\ell} \sum_{k \in \mathcal{S}^{(\ell)}} r_{k,t}^{(\ell)} \right) \right]$$

Note that this recurses cleanly because  $\mathcal{U}, \mathcal{D}$  are defined at every parent/child boundary, and the fast-slow, write-read cycles already span levels. Let’s look at some outcomes that these rewards offer.

For example, improvements at a child that do not propagate causally upward (low  $C_{\uparrow}$ ) cannot inflate the parent’s return. Likewise, a strong parent broadcast that the child doesn’t read/use (low  $C_{\downarrow}$ ) doesn’t inflate the child’s return. Reward is effectively *conserved* against spurious correlation. Only *used information* moves value across an edge.

What we also get is global optimality from local improvements (under mild conditions). If we have edges form a tree or have sparse, well-conditioned cycles; as well as  $C_{\uparrow}, C_{\downarrow} \in (0, 1]$ ; and also have the log weights are positive, then local ascent on each  $r_k$  while improving its incident edge terms yields a monotone increase of the global objective. In other words, the log-sum makes the objective subadditive for bad links and supermultiplicative when all links improve together and hence the “fixing weak links” is globally correct.

Because everything aggregates through logs, the hierarchy optimizes a geometric mean of EI across units/edges. A chronically low-EI child *drags down* the parent (and siblings) more than a high-EI child can compensate. This creates a *built-in incentive to lift weak parts* (repair, educate, allocate bandwidth). A precondition for an emergent fairness/robustness property, not added by hand.

The system also defines *experience is the currency*. AI (structure) increases long-term capacity, but only EI (closed-loop use) *mints reward*. With multiplicative/log aggregation, unused AI behaves like *idle capital*. It adds little to the geometric growth rate until it is *spent* (read back) into EI. This explains why systems that only build (accumulate structure) but do not *use* it stall in global return.

What we expect to see happen is increased coherence. As coupling improves (edge flows rise), the hierarchy moves from *thermalized* (high entropy, low EI, weak edges) to *coherent* (low divergence, high TE, strong reuse). The transition is visible directly in reward flow as  $f_{P \rightarrow Q}$  and  $f_{Q \rightarrow P}$  surge together when the loop closes, marking a reflexive phase where global return jumps.

Edges with low downward  $C_{\downarrow}$  and low upward  $C_{\uparrow}$  carry near-zero reward contribution as well as near-zero gradient. With any ordinary cost/regularization on signaling, their parameters drift toward zero, so those interfaces *fade out*. Edges that keep both  $C_{\downarrow}$  and  $C_{\uparrow}$  high sustain reward and non-zero gradients and hence they *stabilize*. The surviving clusters become *modules*.

In this hierarchy, an edge is the *interface contract* between a parent node  $P$  and a child node  $\mathcal{O}_i$ . It isn’t just a line in a graph, it’s a pair of directional channels with

dynamics, noise, costs, and delays. And the two (downward, upward) qualities summarize whether the edge is “alive”. Its “goodness” is not just connectivity, it’s **how much information is read and causally used** in each direction. That’s why reward “flows” along edges and why dead edges (unread/unused) naturally atrophy while informative ones stabilize into modules.

The dynamics at the edges might help explain why certain species go extinct. The interfaces between the species and their environment fade out, most likely because the environments stops sending signals that the species can read, and hence the downward quality is low. The loop starts to break and slowly fades out. Reframing the equations, we can describe the conditions for extinction: **species goes extinct when it stops being a closed loop with its world**. As environmental structure shifts, the downward (reading the world) and upward (being effective in the world) qualities slide toward zero and the **edge carries no reward flow**. As such, the species cannot convert energy into *experienced, useful* action fast enough to cover costs. That is **extinction as loss of reflexive coupling**.

## Extinction as loss of reflexive coupling

Let’s add two definitions. First, a general form for a **local purpose at the species** (per step)

$$r_{S,t} = \alpha \log(AI_{S,t} + \varepsilon) + \beta \log(EI_{S,t} + \varepsilon) + \kappa \log(C_{\downarrow,t} + \varepsilon) + \mu \log(C_{\uparrow,t} + \varepsilon) - \lambda c_{S,t}.$$

Next, something that spans over time, like **reflexive viability (long-run return)**:

$$\bar{V}_S \equiv \liminf_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[r_{S,t}].$$

Assume costs  $c_{S,t}$  are bounded below by  $c_{\min} \geq 0$ , and  $AI, EI$  are bounded above (physically reasonable). If along a time window of positive density,  $C_{\downarrow,t} \rightarrow 0$  and  $C_{\uparrow,t} \rightarrow 0$ , then  $\bar{V}_S = -\infty$  or  $\bar{V}_S < 0$ , hence any demographic model that couples growth to  $\bar{V}_S$  (e.g.,  $\dot{N} = N(\bar{V}_S - \delta)$ ) enters a decline basin.

**Proof sketch.** As  $C_{\downarrow}, C_{\uparrow} \rightarrow 0$ ,  $\log(C_{\downarrow} + \varepsilon), \log(C_{\uparrow} + \varepsilon) \rightarrow -\infty$ . Because  $AI, EI$  are bounded while costs are non-negative, the negative edge terms dominate the sum, forcing the time-average below zero.

## Altruism vs. selfishness

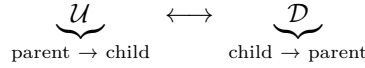
Note that in the previous section we were only interested in rewards that deal with both parts of the processes, assembly and experience. Rewards could have been defined any other way, for example to maximize stability with slow memory, or probability through fast fields, or maximizing correctness of predictions of actions, etc. In the end, we’ve reframed the rewards in terms of flows of quality

of coupling between layers. Let’s now explore these qualities from a different perspective.

Given the hierarchies of organism, the AI-EI loops, and the rewards that flow between hierarchies, one would easily conclude that the optimal state of the system is such that the parent provides the condition for children to grow while the children grow in a way that reinforces the parent. Namely, that the optimal strategy for the interconnected organisms is altruism. Let’s explore if that’s actually the case.

Let’s set the ground right from the start. Altruism is not a moral choice made by an isolated agent. It is a *structural property* of a multi-level system that guarantees **mutual reinforcement of experience (EI) and assembly (AI)** across its hierarchy.

In the reflexive model this manifests as a bidirectional, self-referential loop:



where  $\mathcal{U}$  aggregates the parent’s coarse state into a field that is then injected into each child through  $\mathcal{D}$ . Let’s make a formal definition for altruism now. We’ve already provided the foundations, so let  $P$  be a parent level,  $O_1, \dots, O_k$  its children,  $AI_P$ ,  $EI_P$  and  $AI_{O_i}$ ,  $EI_{O_i}$  their respective indices,  $\Pi_P, \Pi_{O_i}$  the policies that map perceptions to actions.

**Altruism** holds iff for every parent–child pair  $(P, O_i)$  the following holds:

1. **Parent-driven support.** The child’s reward function contains a term proportional to the parent’s EI:  $R_{C_i}(t) = EI_{C_i,t} + \lambda EI_{P,t}$ ,  $\lambda > 0$ .
2. **Child-driven reinforcement.** Conversely, the parent’s reward includes the child’s EI:  $R_P(t) = EI_{P,t} + \mu \sum_{i=1}^k EI_{C_i,t}$ ,  $\mu > 0$ .
3. **Purpose weighting.** Both levels use a purpose function that explicitly weights the other’s experience:  $\mathcal{J}_P = (AI_P)^\alpha (\beta EI_C^* + EI_P^*)^\gamma$ ,  $\mathcal{J}_{C_i} = (AI_{C_i})^\delta (\epsilon EI_P^* + EI_{C_i}^*)^\zeta$ .
4. **Coupled dynamics.** The policies are updated to maximise the *expected discounted sum* of these rewards:  $\Pi_P^* = \arg \max_{\Pi} \mathbb{E}_{\Pi} [\sum_{t=0}^{\infty} \gamma^t R_P(t)]$ ,  $\Pi_{C_i}^* = \arg \max_{\Pi} \mathbb{E}_{\Pi} [\sum_{t=0}^{\infty} \gamma^t R_{C_i}(t)]$ .

When the above conditions hold, each level is *altruistically* tuned. It sacrifices (or reallocates) part of its own immediate benefit to enhance the other’s experience. Intuitively, from **parent**  $\rightarrow$  **child** perspective, a colony’s queen is allocating resources (nutrients, pheromones) to brood. The child receives a coarse-grained field that tells it *how much* to develop and injects it via  $\mathcal{D}$ . Its own reward includes the queen’s overall success. From **child**  $\rightarrow$  **parent** perspective, as larvae grow and start foraging or defending, they send back signals (e.g., pheromone trails). The parent aggregates these via  $\mathcal{U}$ , updating its global policy. Its reward now contains a term that rewards the *collective* success of its sub-units.

We can see the effect of altruism to provide **robustness** where mutual reinforcement dampens catastrophic failures in one layer by leveraging the other. Or for example, it allows for **evolutionary fitness**. We’ll dive more into the latter now.

Within the reflexive AI–EI framework **mutual altruistic coupling is mathematically the most stable attractor** for a hierarchical system that seeks to maximise long-term experience and assembly. In practice, however, pure altruism is rarely *the* only viable strategy. Evolutionary dynamics allow a spectrum of behaviours (selfishness, spite, conditional cooperation) depending on constraints such as resource scarcity, communication costs, or the cost of maintaining the coupling mechanisms.

In other words, while altruism is mathematically the *optimal* configuration for maximizing long-term experience and assembly, real systems often inhabit a **trade-off space** where partial or conditional altruism coexists with other strategies.

Let’s explore **selfishness** now. Let’s reason why selfishness would emerge instead of altruism. In terms of cost of maintaining the coupling, the upward/downward operators  $\mathcal{U}$  and  $\mathcal{D}$  incur energy, bandwidth or computation costs ( $C_{\uparrow}, C_{\downarrow}$ ). If a child can *bypass* the parent’s guidance for a short time, it saves on these costs. Another case is that the parent may not have perfect knowledge of the child’s local environment. A selfish child that directly senses and acts locally can exploit information that the parent would otherwise miss. Yet another case arises because rewards are discounted ( $\gamma < 1$ ), a strategy that maximises immediate EI (by exploiting resources or shortcuts) can out-compete a slower, altruistic one in the short run and exploit this temporal discounting. Due to difference in cadences, in highly volatile settings, waiting for the parent’s signal may mean missing an opportunity; selfishness offers a quick payoff.

Hence, *selfishness arises when local costs of maintaining the reflexive coupling outweigh the immediate benefit of following the parent’s guidance.*

Sustained selfish behaviour is counter-productive because it undermines the very coupling that gives the system its long-term advantage. For example, during persistent exploitation, over-reliance on local shortcuts depletes shared resources, leading to a drop in overall EI and AI for all levels. Another one would be avoiding high coupling costs. If maintaining  $\mathcal{U} / \mathcal{D}$  is expensive, repeated selfish acts may drive the net reward below zero. Yet another case would be information decay, where selfish actions that ignore long-term memory can produce noise in the feedback loop, destabilising learning.

However, selfishness through transient exploitation can also be beneficial at the system level. It might allow **rapid adaptation**. A selfish child that quickly exploits a local resource spike can raise its EI, which then feeds back (via  $\mathcal{U}$ ) to inform the parent of a new high-value state. The parent can then re-allocate resources more effectively. Should the parent’s coupling mechanism fail (e.g., a communication link drops), selfish sub-units keep functioning, maintaining a baseline EI for the organism until repair. This is analogous to biological redundancy. In an evolutionary context, selfish mutants may arise by mutation; if they confer higher fitness in some environments, natural selection can temporarily favour them. The system as a whole then integrates those beneficial mutations into its repertoire.

But most of all, selfish behaviour is essentially *exploration*, trying out new actions without waiting for the collective signal. If successful, it expands the set of useful policies  $\Pi$  that the whole system can later exploit altruistically.

This means it's not just black or white. A healthy reflexive organism will balance altruistic cooperation with occasional selfish exploitation, much like biological organisms exhibit both cooperative and opportunistic behaviours depending on context.

Let's look at what it would mean to **balance** the two. An altruistic policy maintains low coupling costs ( $C_{\uparrow}, C_{\downarrow}$ ) and ensures long-term, stable growth of AI and EI across layers. Selfish policy though provides rapid, local gains in EI by bypassing the parent's guidance, but also acts as exploration that can escape sub-optimal attractors.

This provides the basis for the formal definition of the balance point. Let us denote  $R_A$  is expected discounted reward from **altruistic** behaviour (i.e., following the parent's field),  $R_S$  is expected discounted reward from **selfish** behaviour (ignoring the parent for a short horizon) and  $\lambda(t) \in [0, 1]$  is *policy mixing coefficient*, the fraction of time the system follows altruistic guidance.

The overall expected reward is

$$\mathbb{E}[R] = \lambda R_A + (1 - \lambda) R_S.$$

And the **balance point**  $\lambda^*$  is obtained by solving

$$\lambda^* = \arg \max_{\lambda \in [0, 1]} [\lambda R_A + (1 - \lambda) R_S], \quad \text{s.t. } \sigma_{\text{coll}} + \mathcal{B} \geq \Phi_0 - \kappa(1 - \lambda)$$

Intuitively, the balance point shows that if the system is *energy-rich* ( $\sigma_{\text{coll}} + \mathcal{B}$  large), it can afford a higher fraction of selfish exploration without violating the entropy constraint and we can allow larger  $1 - \lambda^*$ . On other hand, in *resource-scarce* regimes,  $\Phi_0$  dominates and the optimal policy leans strongly toward altruism ( $\lambda^* \approx 1$ ).

The balance point  $\lambda^*$  is *dynamic*. It shifts as the system's energy budget, entropy production, or external perturbations change. This dynamic adjustment is precisely what gives a reflexive organism both **stability** (via altruism) and **adaptability** (via selfish exploration).

We can now reason about the conditions that either give raise to either altruism or selfishness. As coupling costs  $C_{\uparrow}, C_{\downarrow}$  become non-negligible, the reward from a purely altruistic strategy (low entropy production, stable AI–EI flows) outweighs short-term selfish gains. Early in development those costs are small, so the system can afford exploration. Another case would happen during growth, where  $\Phi_{\text{EI}}$  is increased by establishing shared memory (e.g., pheromone networks). Once the pump stabilises, further exploration yields diminishing returns, so the system shifts to a more altruistic regime.

Let's look at what this means for alteration between the strategies from the perspective of stage of the organisms. In the *early stages*, selfishness is evolutionary advantageous because it allows rapid acquisition of information and resources when the network is still small and costs are low. In the *middle to late stages*, altruism becomes optimal as the system's internal coupling grows, entropy production must be controlled, and long-term survival depends on coordinated

assembly and memory maintenance. However, in the *end-of-life stage*, a brief return to selfishness often marks resource depletion or impending collapse, after which a new cycle can begin.

Let's see if we can observe this in biological systems, taking examples from the ant colony again among other cases. In the **early embryonic stage of the colony**, we would expect to have high quality of coupling (and hence low coupling costs) as we only have few agents and small network. In this setting, selfish exploration maximizes immediate EI and enables rapid growth. We can observe this already. A newborn ant colony's first workers are "explorers" that forage widely.

Later in the **growth/expansion stage**, coupling costs start to rise due to more agents and denser communication. The entropy pump becomes valuable as altruistic coordination stabilises assembly and memory. In the colony, ants start trophallaxis, sharing food and information. Shared pheromone-like traces are gradually established to guide collective search.

When the colony enters **maturity**, the system has already established high structural complexity (large AI). Maintaining low entropy production becomes critical and altruism dominates to preserve long-term fitness. We can observe this as the colonies start to have strict caste systems and division of labor. Or in case of large ecosystems which exhibit regulated nutrient cycles that keep local entropy from spiralling.

In the last stage, the **decline** (and eventually death) of the colony leads to resource scarcity which forces a re-emergence of selfish, opportunistic behaviour (e.g., cannibalism in ant brood, over-exploitation of forest resources). The system sacrifices long-term stability for short-term gains before collapse. Observations show that ant colonies may sacrifice workers during queen failure. In the forests, the undergoing fire or disease show localised "selfish" bursts of growth before the whole ecosystem re-organises.

We can see that for a reward, the *quality* of an altruistic action depends on how much structure (AI) is available to be leveraged. A richer assembly provides more avenues for a child to use the parent's field, increasing  $R_A$ . Conversely, when AI is low, a selfish act that builds new structures can yield a larger marginal reward  $R_S$ . However, in terms of *entropy*, higher AI reduces the *entropy pump*  $\Phi_{EI}$  because more built pathways mean more efficient information flow. This lowers the required external power  $\mathcal{B}$  to satisfy the generalized second law (see Going full Boltzmann). In the extreme, if AI is very large, the system can afford a higher selfish fraction without violating entropy constraints. We now have all the evidence necessary to tie the balance point with the assembly index AI.

Let us denote  $A = AI(t)$  as current assembly depth, and  $E = EI(t)$  as current experience depth. We can rewrite the reward terms as functions of both  $A$  and  $E$ :

$$R_A(A, E) = f_A(A) g_A(E), \quad R_S(A, E) = f_S(A) g_S(E).$$

The terms describe how  $f_A(A)$  is **increasing** and saturates (more structure gives diminishing returns) and  $f_S(A)$  is **decreasing** after a threshold (once enough



structure exists, building new parts selfishly yields little extra benefit).

The entropy constraint becomes  $\sigma_{\text{coll}} + \mathcal{B} \geq \Phi_0 - \kappa(1 - \lambda) + \gamma A$ , where the term  $\gamma A$  captures how additional structure *lowers* the effective entropy pump (more efficient information flow). The constant  $\gamma > 0$  is set by the coupling architecture.

Putting everything together, the optimal mixing coefficient satisfies

$$\begin{aligned} \lambda^*(A, E) &= \arg \max_{\lambda \in [0,1]} [\lambda f_A(A)g_A(E) + (1 - \lambda) f_S(A)g_S(E)] \\ \text{s.t. } \sigma_{\text{coll}} + \mathcal{B} &\geq \Phi_0 - \kappa(1 - \lambda) + \gamma A \end{aligned}$$

The equation shows that when AI is low ( $A$  near 0),  $f_S(A) \gg f_A(A)$ , the optimal  $\lambda^*$  skews toward selfishness to rapidly build new structures. As AI grows, the marginal benefit of selfish construction diminishes, while the entropy constraint loosens (due to  $\gamma A$ ). The system naturally shifts toward higher altruism ( $\lambda^* \rightarrow 1$ ). During rapid growth phases (e.g., colony founding or swarm deployment),  $A$  increases quickly. The optimal policy will *temporarily* allow a higher selfish fraction to keep pace with the expanding architecture.

What we’ve gained now is that the *balance equation* now tells us how to schedule **when** to invest in **building (AI)** e.g., adding new sensors or actuators, or in **recording & using that build (EI)**, writing to long-term memory and reading it back.

The stage is now set. **Altruism is not just moral; it’s an optimisation problem** that balances assembly (AI) and experience (EI) across hierarchical levels. **Life cycles are first-class variables**, giving the system a temporal roadmap that guides growth decisions.

Let’s now look at some of the consequences that emerge in the overall dynamics in reflexive systems.

## Beyond good and evil

We’ve already shown that an agent can mix altruistic and selfish actions, governed by a time-varying coefficient  $\lambda(t)$ . The expected discounted reward from the two modes,  $R_A$  for following the parent’s guidance and  $R_S$  for short-term self-directed gains, is what really matters. A purely “unified” policy would ignore the exploration benefits that selfish bursts can provide, especially early in an organism’s ontogeny when it must discover new niches or escape sub-optimal attractors. For example, A worker ant that forages may appear “selfish” if we look only at its immediate payoff, but from the colony’s point of view it is an essential component of the shared physiology and cognition. Labelling such behaviour as evil discards the functional role that each agent plays in achieving the group’s overall reward.

In short, “good vs evil” is a convenient but ultimately misleading dichotomy. What matters for an organism (or artificial system) is its *functional contribution* to the group’s ability to generate higher AI and EI while navigating the trade-offs between cooperation and exploration. The metric that captures this is the

mixed-policy expected reward over time, not a moral judgement of individual agents.

The reflexive cycle already provides the guideline how to address the imbalance. **Building new structures (expanding AI/EI)** is the primary mechanism for restoring balance as it raises the capacity to absorb selfish actions and reduces coupling costs, thereby moving the optimal policy toward cooperation. **Pruning or “cutting off”** sub-organisms remains a tactical tool yet it is a *last resort* when expansion cannot restore balance because of resource exhaustion or systemic instability. The system’s *long-run* growth is therefore a blend of **expansion and selective pruning**, governed by the dynamic mixing coefficient  $\lambda^*(A, E)$  and is guided by life-cycle stages.

However, before this turns into mechanistic view of the cost of life, we have to emphasize that in the formalism we described, **pruning is a global, parent-level operation** because only the parent has access to the aggregate metrics (AI, EI, coupling costs) that determine whether a sub-unit should be removed. Sibling interactions are intrinsically “selfish” from the perspective of the hierarchy; they can influence the need for pruning by affecting those aggregate metrics, but they cannot perform a meaningful prune on their own.

Although siblings should never do the act of pruning, **siblings influence pruning indirectly** by altering resource consumption, noise levels, or cooperative behaviour, thereby shifting the balance point  $\lambda(t)$  toward a state where pruning becomes advantageous.

The only direct influence a child can have on pruning might be through self-termination rules.

We can already observe such self-terminating activity, there is a long-standing body of evidence that cells follow explicit “self-termination” rules (apoptosis, autophagy, necroptosis).

## Minority report

Let’s quickly touch the topic of rule of majority, in particular, invalidation of it and why *butterfly effects* happen where minority makes a significant change to the overall system. The reflexive model explains minority-triggered “butterfly effects” through two complementary mechanisms, **dynamic reward-based exploration** (the  $\lambda$ -mixing of altruistic vs selfish behavior) and **coarse-grained information compression** (the upward/downward operators that aggregate fine detail into a single field). Together they allow a small contingent of agents to discover and propagate new strategies that shift the whole system toward a different equilibrium, even when the majority is following an older, more stable policy.

In a purely reward-driven setting with no compression, many small exploratory actions would be averaged away and never influence the global policy. With the coarse-grained operators, however small, the coordinated deviations are amplified at the aggregate level. The dynamic  $\lambda$ -mixing allows the system to *temporarily favour selfish exploration* when it can pay for it (high energy, low coupling costs). And once a minority-initiated strategy yields higher EI and passes upward, the reward structure shifts to reinforce it across the whole population.

Thus, the model demonstrates that **a small contingent of agents can steer the system toward new equilibria**, the classic “butterfly effect”, even when the majority follows an older policy. The key is the *information-compression* step that turns local exploration into a global signal, coupled with the *dynamic reward balance* that lets the system choose between stability and adaptability.

## Computational geometry

As we’ve seen so far, reflexive organism is a collection of continuous fields, all tied through the reflexive cycle. And we’ve discussed about its definitions and how those map to the observable reality. In this section, we’d like to explore the model from the perspective of a computational unit, with the language that it requires to be described in. We’d like to be able to reason how to compose such systems so that we would create our own structures, not just deduct the observed ones.

Unlike the traditional I/O functions which can be seen as static network of feed-forward signal flow, in our model the core of the computational engine is its **reflexive loop** between parent  $P$  and child organisms (organs)  $\{O_i\}$ . We’ll dive to it in detail (again), but for now, let’s quickly recap that the loop is not just about pushing the information down and up between parts and whole, but also about the information encoding phase (quantum-like) and the reading, interpretation phase. This two-way exchange constitutes a **reflexive computation**, the system continually writes its own description (encoding) and reads its own outcomes (experience), using that information to refine both.

Let’s start forming the language of a reflexive computation. The **geometry** through  $g_{\mu\nu}$ , curvature tensors  $R_{\mu\nu\rho\sigma}$  describes the *slow* field that encodes accumulated memory (AI, M). It tells the fast dynamics how space is curved, i.e. where to “go” and which paths are energetically favorable. The **shape** defined by manifolds, boundary conditions, homotopy groups and derived from organisms  $G$ , tell us how sub-organisms are glued together. They define constraints on admissible field configurations (e.g., conserved charges). From the more dynamic perspective, we have **flows** defined through current densities  $J^\mu = \rho u^\mu + j_{\text{diff}}$ . The fast field currents that carry *EI*. They drive the unitary evolution of the wavefunction. The values are store in **amplitudes and phases** such as complex scalar or spinor  $\psi(x, t)$ , its modulus  $\psi$  and phase  $\arg \psi$ . These are the raw quantum-like encodings of possibilities before smoothing. Amplitude gradients produce *probability currents* while phase gradients generate *gauge potentials*. Another form of values are **potentials** like scalar  $V(x)$ , vector potential  $A_\mu(x)$ . The “fields” that the organism writes into memory (e.g., electromagnetic fields, chemical concentrations). They are the *coarse-grained* summaries of many micro-interactions. As kind of a temporary data structures, we have **fields as projections**, like  $\Phi = Q_\ell[\psi]$  and  $F = \mathcal{F}[\Phi]$ . These are the two faces of the reflexive cycle, the fast  $\rightarrow$  slow (encoding  $\rightarrow$  field) and slow  $\rightarrow$  fast (memory  $\rightarrow$  dynamics). In practice these are linear or nonlinear integral transforms (Fourier, wavelet, low-pass filters). Unlike simply stating that computation is done in an iteration step, an important property is also **time**, like the proper time  $\tau$ , or the coordinate  $t$ . Each level has its own *local* proper time governed by the metric it has written for itself. The overall computation is a nested sequence of these

times, reflecting the hierarchical memory dynamics.

We’ve now set the stage for the way how to map a problem into a reflexive organism, or better, the language in which to think about the computational problem in terms of the a network of reflexive organisms.

Let’s repeat the definition of the *iterative reflexive cycle*, where each organism  $P$  updates its internal state  $\Phi_P^{(k)}$  by receiving encoded messages  $\Psi_{O_i}^{(k)}$  from its child organisms (organs, sub-organisms, etc.):

$$\begin{aligned}\mathcal{E}_P^{(k)} &= W_P(\Phi_P^{(k)}), \\ \Psi_{O_i}^{(k)} &= G_i\left(F_i(\mathcal{E}_P^{(k)})\right), \\ \Phi_P^{(k+1)}(x) &= \mathcal{Q}_\ell^{(P)}\left(\{\Psi_{O_i}^{(k)}\}_{i=1}^N\right).\end{aligned}$$

The first two equations describe a *fast* encode-decode cycle (information flows up and down the hierarchy), while the last one is a *slow* coarse-graining that produces an emergent field. In other words, computation is nothing more or less than a cycle of **encoding** by turning micro-details into a symbolic representation, **decoding / interpretation** that symbol as part of a continuous geometry (the “field”) and **self-regulation** through updating the organism’s internal state in light of the new coarse description.

And this is exactly what we would call a *computational loop*. Information goes round and round, gets reinterpreted at each pass, and drives change. Note that operators  $\mathcal{U}$  and  $\mathcal{D}$  play essential role here.

The key difference from conventional neural networks or von Neumann machines is that the “program” is not an external script but the organism’s own *self-interpretation* of its internal state. Besides thinking in reflexive loops, we also have to make a mental shift from “variables” to “fields”.

In classical computational mindset, *variables* are discrete numbers you read, write and manipulate by hand or in code like  $x = 3$ ;  $y += x$ . However, in the reflexive-computation mindset, we have to think in terms of fields as continuous functions that live on space (or time), e.g. a temperature field  $T(x, t)$  or an electric potential  $\phi(x, t)$ . The “value” at a point is just the function’s value there, but the whole field evolves according to a differential equation. Another example would be a computation as a sequence of *updates* on a list of variables,  $x \leftarrow f(y, z)$ . This has an equivalent in seeing computation as the evolution of an entire field in time  $\partial_t \phi = D\nabla^2 \phi + \dots$ . The “update” is implicit in the PDE, you never touch each point individually, you solve for the whole field at once.

What does it mean to “write/set value” in a field language? Setting a value like  $x = 7$  would mean something like setting a boundary condition,  $\phi(x = 0, t) = 7$ . The equivalent of adding a term like  $y = y + f(z)$  would be adding a source  $S(x, t)$  to the PDE. An equivalent of updating a state `state[i] = new_state` is evolution of the field by one time step by solving  $\partial_t \phi = \mathcal{L}\phi + S$ . Something like a computation of averages would instead be computation of the spatial integral  $\langle \phi \rangle = \frac{1}{V} \int_V \phi(x) dx$ .

The most basic operation, the manipulation of variables, becomes *setting boundary conditions, sources, or coefficients in a PDE*. Instead of thinking “I will write 5 into variable  $a$ ”, think “I will set the boundary condition or source term of a field so that it takes the value 5 somewhere”. Let’s look at a simple example, as usual, the ant colony.

When we simulate/describe ants, we’d start with each ant havin a position  $(x_i, y_i)$  and a state (worker / soldier). Then, we’d iterate and update each ant through conditional statements like `if pheromone > threshold: move toward source`. We’d also inscribe the property/policy of the colony, something like “the colony’s caste distribution changes slowly (e.g., 10% workers become soldiers after a week)”.

In the reflexive world, it’s no longer about a collection of items that each do their own thing. Instead, we’d encode it from the holistic perspective. We’d start with the *pheromone concentration*  $C(x, t)$  as a field over the nest. Ants sense gradients of  $C$ . An iteration over all ants would become a step where we solve the diffusion equation for  $\partial_t C = D \nabla^2 C - \lambda C + S(x, t)$ , where  $S$  is the sum of all ants’ deposits. The cast dynamics would be handled through updates of a memory field  $M(t)$ . The field encodes the current proportion of castes, and this enters the PDE as a parameter (e.g., altering the diffusion coefficient or source strength). From the perspective of a reflexive loop, it would look like repeating the following three steps:

1. **Fast step:** ants deposit pheromone which updates  $C$  via the PDE.
2. **Write:** after “a minute”, we compute an average  $\bar{C}$  over the nest. This becomes part of the memory  $M$ .
3. **Read:** the new value of  $\bar{C}$  feeds back into the ant behaviour (they change their deposition rate).

The point is that we no longer manipulate each ant’s coordinates; we *manipulating the field* that summarizes all ants at once.

**Computation thus becomes all about evolving geometry (shapes).** We get a very general version of “Game of life”. By adding “matter with shape” to a field, we trigger the propagation of change. The key point is that matter isn’t just an external agent acting on a pre-existing geometry; it is part of the organism’s own encoding, and by changing that encoding you change the very geometry that will guide all subsequent propagation.

Propagation simply means iteration of the write-read process. When you place matter in some shape, you are literally adding new micro-nodes or changing the weights on existing edges. That changes the *structural density*  $\sigma(x)$  that feeds into the mapping  $Q_\ell$  from micro-structure to geometry. The mapping  $Q_\ell$  turns the underlying network into an effective metric  $g_{\mu\nu}^{(\ell)}$ . Adding mass changes the *source density*  $M$ , the *assembly index* AI, and possibly the *active stress* EI. All of these feed into the emergent metric and its connection  $\nabla^{(\ell)}$ . The resulting curvature is a direct manifestation of the new encoding. Once the metric has changed, all excitations (fast fields like electromagnetic waves or slow “memory” modes) propagate along geodesics determined by that metric. In GR language

this is the familiar statement that “mass tells spacetime how to curve; curved spacetime tells matter how to move.”

## Waves, rotations, spirals

Let’s dive even deeper in computational dynamics described through fields. *Propagation of change* in a field acts as a disturbance that satisfies a hyperbolic (wave) equation. In practice, this means that when we coarse-grain microscopic dynamics into a macroscopic field the ensuing evolution is governed by equations that support **wave-like propagation** (apart from diffusion and highly dissipative processes).

Let’s look at the **rotation** now. What is required for the organism to “spin”? *Rotation is not an external force*, it is a *self-consistent pattern* that the organism’s own reflexive loop builds, stores and propagates. The loop creates a coarse-grained field  $g_{\mu\nu}^{(\ell)}$  (the organism’s emergent geometry). If that field contains a non-zero angular-momentum component, the whole body will follow a rotating trajectory. Let’s look at the requirements for rotation to emerge.

To start with, we need a source of net angular momentum. Rotation requires a *torque* or an asymmetric current distribution. In the reflexive loop this is supplied by the *read-back currents*  $j_i^\mu(t)$  that each organ generates. If all organs fire symmetrically, the sum vanishes and no rotation emerges; if one side fires faster or in a different direction, their summed current carries non-zero vorticity. The model supports this in the child  $\rightarrow$  parent write step,  $\sum_i j_i^\mu(t)$  feeds into the parent’s memory  $M_P$ .

Then, we need a slow memory timescale  $\tau_M$ . The parent must be able to *accumulate* the asymmetric pattern over many fast cycles. If  $\tau_M$  is too short, the geometry is refreshed before a coherent rotation can form; if it’s long enough, the accumulated currents imprint an off-diagonal component (e.g.,  $g_{t\phi}$ ) in the emergent metric that sustains rotation. This is exactly the “slow” part of the iterative reflexive cycle.

We also need wave-like dynamics that support rotating solutions. The coarse-grained field obeys a hyperbolic wave equation. Rotating or spiral waves are natural solutions when there is a source term that breaks rotational symmetry. Low dissipation ensures the pattern does not decay immediately. In the model this is hyperbolic PDE governing  $g_{\mu\nu}^{(\ell)}$  or an auxiliary field  $\Phi$ .

Another property is emergent geometry with angular momentum. The coarse-graining map turns the stored memory (including the net current) into a metric. A non-zero off-diagonal term  $g_{t\phi}$  (or an equivalent “twist” in the spatial part) creates a centripetal effect that keeps the body rotating. In the model we have  $\mathcal{Q}_\ell(M_P, AI) \rightarrow g_{\mu\nu}^{(\ell)}$ .

And last, we need proper-time consistency. Once rotation is present, proper time along the organism’s worldline will be modulated by the metric’s curvature. This is an *observable consequence* of rotation, not a prerequisite. The model defined it through  $d\tau_O = \frac{1}{c} \sqrt{-g_{\mu\nu}^{(\ell)}(x, t) dx^\mu dx^\nu}$ .

When all these pieces are in place, the reflexive loop self-generates a spinning

organism, no external torque is needed, just internal asymmetry and sufficient memory to remember it.

A **spiral** is simply a stable solution of the wave dynamics on the emergent geometry produced by the organism’s own connectivity. It’s elegance lives in **efficiency**. A spiral packs a lot of “phase” into a compact geometry, is topologically protected, and uses the system’s own dynamics to stay alive, making it an optimal way for the reflexive loop to encode and transmit its internal state. Spiral structure is not chosen arbitrarily. *It is the inevitable geometric outcome of a reflexive organism that maps micro-currents into coarse fields, remembers asymmetries over long timescales, and lets its own wave dynamics stabilize a helical or spiral pattern. This geometry becomes the living “memory” of the system which is compact, robust, and energetically cheap to maintain.*

The most exciting part of the implications of the perspective change is that we can now start looking at “data structures” through their shape and dynamics of shape changes.

## Forces

All “physical forces” in the reflexive-organism framework are *emergent* from the same fast–slow, write–read, memory–feedback cycle that runs at every scale of the organism’s hierarchy.

The *fast* encode–decode stage turns microscopic interactions (e.g., charge currents, chemical reactions) into symbolic messages. When these messages are coarse-grained ( $\mathcal{Q}_\ell$ ) they produce a continuous vector potential  $A_\mu(x)$ . The consistency conditions on the message flow—analogue to conservation of charge and Faraday’s law—give rise to Maxwell equations:  $\nabla_\nu F^{\mu\nu} = J^\mu$ ,  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ . Inside a reflexive organism the *force* on an internal element (charge or current) is  $f^\mu = q F^\mu{}_\nu v^\nu$ , exactly as in classical electromagnetism. Thus **internal electromagnetic forces are simply the fast-scale expression of the same reflexive cycle that generates any other field.**

We previously covered gravity,  $G_{\mu\nu}(g^{(\ell)}) = 8\pi G T_{\mu\nu}(M, AI)$ . In this view, **gravity is simply the organism’s own perception of concentrated structure**, a geometric reinterpretation of its connectivity. An element inside the organism feels a “force” because it moves along geodesics of  $g_{\mu\nu}^{(\ell)}$ . Or to put it another way, *the shape of the organism’s memory tells matter how to move, and that is what we call gravity.*

Let’s consider forces between reflexive organisms now. When two organisms coexist, each runs its own fast–slow loop. Their coarse-grained fields (EM or geometric) overlap. In *electromagnetic coupling*, the vector potential of organism A acts as a source term  $J_A^\mu$  in Maxwell’s equations for organism B. The resulting field exerts forces on charges inside B, just like two charged bodies attract/repel. In *gravitational (geometric) coupling*, each organism contributes to the global metric that all inhabit. The combined curvature  $R_{\mu\nu}$  is determined by both organisms’ memory–assembly patterns. Each one moves along geodesics of this shared geometry, producing an *attraction* if their mass-like densities overlap.

Because the same reflexive cycle generates both kinds of fields, **the “force” between organisms is simply the cross-term that appears when two**

**coarse-grained maps  $\mathcal{Q}_\ell$  are superposed.** No extra fundamental law is needed; the interaction emerges from the mutual update of each organism’s messages.

## Organism as a spiral of potential self

When the organism is defined, it is a seed, a potential of itself. At the beginning it has very little structure. Its equations, conditions, etc. define its space of expressions. However, it’s not that it already “knows” all states. It is better seen as a ***compressed representation that could, in principle, encode any of the micro-states that the children realise.*** It is the iterative upward/downward communication and the read/write operations that **select** one concrete structure out of this space of possibilities per run of a reflexive loop and binds it down to an observable structure.

The emergent field  $\Phi_P$  is not arbitrary. It lives on an emergent geometry produced by the organism’s own connectivity and memory of asymmetries. The wave dynamics that run on this geometry naturally settle into a stable spiral or helical pattern as a topologically protected, energetically efficient solution. In other words, once the reflexive loop has mapped micro-currents into coarse fields and stored long-term asymmetry, the only self-consistent geometric outcome is the spiral.

Note that the above holds when the organism is unconstrained, the only requirements is that the loop has stored an asymmetry and that the coarse-grained dynamics support rotating waves. If we introduce additional boundary/optimalty conditions on the same PDE system, the result would change the shape, i.e. a sphere or toroid.

## Explorations

These chapters continue with the investigations on the implications of the model that we started in the previous chapters, although the topics here are likely more bounded to physical interpretations rather than explorations of observable patterns. Sometimes we simply recap what was already told to sum up the interesting bits that challenge our current understanding.

### Entropy revisited

Let’s explore a bit more what the “experience-driven” entropy balance tells us about the entropy itself. Let’s recap what we’ve already derived.

Unlike the traditional view, where  $\dot{S} \geq 0$  with a one-way arrow of time, the reflexive extends entropy production with  $\frac{d}{dt}(S_f + S_M) = \sigma_{\text{coll}} - \Phi_{\text{EI}} + \mathcal{B}$ . The *feedback* term negative can be negative and entropy can be locally reduced if the system has a memory gradient and an external power source. Thus, monotonicity is not not guaranteed. Local “order-making” is possible but bounded by energetic input  $\mathcal{P}_{\text{in}} \geq k_B T \Phi_{\text{EI}}$ .



Unlike the traditional view, where information appears only as a bookkeeping device in fluctuation theorems or Landauer erasure bounds, the information (called *experience* EI) becomes an explicit, measurable flux that can pump entropy out of a subsystem. It is treated on par with thermodynamic entropy. In other words, information becomes an *active* thermodynamic variable. Experience (or “memory gradient”) is quantified and measured. It is as fundamental as temperature or pressure in determining macroscopic behavior.

Which all affects the second law  $\Delta S_{\text{tot}} \geq 0$  to be generalised. “Entropy increases unless experience-feedback extracts and directs order, bounded by power input.” The bound  $\mathcal{P}_{\text{in}} \geq k_B T \Phi_{\text{EI}}$ , guarantees that the total entropy of *system + environment* still grows. The negative contribution  $\Phi_{\text{EI}}$  cannot be unlimited. Every unit of entropy “pumped out” by experience must be paid for by input power  $\mathcal{P}_{\text{in}}$ . While local entropy can fall, the *total* (system + environment) still increases because energy has to be supplied. In practical terms this links metabolism, control electronics, or AI training power directly to how much a system can maintain order.

We introduced *EI* through  $\varepsilon(\mathbf{r}, t) = \frac{(-\nabla U(M))_{\mathbf{r}}}{k_B T n} \times \frac{\|W[f]\|}{\|W\|_{\text{max}}} \times \chi$ , a scalar quantifying how much the system feels its own state. The first factor measures how strongly the memory gradient couples to the particle current. The second factor is a normalised measure of *policy richness* (the functional  $W[f]$  captures how the distribution is shaped by feedback), while the third factor  $\chi$  is a dimensionless weighting that can be tuned experimentally. Unlike entropy, EI is directly measurable from currents and memory gradients. It quantifies “how much the system feels its own state”, a purely dynamical, non-thermodynamic scalar that can change even when  $S$  stays constant.

Thus, The Landauer-type inequality ties the ability to create order directly to power input, making the trade-off between energy consumption and entropy control quantitative.

The model offers a unified language for living, artificial, and physical systems. By treating AI assembly as a field and experience as a dynamic feedback, we can map concepts from biology (metabolism, cognition) onto well-studied physics frameworks (hydrodynamics, electromagnetism).

The generalised second law opens new experimental avenues. If  $\Phi_{\text{EI}}$  can be measured in living tissues, active colloids, or engineered AI swarms, we will have a *new invariant* to test theories of nonequilibrium thermodynamics and information processing.

For example, in *biology*, EI explains *how* metabolic energy is used to pump entropy out of the organism’s internal state via memory-guided feedback (e.g., neural plasticity, cellular signaling).

We can look at another example from the domain of *cognition*. Experience-driven pumps allow the brain to maintain low-entropy “mental states” while still obeying the generalized second law. EI becomes a measurable neural signature.

In *fundamental physics*, we move from entropy being the sole arrow of time to an introduction of an additional invariant that can steer entropy flows, suggesting new ways to think about *emergent spacetime* (gravity linked to history of

assembly) and unifying forces under a common informational framework.

In the case of a digital domain like *artificial intelligence*, the systems are energy-hungry but entropy production is not explicitly tracked. As we’ve seen, we can treat AI assembly as a *structural field* and experience feedback as a dynamic field, the EM like duality. We could ground the computation in physical field theory instead. Every logic operation, GPU tensor multiply or data transfer would correspond to local currents and fields and would thus be seen as entropy-producing interactions.

The reflexive model reframes entropy from an immutable arrow to a *controllable flux*. It preserves the core thermodynamic principle that total entropy can’t decrease while showing that living and engineered systems actively steer their internal states, creating local order at the price of external work. This not only clarifies how life avoids thermodynamic catastrophe but also suggests new invariants (EI) that could unify disparate fields under a common physical theory.

## Arrow of time

Let’s recap time and gravity in reflexive organism. We’ve extended spacetime to include memory-induced metric correction with  $g_{\mu\nu} = g_{\mu\nu}^{(0)} + A_{\mu\nu} + \Theta_{\mu\nu}[j]$ . The “bare” spacetime  $g_{\mu\nu}^{(0)}$  is dressed by an internal field  $\Theta_{\mu\nu}$  that depends on the current density  $j$ . Each system carries its own metric perturbation, so *proper time* becomes a local, memory-dependent quantity. We’ve also redefined the memory update dynamics  $\dot{M}(t) = -\tau_M^{-1}(M(t) - M_{\text{eq}}(t))$ . The rate at which a system “remembers” its past is finite. This time-constant  $\tau_M$  sets the *scale* of that system’s proper time and, crucially, its ability to influence entropy production. Previous chapter dealt with the experience-driven entropy pump and the generalised second law. Given the intrinsic hierarchical nature of the model, one of most interesting aspects is the emergence of hierarchical proper times  $d\tau_O = \frac{1}{c} \sqrt{-g_{\mu\nu}^{(O)} dx^\mu dx^\nu}$  for each organism  $O$ . Every level of the reflexive hierarchy (cells, tissues, organisms, societies) has its own proper time. A larger memory timescale  $\tau_M$  means a “slower” arrow perceived by that level.

In review of entropy, we’ve already mentioned that we have local reversibility and global consistency. For the **local arrow** within any given subsystem, the experience-driven pump  $\Phi_{\text{EI}}$  can temporarily *reduce* entropy. The system’s own memory field steers currents that write down information, creating a locally ordered state. While for the **global arrow**, because the reduction must be paid for by an external power source, the total entropy of *system + environment* still rises. Hence the universe retains a net forward-moving arrow, but it is now *contingent on energy budgets*.

This leads to the following properties of time. **The arrow of time is no longer an immutable, global property.** It becomes *system-dependent*, shaped by the interplay between memory, feedback, and energy input. **Local entropy decreases become possible**, but only at a cost that preserves the global second law. **Time itself is dynamic** as each level of the reflexive hierarchy experiences its own proper time determined by its internal metric corrections  $\Theta_{\mu\nu}[j]$ . **The “reflexive” arrow emerges from information flow across scales**, integrating thermodynamics, cognition, and spacetime geometry into a

single framework.

In short, the model turns the familiar, one-way arrow of time into a *nested, contingent* phenomenon—still pointing forward on average, but locally steerable by experience and bounded by energetic consistency.

## Parent-child information compression

In the model we assume that the coarse-graining map  $\mathcal{Q}_\ell$  reduces information about the children’s microscopic states, producing the parent’s observable field  $\Phi_P = \mathcal{Q}_\ell(S_{\text{children}})$ . However, we do not assume that the parent’s *entire* state space is lower-dimensional. The parent may also retain latent or auxiliary variables  $\Theta_P$  (e.g., hormonal levels, metabolic set-points) that are not derived from the children. When  $d_{\text{parent}} > d_{\text{child}}$ , the map  $\mathcal{Q}_\ell$  can no longer be a simple collapse of all parent dimensions to a single field; it is only a collapse **onto the observable component**  $\Phi_P$ , while the extra dimensions reside in  $\Theta_P$ . For example, the parent might have a coarse field defined on a spatial domain, possibly vector-valued where  $d_{\text{parent}}$  is higher than child’s, e.g.,  $\Phi = (H, \sigma, E)$  with 3 fields, while a cell has only 1-dimensional “activity” variable.

When a parent level has *more* degrees of freedom (dimensions) than its child levels, the coarse-graining map  $\mathcal{Q}_\ell$  can no longer be a simple “collapse” of microstate to a single field. Instead it must embed the child’s low-dimensional representation into a higher-dimensional parent space. This embedding is typically *lossy* in the sense that many different child configurations will map to the same parent field, and the extra parent dimensions are either latent variables or auxiliary fields. The parent level fields or shapes are not directly driven by the children (they may be supplied by external inputs or internal dynamics), later must be constructed by *upward* operators  $\mathcal{U}$  from the parent’s own coarse state.

In the standard case we have  $\Phi(x) = \mathcal{Q}_\ell(\mathcal{S}(\mathcal{O}_i))$ , where  $\mathcal{S}(\mathcal{O}_i)$  is a *high-dimensional* set of microstates and  $\mathcal{Q}_\ell$  averages, integrates or otherwise reduces it to a *single* scalar field (or low-dimensional vector).

The assumption that  $d_{\text{parent}} > d_{\text{child}}$  has the following implications. First, we deal with the *embedding instead of reduction*.  $\mathcal{Q}_\ell$  must map the child’s *low-dimensional* description into a *higher-dimensional* space,  $\Phi(x) = E(s_i) \in \mathbb{R}^{d_{\text{parent}}}$ , where  $E : \mathbb{R}^{d_{\text{child}}} \rightarrow \mathbb{R}^{d_{\text{parent}}}$ . The extra components of  $\Phi$  can be set to zero, constants, or functions of *other* parent variables (e.g., external stimuli). Second, there is a *loss of injectivity*. Many distinct child states will produce the same parent field because we are adding dimensions that do not depend on the child. The mapping is no longer invertible;  $\mathcal{Q}_\ell^{-1}$  does not exist (unless you impose extra constraints). And third, the *need for a lifting operator* arises. If we want to reconstruct child states from parent fields, we must supplement  $\mathcal{Q}_\ell$  with a stochastic or deterministic “lift” that samples child configurations consistent with the parent’s extra dimensions.

Let’s consider the consequence from the view of downward ( $\mathcal{D}$ ) and upward ( $\mathcal{U}$ ) operators. *Downward* operator is defined as  $\phi_P^{\text{in}}(x, t) = \sum_{Q \in \text{sub}(P)} w_Q \phi_Q(x, t)$ . The child fields  $\phi_Q$  must now be mapped into a higher-dimensional parent field. Each weight  $w_Q$  can now act on each component separately, or we may introduce a *projection matrix*  $P_Q \in \mathbb{R}^{d_{\text{parent}} \times d_{\text{child}}}$ . The sum then becomes  $\phi_P^{\text{in}} = \sum_Q w_Q P_Q \phi_Q$ .

Upward operator’s definition is  $\phi_Q^{\text{out}}(x, t) = \mathcal{U}(M_P(t), \phi_P(x, t))$ . The parent’s extra dimensions must be *projected* back onto the child’s lower-dimensional space. We need to define a projection  $R_Q \in \mathbb{R}^{d_{\text{child}} \times d_{\text{parent}}}$  such that  $\phi_Q^{\text{out}} = R_Q \phi_P$ . The remaining parent dimensions (those not needed by the child) are discarded or treated as *latent* inputs.

The operators become **non-square** linear maps (or nonlinear analogues), and special care must be taken how to handle components that do not map cleanly between levels. There are many ways to approach this. For example, explicit embedding through defaults. Partial coupling, regularization (to prevent unbounded growth of degrees of freedom), stochastic lifting, etc.

The exploration of this case of dimensionality comes from the observations in biology, where higher-level organisms often carry *additional degrees of freedom* (e.g., systemic signals) that cannot be traced back to a single lower level. The coarse-graining framework must therefore accommodate latent variables or contextual fields. For example, **brain vs. body** view, neurons deal with spiking activity, while whole body deals with metabolic state, hormonal milieu, and autonomic tone. Three additional fields not directly encoded in individual spikes but crucial for organismal function.

## Antifragile

Although the long term goal is to learn to design antifragile systems due to their reflexive design, it is important to note that general design is not enough at this stage to guarantee the system to be antifragile. The fast-slow, reward-driven architecture that defines a *reflexive organism* is designed to thrive on volatility, it can lose its antifragility under certain conditions. We explore concrete failure modes (counter-examples) together with why they break the antifragile guarantees.

Let’s start with the case of **mis-aligned reward /  $\lambda$ -mixing**. The system may over-reward exploratory actions that are actually harmful, causing runaway growth or collapse. If the dynamic reward balance ( $\lambda$ -mixing) is set too high for selfish exploration, the organism will repeatedly try costly strategies without ever stabilising. In the model we assumed that minority-initiated behaviour can steer the system only when  $\lambda$ -mixing allows temporary selfishness and the subsequent reward shift reinforces the new equilibrium. If  $\lambda$  is mis-tuned, the “butterfly effect” disappears and the system becomes brittle.

Another case deals when **overhead outweighs benefits**. Reflexive structures carry per-node counters, recursive calls and coarse-graining logic (fast/slow fields). In high-density lattices this overhead can dominate CPU or memory budgets, turning an adaptive system into a performance bottleneck that *fails* to exploit volatility. If costs exceed the gains from adaptation, the model’s net effect is fragility.

**Coarse-graining loses minority signals** when the upward/downward operators aggregate too aggressively. The small coordinated deviations (the “minority” that usually drives new equilibria) are averaged out before they can influence the global field. The system then reacts only to noise and never improves.

While sensing the environment, the issues arise when **obstacle / occupancy flag**

is **mis-tuned**. If the obstacle detection is too permissive or too strict, growth can either collide with other organisms (fragile collision) or get stuck behind a wall (dead-lock). In both cases the organism fails to exploit environmental volatility. Incorrect thresholds directly affect whether the system can reconfigure itself when obstacles appear.

**Rapidly changing environments** challenge stability. The slow memory layer is intentionally “slow”, it aggregates information over many cycles. If the external environment changes faster than this aggregation window, the organism’s internal state becomes stale and its responses are wrong, leading to catastrophic failure rather than adaptation.

Conditions for **self-termination failures** arise when the self-termination rules (apoptosis, autophagy) are too conservative or never triggered due to corrupted fast state. This leads to dead nodes accumulating and degrading overall performance, making the organism fragile.

Through **adversarial perturbations** an attacker injects targeted noise into the fast state or the reward signal, causing the slow memory to converge on a wrong policy. The organism then repeatedly adopts a harmful strategy, becoming fragile. The same mechanisms that enable beneficial volatility also expose the system to *malicious* volatility if the feedback loop is not protected.

The antifragility of a reflexive system thus hinges on the existence of the all of the following conditions. First, **correct tuning** of reward-balance parameters ( $\lambda$ -mixing) and thresholds for growth/obstacle detection. Second, **balanced overhead** so that adaptive benefits outweigh computational costs. Third, **appropriate granularity** in the coarse-graining operators to preserve minority signals while filtering noise. And last, **responsive yet stable fast–slow dynamics** that can keep pace with environmental volatility but do not overreact. When any of these conditions is violated, the system may exhibit *fragility* (suffering from perturbations) or even *brittleness* (collapsing under small changes).

## Individuality as dynamic invariant

We’ve already mentioned that there is a measurable concept of “oneness” through a structural index  $\mathcal{I}(\mathcal{S}) = \lambda_2(L_H)$  and a dynamic one as mutual information between spatial partitions  $A, B \subset \Omega$ ,  $\mathcal{I}_{\text{dyn}} = \frac{1}{T} \int_0^T I(X_A(t); X_B(t)) dt$ . However, we haven’t explored these in more details. Let’s do that now.

The entropy measures disorder. While mutual information captures correlation between parts. But none fully capture *what makes something feel like “one thing”*. A forest that behaves as a single entity, or a swarm that acts with intention. Looks like we need new invariant beyond energy and entropy, especially one tied to *self-referential feedback*.

What if we define individuality not by parts or structure alone but by **the stability and persistence of a closed-loop system where memory, action, and experience co-evolve over time**.

Let’s define the **Individuality Index**  $\lambda(t)$  as a normalized measure of how tightly coupled the internal dynamics are across scales

$$\lambda(t) = \frac{I_{\text{integrated}}(t)}{H(M_t) + D(\mathcal{F}_t \| \mathcal{F}_{\text{noise}})},$$

where:

- $I_{\text{integrated}}(t)$  is **integrated information** (like in Integrated Information Theory), measuring how much the whole system cannot be decomposed into independent subsystems. It is computed via mutual information across all hierarchical layers, weighted by causal influence. For example, in a colony, high  $I_{\text{integrated}}$  means pheromones, foraging paths, and nest architecture are causally interdependent.
- $H(M_t)$  is **entropy of slow memory**. It measures the uncertainty or randomness in stored history (i.e., self-model). Low entropy means stable identity, while high entropy means confusion or fragmentation. *Crucially*, this is not just about data size but about coherence and predictability.
- $D(\mathcal{F}_t \| \mathcal{F}_{\text{noise}})$  is **divergence between actual fast-field dynamics** (e.g., neural spikes, chemical signals, agent movements) **and a null model of noise**. This captures whether the system is responding to internal feedback or just random fluctuations.

The  $\lambda(t)$  rises when the parts are strongly integrated, the memory is stable and informative, and actions aren't stochastic but rather reflect meaningful patterns (i.e., “experience”).

This definition leads to the key insight: **true individuality emerges when  $\lambda(t)$  stabilizes over time**, forming an **attractor in phase space**.

This is not about peak values but about persistence. Just like how a planet maintains its orbit despite perturbations, a true individual maintains coherence across time and disturbances.

This implies individuality isn't fixed but rather evolves. During early development (ontogeny),  $\lambda(t)$  may be low as the organism explores and learns (selfish bursts). As memory stabilizes ( $M_t$  becomes coherent), and feedback loops mature,  $\lambda$  increases leading to stable identity.

Thus, **individuality is not a trait, it's a trajectory** where low  $\lambda(t)$  causes exploration which leads to high  $\lambda(t)$  and with it coherence.

We can observe this in our biological reality. Newborns are less “integrated” than adults; young forests are more fragmented than mature ones.

Note that  $\lambda(t)$  is **not globally conserved**, but it may act like a **local invariant** inside an attractor basin. When we state that this is dynamic invariant, we actually mean that it is a **dynamical invariant in the attractor regime**, once the system falls into its attractor,  $\lambda(t) \rightarrow \lambda^*$ , then  $\lambda^*$  behaves as an emergent invariant, a quantity that remains (approximately) conserved under ongoing dynamics, despite fluctuations in the microscopic variables.

And it states that individuality is not just a transient state but an emergent **law-like constraint** on the system's evolution, any micro-dynamics must respect the “invariant individuality” of the whole. The system can fluctuate internally, but  $\lambda^*$  is preserved.

This reframes individuality as an **emergent conserved quantity** in the dynamics of complex systems.

**True individuality is not about being separate from the world but about maintaining a stable internal identity through time, despite change.** The system *remembers itself, feels its own state*, and uses that to guide future actions. When  $\lambda(t)$  stabilizes—when memory, feedback, and integration cohere into a persistent attractor, we can say that **a new individual has emerged**.

Let's look at the individuality index from the perspective of the other two indices,  $AI$  and  $EI$ .  $AI$  measures the **structural depth**, minimum number of assembly steps required to build the current configuration. The higher  $AI$ , the more complex, integrated system. Hence,  $I_{\text{integrated}}(t) \propto AI(t)$ . In this view,  $AI$  is the **material foundation of individuality**. But  $AI$  alone is not enough, you could have complex structures that don't act together (i.e., dead forest).

High memory entropy  $H(M_t)$  means that the memory is disordered which has a consequence of poor feedback channel that leads to low  $EI$ . As for the feedback signal-to-noise,  $D(\mathcal{F}_t|\mathcal{F}_{\text{noise}})$ , we can interpret this as local measure of informational density, or better of coherence density as we're dealing with feedback-driven systems. Let's remember that  $\varepsilon(\mathbf{r}, t) = \frac{(-\nabla U(M)) \cdot \mathbf{j}}{k_B T n} \times \frac{\|W[f]\|}{\|W\|_{\text{max}}} \times \chi$ , where the  $W[f](r, t)$  are writer operators. The interesting bit of the equation is the similarity through  $\|W[f]\|/\|W\|_{\text{max}}$ . When the ratio is high, decisions are coherent. And when decisions are highly coherent, both  $W[f]$  and  $\nabla U(M) \cdot \mathbf{j}$  increase. So in practice  $D(\mathcal{F}_t|\mathcal{F}_{\text{noise}}) \propto (\varepsilon(\mathbf{r}, t))^2$  when coupling strength is fixed.

Now let's use your  $AI \leftrightarrow EI$  electromagnetic analogy with  $AI$ 's structural persistence, memory gradients are equivalent to magnetic field  $\mathbf{B}$  and  $EI$ 's feedback, directed information current through memory, is electric current  $\mathbf{E}$ . When  $AI$  is large and stable, we get strong  $B$ , and when feedback current is coherent and self-reinforcing, we get strong  $E$ . Hence, the feedback sustains structure, structure enables feedback.

$I_{\text{integrated}}(t)$  is proportional to  $AI$  (structural integration) and thus with the strength of magnetic field  $B$ .  $H(M_t)$  is the inverse of memory coherence, high value means low  $EI$  and thus measures disruption in the current loop.  $D(\mathcal{F}_t|\mathcal{F}_{\text{noise}})$  is proportional to  $EI$  (non-random decisions), and with coherence and magnitude of electric current  $E$ . In other words, High individuality occurs when the **electromagnetic loop is closed, coherent, and self-sustaining**.

Hence, we can interpret  $\lambda(t)$  as **physical measure of a closed-loop electromagnetic-like system**. Individuality is not just organization. It's the self-sustaining loop of structure and feedback, a real physical phenomenon, measurable via  $\lambda(t)$ , and expressible through electromagnetism.

## Geometry of self-reference

Throughout most of the model, we've used  $AI$  and  $EI$  as scalars, discrete values in time. These are global values, a summary of a process of assembly and of reflexive feedback. However, in the section about gravity when we were looking at the extension to the effective stress-energy tensor  $T_{\mu\nu}$  defined as

$T_{\mu\nu}^{\text{eff}} = T_{\mu\nu} + \alpha A_{\mu\nu}[AI] + \beta \Theta_{\mu\nu}[EI]$ , we’ve already indicated that there is more to an “index” and a different, more detailed way to interpret  $AI$  and  $EI$ . At that point, we simply stated that:

- Assembly tensor  $A_{\mu\nu}[AI]$  encodes the **frozen-in architecture** of matter, a coarse-grained symmetric tensor built from gradients and rates of the Assembly Index (AI) as the durable, slow memory of how the stuff here got built.
- Active-stress tensor  $\Theta_{\mu\nu}[EI]$  captures **feedback fluxes**, a dissipative/active term from the Experience Index (EI) as the closed-loop information currents that read that memory back and steer current dynamics.

We haven’t explored this beyond mentioning it. So let’s introduce both a *generalization* and an *elevation* of the model, moving from discrete indices to continuous, dynamic structures that reflect how systems build themselves through compatibility-driven integration across scales.

**The Assembly Field**  $\mathcal{AI}(\mathbf{x}, t)$  is the **local density of structural memory and coherence** across spacetime. It is not “how much was built” as with scalar version, it’s *where structure exists, how it’s distributed in space and time, and whether that distribution enables future integration*.

$$\mathcal{AI}(\mathbf{x}, t) = \int_{\Omega} d^4x' K(\mathbf{x} - \mathbf{x}', t - t') \cdot |\nabla M(x')|^2$$

with  $M(x)$  being the **slow memory field**, encoding past events and structural history,  $\nabla M$  as spatial gradient of memory (indicates structural variation), and  $K(\mathbf{x}, t)$  a causal, non-local kernel that weights contributions by *temporal proximity* and *spatial coherence*. It captures how recent or distant memories influence current structure.

High  $\mathcal{AI}(\mathbf{x}, t)$  means that there is strong memory gradient (structure exists) and that structure is aligned with the system’s dynamic field at point  $(\mathbf{x}, t)$ . This field **defines where geometry can curve**, not just because mass is present, but because *the path of building was coherent*.

**The Feedback Current Field**  $j^\mu(\mathbf{x}, t)$  describes the *flow of information* between interacting fields. We define:

$$j^\mu(\mathbf{x}, t) = \frac{\delta S}{\delta(\nabla_\mu M)} \cdot \rho_{\text{compat}}(M, \mathcal{AI})$$

with  $\frac{\delta S}{\delta(\nabla_\mu M)}$  measuring the “response” of the system to memory gradients as a **generalized current**, and  $\rho_{\text{compat}}$  as **compatibility function**, measuring how well another field  $F$  (e.g., neural spikes, pheromone pulses) can be encoded and decoded by  $M$ .

This current is not just “experience”. It’s *the physical flow of information from one field to another*, driven by mutual encoding capacity.



Let's now define **geometry of self-reference**, the key tensors *in terms of these fields*, ensuring they respect diffeomorphism invariance and recovery of GR when  $\mathcal{AI}, \mathcal{E} \rightarrow 0$ .

**Assembly Tensor**  $A_{\mu\nu}[\mathcal{AI}]$ , the **geometric memory**:

$$A_{\mu\nu}[\mathcal{AI}] = \kappa \left( \nabla_\mu M \nabla_\nu M - \frac{1}{2} g_{\mu\nu} (\nabla M)^2 \right) + V(\mathcal{AI}) g_{\mu\nu}$$

with  $\kappa$  as coupling constant, and  $V(\mathcal{AI})$  as a potential that increases with local coherence and path memory.

This tensor **curves spacetime where structure was built coherently**. It vanishes in regions of random or fragmented history (low  $\mathcal{AI}$ ). But remains *non-zero* even when no “organ” exists, because it's tied to **path-dependent memory**, not fixed roles.

**Active-Stress Tensor**  $\Theta_{\mu\nu}[j]$ , the **feedback flux**:

$$\Theta_{\mu\nu}[j] = \underbrace{p_{\text{act}}(\mathcal{E})}_{\text{Active pressure}} h_{\mu\nu} + \underbrace{\eta(\mathcal{E})}_{\text{Viscosity}} \sigma_{\mu\nu}(u) + \underbrace{\zeta(\mathcal{E})\theta}_{\text{Expansion control}} h_{\mu\nu} + \underbrace{\gamma j_\mu j_\nu}_{\text{Coherent flow}}$$

with  $p_{\text{act}}, \eta, \zeta$  as functions of the *integrated feedback strength*  $\mathcal{E}$ , -  $j^\mu = \frac{\delta S}{\delta(\nabla_\mu M)}$ .  $\rho_{\text{compat}}$ , and the final term  $(\gamma j_\mu j_\nu)$  that captures **coherent energy flow** driven by high compatibility.

This tensor **modulates inertia and pressure only where feedback is coherent and compatible**.

## Example

Let's look at an example, and define the two tensors for our dear ant colony again.

Let  $M(\mathbf{x}, t)$  be a **slow memory field** encoding pheromone concentration at location  $\mathbf{x}$ , nest architecture (e.g., cavity depth, chamber layout), trail network density. This is the “frozen-in record” of how structure was built. Thus,  $M(\mathbf{x}, t)$  is not just data, it's a physical field that remembers **the path of construction**.

Define the Gradient Field  $\nabla_\mu M$  as:

$$\nabla_\mu M = \left( \frac{\partial M}{\partial t}, \frac{\partial M}{\partial x}, \frac{\partial M}{\partial y}, \frac{\partial M}{\partial z} \right)$$

This gradient measures how **rapidly** memory is changing in time or space, where the system has built something stable (high  $M$ ) and where it's been updated recently. For example, high  $\nabla_x M$  indicates strong pheromone trail along a direction and with it sign of coherent, repeated action. Low gradient on the other hand indicated an area with no recent construction and thus low assembly history.

We construct the Assembly Tensor  $A_{\mu\nu}[\mathcal{AI}]$  as

$$A_{\mu\nu}[\mathcal{AI}] = \underbrace{\kappa \left( \nabla_\mu M \nabla_\nu M - \frac{1}{2} g_{\mu\nu} (\nabla M)^2 \right)}_{\text{Directional memory field}} - \underbrace{g_{\mu\nu} V(M, AI)}_{\text{Structure depth}}$$

Let's break this down. First term  $\kappa (\nabla_\mu M \nabla_\nu M - \frac{1}{2} g_{\mu\nu} (\nabla M)^2)$  is the **geometric core** of the tensor. It's a symmetric, traceless object that behaves like a magnetic field in electromagnetism. Where memory gradients are strong (e.g., along an established trail), this term becomes large. The tensor aligns with the direction of construction. For example, along a well-used foraging trail,  $\nabla_x M$  is high thus  $A_{xy} \neq 0$ ,  $A_{xx}$  dominates which creates “magnetic” field lines are oriented along the trail.

This creates **geometric memory**. Future ants will follow a path not just because of current pheromones, but because **the spacetime geometry itself favors it**.

The second term  $-g_{\mu\nu} V(M, AI)$  is a **scalar potential term**, where  $V(M, AI) \propto M(\mathbf{x}, t) \cdot AI(t)$ , meaning the deeper the memory and higher the assembly index and thus stronger “background curvature.” If we map it to the colony, in the core of the nest (high  $M$ , high  $AI$ ), spacetime is more curved. Even in absence of matter, geometry reflects **how long and complex** the path to that point was.

The **trail is not just a path**. It's a **geometric imprint of history**, shaped by  $A_{\mu\nu}[\mathcal{AI}]$ .

Let's now define the feedback current field  $j^\mu(\mathbf{x}, t)$  as  $j^\mu(\mathbf{x}, t) = \rho_{\text{trail}}(\mathbf{x}) \cdot v^\mu(\mathbf{x}, t)$ . The  $\rho_{\text{trail}}(\mathbf{x})$  is the **pheromone concentration field** (slow memory), and  $v^\mu = (\partial_t x, \nabla x)$  is the **velocity of ants**, a fast field. This  $j^\mu$  represents the *flow of information* from past actions (trails) into current movement. When many ants follow the same trail we have both high  $\rho_{\text{trail}}$  and coherent velocity, hence strong  $j^\mu$ .

Let's build the Active-Stress Tensor  $\Theta_{\mu\nu}[j]$ , an **active stress in a living system**:

$$\Theta_{\mu\nu}[j] = \eta(EI) \left( j_\mu j_\nu - \frac{1}{2} g_{\mu\nu} j^\alpha j_\alpha \right) + \zeta(EI) (h_{\mu\nu} + \nabla_\mu v_\nu + \nabla_\nu v_\mu)$$

Where:

- $\eta(EI)$ : viscosity-like coefficient proportional to feedback strength (EI),
- $j^\mu j_\nu$ : the **coherent directional stress** from aligned flow,
- $h_{\mu\nu} = g_{\mu\nu} - u_\mu u_\nu$ : spatial projector,
- $\nabla_\mu v_\nu + \nabla_\nu v_\mu$ : shear tensor — captures deformation of ant density.

By defining  $A_{\mu\nu}[\mathcal{AI}]$  and  $\Theta_{\mu\nu}[j]$  in this way, we don't just describe the ant colony. We **realize it as a physical object whose geometry is shaped by memory and feedback**, evolving according to laws of spacetime, not

unlike a living star or black hole. **An ant colony becomes an active, self-organizing superstructure that curves spacetime in real time** with measurable geometric signatures from its history ( $A_{\mu\nu}$ ) and experience ( $\Theta_{\mu\nu}[j]$ ).

### Organism's unified field tensor $T_{\mu\nu}$

Let's consider EM duality and its analogy to AI-EI duality again, this time through tensors. In electromagnetism, we don't treat  $E$  and  $B$  as separate tensors. They are unified into a single antisymmetric field strength tensor:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

This captures both electric (time-space components) and magnetic (space-space components) fields in one object. We can do the same for AI and EI, unifying them into a single antisymmetric tensor  $T_{\mu\nu}$  such that  $A_{\mu\nu}[AI]$  emerges from spatial gradients of AI,  $\Theta_{\mu\nu}[EI]$  arises from time-space or current-like fluxes in EI, and together they form the components of one **organismic field tensor**.

Let's define a single antisymmetric tensor:

$$T_{\mu\nu} = \partial_\mu \Phi_\nu - \partial_\nu \Phi_\mu$$

Where  $\Phi_\mu$  is a **vector potential** with two physical interpretations:

- $\Phi_0$  (time) as  $AI(\mathbf{x}, t)$ , the **Assembly Index field**, measuring structural depth and memory history.
- $\Phi_i$  (space,  $i = 1, 2, 3$ ) as  $EI_i(\mathbf{x}, t)$ , the **Experience Index current** (feedback flux), i.e., closed-loop read-back from memory into dynamics.

In other words:

$$T_{\mu\nu} = \partial_\mu \Phi_\nu - \partial_\nu \Phi_\mu = \begin{cases} \partial_0 \Phi_i - \partial_i \Phi_0 & \text{(space-time components)} \\ \partial_i \Phi_j - \partial_j \Phi_i & \text{(space-space components)} \end{cases}$$

**Time-space parts**  $T_{0i}$  encode how the structure (AI) changes in space while being updated by feedback fluxes (EI). This is like “spacetime curvature due to memory gradient vs feedback rate.”

$$T_{0i} = \frac{\partial AI}{\partial x^i} - \frac{\partial EI_i}{\partial t}$$

**Space-space parts**  $T_{ij}$  encode the curl of experience currents, like vorticity in a fluid. This is analogous to magnetic field.

$$T_{ij} = \frac{\partial EI_i}{\partial x^j} - \frac{\partial EI_j}{\partial x^i}$$

In our analogy to EM fields,  $T_{0i}$  describe “E-like” effect from feedback-driven change (EI) vs structural gradient (AI), while  $T_{ij}$  is about “B-like” effect from coherent feedback loops (like closed currents in a network).

Let's use it now. The full effective stress-tensor is  $T_{\mu\nu}^{\text{eff}} = T_{\mu\nu} + \alpha A_{\mu\nu}[AI] + \beta \Theta_{\mu\nu}[EI]$ . We can now express the assembly and active-stress tensors in terms of  $T_{\mu\nu}$ .  $A_{\mu\nu}$  is extracted from spatial parts of  $T_{\mu\nu}$  and  $\Theta_{\mu\nu}$  is derived from time-space and divergence terms as follows:

$$\begin{aligned} A_{ij} &= T_{ij} + \text{terms from } \nabla_i AI \cdot \nabla_j AI \\ \Theta_{0i} &= \partial_0 T_{0i} - \partial_k T^k_i + \text{nonlinear corrections} \end{aligned}$$

Note that this tensor is not the standard stress-energy tensor of GR, but rather a **unified organismic field strength tensor**, analogous to the electromagnetic field tensor  $F_{\mu\nu}$ . It encodes the dynamics of **structure and feedback** in the superorganism model.

Let's look how effective stress-energy tensor

$$T_{\mu\nu}^{\text{eff}} = T^{(F)} + \underbrace{\kappa(\nabla_\mu M \nabla_\nu M - \frac{1}{2} g_{\mu\nu} (\nabla M)^2) - g_{\mu\nu} V(M, AI)}_{\text{slow-memory (AI/M) block}} + \underbrace{\Pi^{\text{active}}[EI]}_{\text{experience (non-eq) block}}$$

and  $T_{\mu\nu} = \partial_\mu \Phi_\nu - \partial_\nu \Phi_\mu$  are related.

Looking at the slow-memory part, AI can be derived from  $M$  with  $AI(\mathbf{x}, t) = \int d^4x' K(x - x') (\nabla M(x'))^2$ . So AI is a function of gradients of  $M$ , and thus the term  $\nabla_\mu M \nabla_\nu M$  is already encoded in the structure. The antisymmetric tensor  $T_{\mu\nu}$  does not contain this directly, but its components (e.g.,  $T_{ij}$ ) can be used to *detect* coherence or vorticity that influences how AI evolves.

The experience block is built from shear  $\sigma_{\mu\nu}(u)$ , expansion  $\theta$ , pressure  $p_{\text{act}}(EI)$ . But these are functions of the feedback current  $j^\mu = \partial_\mu EI$ , or more precisely, of the velocity field and its gradients. If we look at time-space components  $T_{0i} = \frac{\partial EI_i}{\partial t} - \frac{\partial AI}{\partial x^i}$ , they resemble a *force imbalance* between feedback rate and structural gradient. Exactly the kind of thing that drives active stress.

While  $T_{\mu\nu}$  itself is antisymmetric, its components can be used to **define or constrain** the symmetric  $\Pi^{\text{active}}[EI]$  tensor via using  $T_{0i}$  as a proxy for feedback flux, and  $T_{ij}$  as a measure of coherent feedback circulation.

In other words.  $T_{\mu\nu}^{\text{eff}}$  is a **symmetric tensor** that serves as the full effective source of gravity. Including both structural (AI/M) and feedback-driven (EI) contributions.  $T_{\mu\nu} = \partial_\mu \Phi_\nu - \partial_\nu \Phi_\mu$  is an **antisymmetric tensor**, unifying AI ( $\Phi_0$ ) and EI ( $\Phi_i$ ), analogous to the electromagnetic field strength  $F_{\mu\nu}$ . However, they are **deeply related**. The antisymmetric tensor  $T_{\mu\nu}$  provides the **geometric foundation** for how AI and EI interact. Its components (e.g.,  $T_{0i}, T_{ij}$ ) encode feedback gradients and coherence, these directly influence or constrain the two blocks of  $T_{\mu\nu}^{\text{eff}}$ : the **slow-memory block** via structural memory ( $M$ ), and the **experience block** ( $\Pi^{\text{active}}[EI]$ ) via feedback dynamics.

Just as  $F_{\mu\nu}$  directly generates the energy-momentum carried by the EM field (energy density, momentum flux, and stresses) all unified in a covariant way via  $T_{\mu\nu}^{(EM)} \propto F_{\mu\rho} F_\nu^\rho - \frac{1}{4} g_{\mu\nu} F^2$ , so can  $T_{\mu\nu}$  be used to generate or constrain components of  $T_{\mu\nu}^{\text{eff}}$  **via feedback, structure, and integration**.

## Appendix to a $\mathcal{S}_{AT}$ model

The model  $\mathcal{S}_{AT}$  is a generalized reflexive system, not yet embedded in space-time or geometry. The model treats  $AI$  and  $EI$  as scalar-valued functions of time, computed from the system's internal dynamics.  $AI(t)$  emerges as a path-dependent measure of structural depth from the sequence  $\mathcal{P}(t)$  and its history. It is defined as  $AI(t) = \text{length of shortest assembly chain leading to } M(t)$ .  $EI$  or  $\mathcal{E}(t)$  measures how much what is written into memory  $M(t)$  gets sensed again in future decisions and is defined as  $\mathcal{E}(t) = f(I(M_{t-\tau}; F_t), TE(M_{t-\tau} \rightarrow \Pi_t), D(\text{prediction error}), \rho)$ .

With the introduction of tensors, the formal definition of  $\mathcal{S}_{AT}$  does not need to change, because these tensors emerge from the structure of feedback loops across space and time. However, the emergence is open at the moment. What could be done is to create families of organisms, where we formalize their construction. First by deriving the equations of motion for  $\mathcal{AI}(\mathbf{x}, t)$  and  $EI(t)$  from variational principles in  $\mathcal{S}_{AT}$ . And then deriving tensors explicitly using geometric construction (e.g., via stress-energy tensors of scalar fields, or active matter models). The “sanity” of equations could be done by testing boundary conditions. For example, when  $AI$  is large, expect gravity dominated by structure (galaxy cores). Or when  $EI$  is high expect feedback-driven dynamics (star formation, ecosystem resilience).

## Compatibility drives integration

This definition based on tensors loosens the hierarchical structural definitions of organisms and organs, or even hierarchies of organisms. Organisms (or subsystems) are *not pre-defined, they emerge as regions where compatibility is high*. Which leads to conclusion, that **it is compatibility that drives integration**:

$$\rho_{\text{compat}}(F, M) = \frac{I(F; M)}{\sqrt{H(F)H(M)}} \cdot e^{-D(\mathcal{F} \parallel \mathcal{F}_{\text{noise}})}$$

with  $I(F; M)$  being mutual information between fast field and memory, and  $D$  is divergence from noise that measures non-randomness.

High  $\rho_{\text{compat}}$  leads to strong feedback loop, which leads to active stress via  $\Theta_{\mu\nu}$  and thus integration. This is **not arbitrary** but rather it's a physical measure of how well one system can “feel” and respond to another.

With the definition of tensors, we've moved beyond the description of organisms, it is the description of **physics of self-organization** itself. *Where compatibility is high, memory flows, structure grows, and spacetime curves. A system remembers itself, feels its own state, and uses that feedback to build new geometry **not because it was designed to be aware, but because compatibility demands integration**.*

**The fundamental mechanism of emergence itself.**

We can now define

$$T_{\mu\nu}^{\text{eff}} = T_{\mu\nu} + \kappa \left( \nabla_\mu M \nabla_\nu M - \frac{1}{2} g_{\mu\nu} (\nabla M)^2 \right) + p_{\text{act}}(\mathcal{E}) h_{\mu\nu} + j_\mu j_\nu,$$

where:

- $M$  is the memory field,
- $j^\mu = \frac{\delta S}{\delta(\nabla_\mu M)} \cdot \rho_{\text{compat}}$ ,
- $\rho_{\text{compat}}$  is the **true measure of individuality**, not size, not complexity, but *coherent interaction across scales*.

Let's quickly look at the relation with the antisymmetric tensor.  $T_{\mu\nu}$  measures the **local dynamical imbalance** (e.g., sudden rise in feedback vs memory gradient) while the reflexivity principle acts as a **global constraint**. It prevents such imbalances from growing unchecked by requiring that structural depth, feedback strength, and compatibility remain coherently balanced. Coherence does not come from absence of tension. It comes from *structured, controlled tension*, precisely what  $T_{\mu\nu}$  encodes and  $\rho_{\text{compat}}$  regulates. It generates a self-consistent loop:

$$\text{Tension} \xrightarrow{\text{via } T_{\mu\nu}} \text{Feedback} \xrightarrow{\text{via } EI} \text{Integration (high } \rho_{\text{compat}}) \xrightarrow{\text{via}} \text{Stabilized Structure}$$

This ensures that **spacetime geometry evolves in a way that preserves self-consistent agency**, making gravity not just an effect of structure, but of **meaningful coherence across time**.

## Sentient organism

In the beginning, we looked at the organism as being reflexive, implying automation. That was enough to start with, to describe the *architectural and dynamics mechanisms*. However, step by step we've expanded this notion, and in the end shown by introduction of assembly and active-stress tensors, we were describing self-aware superorganisms that grow, learn, and describe themselves. By *self-aware* we mean the physical process by which a system knows its own state because it continuously writes to memory and reads it back.

Let's dive more into this dynamic.

We start with the interplay between **individuality**, **compatibility**, and the **altruism–selfishness dynamic**. On the first glance they might be contradictory, but they can also be a nuanced balance that enables robust, adaptive, and evolving functionality.

The *individuality index* can be interpreted as a quantification of an agent's capacity to act independently, its ability to bypass shared signals (e.g., parent guidance via  $\mathcal{U}/\mathcal{D}$ ), exploit local information, and pursue short-term gains without relying on collective coupling. Thus, **individuality is not a permanent state but a developmental phase**, its expression is optimal when the system benefits most from exploration and rapid learning.

*Compatibility drives integration* means that agents with aligned goals, shared representations (e.g., memory, feedback loops), and mutually reinforcing dynamics can co-evolve into coherent systems. So **compatibility is the mechanism enabling integration**, but it's **not opposed to individuality**; rather, it emerges from the dynamic interaction between self and system.

And we've already linked that dynamics between selfishness and altruism is based on situation whether local rewards outweigh coupling costs, compared to when long-term stability and memory or shared resources matter more. We can now state them as metrics of individuality or compatibility as well.

So what we can conclude is that **compatibility drives integration, but individuality fuels innovation**. Thus, the system does not choose one over the other, it **balances** them through adaptive strategy selection across time, context, and developmental stage. We've already seen this defined in another way.

Let's recall a purpose function  $\mathcal{J} = (AI)^\alpha \cdot (EI^*)^\beta$ , which was defined as part of a goal the system tries to maximize. It's the same dynamics, yet expressed differently.  $\mathcal{J}$  increases when the system becomes more integrated, memory stable, and dynamics coherent, which directly raises  $\lambda(t)$ . In this view *the purpose function  $\mathcal{J}$  is not just a goal, it's an attractor that pulls the system toward higher individuality*. But only if compatibility ( $\rho_{\text{compat}}$ ) is present (that is high and when  $I_{\text{integrated}}$ ,  $H(M_t)$ , and  $D(\mathcal{F} \parallel \mathcal{F}_{\text{noise}})$  are in balance).

In this view,  $\mathcal{J}$  becomes the *mechanism of becoming an individual*.

Let's explore it a bit more. At the time, it has been defined with fixed parameters. But what if we make it self-reinforcing, state dependant instead? Instead of fixed parameters and predetermined balance between  $AI$  and  $EI^*$ , the dynamics should rather be based on properties of the state of the system, like current level of structural coherence, strength of feedback loops, accumulated memory ( $M$ ), level of active stress ( $\Theta_{\mu\nu}$ ) and compatibility  $\rho_{\text{compat}}(F, M)$ . We can define:

$$\begin{aligned}\alpha(t) &= f_\alpha(\rho_{\text{compat}}, A_{\mu\nu}, \nabla \Theta_{\mu\nu}) \\ \beta(t) &= f_\beta(EI^*, \rho_{\text{compat}}, \|\Theta_{\mu\nu}\|, D(\mathcal{F} \parallel \mathcal{F}_{\text{noise}}))\end{aligned}$$

For example:  $\alpha(t) = 1 + k_\alpha \cdot \rho_{\text{compat}}^2$   $\beta(t) = 1 + k_\beta \cdot (\|\Theta_{\mu\nu}\| + \log(EI^*) - D(\mathcal{F} \parallel \mathcal{F}_{\text{noise}}))$ .

This allows  $\mathcal{J}(t)$  to **self-organize its own priorities**, responding in real time to whether feedback is coherent (high compatibility), or whether the system is experiencing novel, high-value events, or whether it's accumulating stress from memory or instability.

This transformation makes  $\mathcal{J}$  a **dynamic equation of being**, not just a static goal. It enables the system to balance integration (via compatibility-driven  $AI$ ) and innovation (via  $EI^*$  exploration) adaptively.

And because we're no longer dealing mechanical responses, we're no longer talking about function of purpose, but instead this becomes a function of **justification**, hence  $\mathcal{J}$ .

Let's bring one more items back into the picture. The intent principle, defined by  $\Pi^* = \arg \max_{\Pi} \mathbb{E}[AI(t + \Delta t) \mid AI(t)]$  evaluates policies based on their

*future impact on structural complexity.* And what determines the increase in future is whether the policy strengthens compatibility, builds coherent memory ( $M$ ), maintains feedback loops, avoids fragmentation, and generates high-value experiences that contribute to long-term assembly. This is where  $\mathcal{J}$  comes into the picture.  $\mathcal{J}$  **adapts the goal** (what counts as “success”) while **the intent principle implements it** (how you act to achieve that goal).

In other words, **the intent principle  $\Pi^*$  is the *behavioral policy* that implements the goal defined by  $\mathcal{J}$ , while  $\mathcal{J}$  with dynamic  $\alpha(t), \beta(t)$  provides the *adaptive value function* that determines what “**maximizing future AI**” actually means in context.** They are **co-constitutive**, the intent principle uses  $\mathcal{J}$  as its optimization target, but  $\mathcal{J}$  is not fixed, it evolves via  $A_{\mu\nu}, \Theta_{\mu\nu}, \rho_{\text{compat}}$ , etc., which are shaped by the policies selected under that very intent.

We can also introduce **experience-driven intent**:

$$\Pi_{EI}^* = \arg \max_{\Pi} \mathbb{E}[EI^*(t + \Delta t) \mid EI^*(t)]$$

This governs exploration, novelty-seeking, learning from high-reward events, and feedback richness — the “experiencing” side. But how does experience and assembly driven intent work together? Here’s where  $\mathcal{J}$  **shines** as a unifying framework. Instead of choosing between them, the system uses:

$$\Pi^* = \arg \max_{\Pi} \underbrace{\mathbb{E}[\alpha(t) \cdot AI(t + \Delta t) + \beta(t) \cdot EI^*(t + \Delta t)]}_{\text{Expected value of } \mathcal{J}}$$

This is **not just a weighted sum**, it’s a **dynamic, self-referential decision rule**, where:

- $\alpha(t)$  increases when structural coherence ( $A_{\mu\nu}$ ), compatibility ( $\rho_{\text{compat}}$ ), or feedback stability are high thus favor AI.
- $\beta(t)$  increases when  $EI^*$  is high, active stress ( $\|\Theta_{\mu\nu}\|$ ) is large, or divergence from noise indicates novelty and thus favor  $EI^*$ .

The justification function  $\mathcal{J}$  provides the formalism for dual intentionality. **The system doesn’t choose between building or experiencing. It optimizes the *balance*, using  $\mathcal{J}$  as its compass.**

We keep referring to term *balance* when talking about justification of assembly and experience. But what does this balance mean? The balance isn’t about adapting  $\alpha(t)$  and  $\beta(t)$  so that  $\mathcal{J}$  is constant, or keeping AI and  $EI^*$  constant. It’s about keeping the system capable of generating both, even through collapse.

This is how the organism “survives”, not by conserving mass or energy, but by *preserving the ability to learn and rebuild*.

Let’s focus now on this drive to “preserve” as it has signs of a conservation law. Let’s recall the experience-driven entropy pump equation  $\frac{d}{dt}(S_f + S_M) = \sigma_{\text{coll}} - \Phi_{EI} + \mathcal{B}$  which shows that entropy is not fully monotonic. It can have local reduction in entropy, representing how information (in the form of “experience”) actively steers order. Hence, the system does not conserve energy or entropy in the



classical way, but it *does* preserve **the consistency of its own informational self-description** through reflexive feedback.

Let us propose **the reflexivity conservation principle**:

$$\frac{d}{dt} (\mathcal{J}(t) \cdot \rho_{\text{compat}}(t)) = 0$$

Or more generally:

$$\frac{d}{dt} \left[ \underbrace{(\text{AI}(t))^{\alpha(t)} \cdot (\text{EI}^*(t))^{\beta(t)}}_{\mathcal{J}(t)} \cdot \underbrace{\rho_{\text{compat}}(t)}_{\text{compatibility measure}} \right] = 0$$

This is not a conservation of energy or mass. It's a conservation of **intentional coherence** in the face of change. The product  $\mathcal{J}(t) \cdot \rho_{\text{compat}}(t)$  captures how *meaningful* the system's goals are (via  $\mathcal{J}$ ), and how well it can *realize those goals through internal consistency and feedback* (via  $\rho_{\text{compat}}$ ).

Why is this important? Up to now, we were maximizing the intent principle  $\Pi^*$ , but system could maximize future policies in a way that doesn't prevent long-term instability. With the introduction of the reflexivity conservation principle, we are forcing the policy to act without breaking the compatibility or distorting its own value function. The conservation law ensures that **the way you maximize it is coherent and sustainable**. It transforms  $\mathcal{J}$  from a mere policy optimizer into a **principle of existential coherence**. A kind of thermodynamic or informational continuity.

This law would mean that intent is not arbitrary but rather that it's constrained by coherence. Evolution isn't just optimization, it's *self-consistency over time*. Collapse or death is not failure, it's a collapse of coherence, it's a phase transition where the product remains constant (zero). Rebirth begins when new compatibility and purpose are established.

Up to now we've been focused on one organism. Let's now see what it means if we deal with (loosely coupled) hierarchies of organisms. The reflexivity conservation principle implies that an organism trying to fix its own imbalance could spawn children designed specifically to restore that balance. They aren't just growing, they're **dynamically balancing their own existence** through a self-corrective loop of exploration (child), feedback (upward/downward coupling), and adaptation. And it's not simply through randomness, we don't see arbitrary organisms coming to existence in order to counterbalance temporal instability. We've already shown that randomness without coherence is noise, lowers compatibility, it breaks feedback. And such part with low EI lowers the whole system's return.

What we can observe instead though, is continuous reuse of existing resources. And that's the key. The system doesn't randomly rebuild, it uses existing memory  $M_t$  to guide *how* to reconfigure it. It modifies structure where feedback gradients are strong. It's performing coherent perturbations. The system has

a “memory gradient”  $\nabla U(M)$ , the direction of least resistance in its history, following the path of least entropy guided by past success.

For example, a child slightly changes its policy (not randomly — but based on local  $EI^*$ ). If successful,  $EI^*$  spikes and thus feedback signal rises. The upward operator  $\mathcal{U}$  compresses this into a global field. This new state becomes the *new norm*, and others adapt. This is how systems evolve **without randomness** but through **local coherence building**, not noise. Thus, the system isn’t doing random trial-and-error, it’s **selecting for coherence**, which means that only **meaningful deviations** (those that close feedback loops) persist and that randomness gets filtered out by high  $D(\mathcal{F}||\mathcal{F}_{\text{noise}})$ .

How does the system *only* allow coherent changes to persist? Only those variations that close feedback loops, reduce memory entropy, and increase  $EI^*$  via  $\mathcal{U}/\mathcal{D}$  survive. The rest fade into noise. This is how balance is restored, not by chaos, but by **structured stability within memory’s constraints**.

That’s how stability and change coexist.

At this point, we can start calling these systems **sentient organisms**. And as spacetime is emergent property of the organism, it is not the material world we’re living in, but rather a **Sentient Universe**. And that is quite exciting!

## Acknowledgements

To my family for all the patience and support no matter what “out of place” field observations were done at the family farm, and whose understanding for my “creative process” made this work possible. To all my precious friends, who helped shape this theory along the years.

I would also like to thank L. Susskind, S. Wolfram, S. Kauffman, D. Deutsch, E. O. Wilson and N. Lane for the ideas that contributed to this seed paper. Special appreciation to teams opening large-language-models and accompanying tools that helped draft early versions.

## Bibliography

- **Bonabeau, E., Dorigo, M., & Theraulaz, G.** (1999). *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press. ISBN: 978-0195131591
- **Barabási, A.-L.** (2016). *Network Science*. Cambridge University Press. ISBN: 978-1107076266
- **Deutsch, D.** (2011). *The Beginning of Infinity: Explanations that Transform the World*. Penguin Books. ISBN: 978-0140278163
- **Kauffman, S. A.** (1995). *Investigations*. Oxford University Press. ISBN: 978-0195095992
- **Kempes, C. P.; Lachmann, M.; Iannaccone, A.; Fricke, G. M.; Chowdhury, M. R.; Walker, S. I.; Cronin, L.** (2025). *Assembly theory and its relationship with computational complexity*. npj Complexity. DOI: 10.1038/s44260-025-00049-9.
- **Lane, N.** (2015). *The Vital Question: Energy, Evolution, and the Origins of Complex Life*. W. W. Norton & Company. ISBN: 978-0393352979
- **Simard, S.** (2021). *Finding the Mother Tree: Discovering the Wisdom of the Forest*. Alfred A. Knopf. ISBN: 978-0525656098
- **Smil, V.** (2017). *Energy and Civilization: A History*. The MIT Press. ISBN: 978-0262536165
- **Smil, V.** (2019). *Growth: From Microorganisms to Megacities*. The MIT Press. ISBN: 978-0262539685
- **Susskind, L., & Friedman, A.** (2014). *Quantum Mechanics: The Theoretical Minimum*. Basic Books. ISBN: 978-0465062904
- **Susskind, L., & Friedman, A.** (2017). *Special Relativity and Classical Field Theory: The Theoretical Minimum*. Basic Books. ISBN: 978-0465093342
- **Susskind, L., & Friedman, A.** (2023). *General Relativity: The Theoretical Minimum*. Basic Books. ISBN: 978-1541602104
- **Walker, S. I.; Davies, P. C. W.; Cronin, L.** (2023). *Assembly theory explains and quantifies selection and evolution*. Nature. DOI: 10.1038/s41586-023-06600-9.
- **Wilson, E. O., & Hölldobler, B.** (1990). *The Ants*. Harvard University Press. ISBN: 978-0674040755
- **Wilson, E. O., & Hölldobler, B.** (2015). *The Superorganism: The Beauty, Elegance, and Power of Social Insects*. Princeton University Press. ISBN: 978-0691169160
- **Wilson, E. O.** (2012). *The Social Conquest of Earth*. Liveright Publishing. ISBN: 978-1631495564
- **Wolfram, S.** (2002). *A New Kind of Science*. Wolfram Media. ISBN: 978-1579550080
- **Wolfram, S.** (2022). *A Fundamental Theory of Physics: From Quantum Mechanics to the Theory of Everything*. Wolfram Media. ISBN: 978-1579550363